Thailand Statistician
January 2019; 17(1): 41-74
http://statassoc.or.th
Contributed paper

An Exploratory Data Analysis for Average Treatment Effect Estimation based on Partial Balancing and Simultaneous Inference of Regression Models

Rowena F. Bastero*
Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland 21250 USA.
*Corresponding author; e-mail: rowenab1@umbc.edu

Received: 18 March 2018
Revised: 22 June 2018
Accepted: 26 July 2018

Abstract

In order to provide significant outcomes, it is imperative that health care professionals, medical practitioners and policy-makers acquire evidence of the effectiveness of different treatments and programs. This is most commonly done by looking at treatment and control groups and determining if the treatment has a causal effect on the outcome. Ideally, treatment assignment is performed through randomization so that the groups formed are comparable with respect to their features. However, some factors such as cost, time, and ethical issues behind the treatment, may make it difficult to assign treatments at random. This leads to the use of observational studies instead of randomized studies in assessing the causal effect. While an observational study has the same intent as any randomized study, which is to estimate a causal effect, it differs in one major design issue: the lack of randomization in the allocation of units in the treatment and control groups. Due to this, systematic differences in the covariates of the treatment and control groups may exist which pose an inherent problem in estimating average treatment effect (ATE). While the use of propensity score is standard in this kind of situation, in this paper, an exploratory method is presented to determine the ATE in groups that are made homogeneous apriori with respect to the categorical variables. The proposed method begins with forming homogeneous subgroups in terms of qualitative features which reduces any bias induced by systematic differences in the covariates between groups. Regression models on the treatment effect given the continuous covariates are then separately generated for the treatment and control groups. The magnitude of difference in both models is then determined using simultaneous inference in regression, which creates confidence bands that provide graphical representations of the ATE on subgroups formed. This procedure presents an advantage in being able to determine the nature of covariates where the ATE is significantly positive; hence, one is able to provide effective solutions to a more personalized level. Two real data sets are utilized to illustrate the proposed procedure. The data analyses show the capability of the method to establish for which covariate regions the treatment is effective or not.

Keywords: confidence bands, covariate regions, propensity score, simultaneous regression inference.

1. Introduction

In causal inference, the potential outcome framework is founded on the idea that every unit has a pair of potential outcomes $\left(Y^{(0)}, Y^{(1)}\right)$ which refers to the unit's response had it been assigned to the control and treatment group, respectively. However, in observational studies, having both outcomes at once is impossible and consequently, one is always missing. The observed value for each subject is then defined as

$$Y_i = D_i Y^{(1)} + (1 - D_i) Y^{(0)} \tag{1}$$

where $D_i = 0$ if the i^{th} unit is in the control group and $D_i = 1$ if the i^{th} unit is in the treatment group. Under this setup, the treatment effect for each subject is $Y^{(1)} - Y^{(0)}$ and the average treatment effect (ATE) is defined by $E\left[Y^{(1)} - Y^{(0)}\right]$, which is the target estimand for inference.

In randomized controlled trials, the treatment allocation is not confounded by the covariates. This allows for the direct comparison of outcomes between the treatment and control groups. However, in observational studies, systematic differences in the covariates may exist between the two groups; hence, estimates derived by direct comparison are biased because differences in the response between the two groups may be attributed to the disparity in covariates and not the treatment effect itself. Under such situations, methods based on propensity scores (Rosenbaum and Rubin 1983; Austin 2011) have been proposed to remove bias caused by confounding. These methods proper estimation of ATE so that the difference in treatment outcomes realized is truly attributed to the treatment and not on the said systematic differences in covariates between groups.

However, propensity score-based methods may not fully guarantee balance within the two groups in terms of covariate information. It would be of interest to look into the behavior of the ATE given certain constraints on the covariates. In this paper, a technique is proposed that allows assessment of the treatment effect based on the subject covariate profile. In the proposed method, the natural imbalance between the two groups is addressed by forming subgroups that are perfectly balanced with respect to the categorical features of the data. Separate regression modeling of the outcome of the treatment and control groups allow for estimation of the treatment effect on partially homogeneous groups. The method based on simultaneous inference aids in determining the magnitude of difference in treatment and control and identifying the regions of the covariates where the treatment is effective. The combination of these components is geared towards providing a better procedure for defining effectiveness of a treatment in an individualized level.

This exploratory mechanism is advantageous for two reasons. First, partial balance is achieved due to the creation of independent classes based on the categorical variables. This eliminates any bias induced by the categorical covariates since units in the generated classes are made as similar as possible with respect to their categorical characteristics. Second, more extensive analysis of the subjects' covariate profile is performed. This provides a comprehensive idea on the behavior of the covariates for which positive and negative ATEs are realized. Consequently, medical practitioners are presented with a way to identify patients for whom the intended treatment will be effective. However, as a limitation, this procedure is only applicable for small number of covariates; although, this is typical of observational studies. Variable selection may also be performed for selection of a smaller subset of variables. A more detailed discussion of simultaneous inference is presented in Section 2. The proposed exploratory analysis is discussed in detail in Section 3. Real data sets are also considered to illustrate its use, shown in Section 4. Concluding remarks are in Section 5.

2. Simultaneous Inference in Regression Models

The average treatment effect can be assessed using regression models of the treatment and control groups which describe the relationship of the outcome Y on the same set of predictors $x_1, x_2, ..., x_p$. In general, suppose the two linear regression models are

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e}_i, \ \ i = T, C \tag{2}$$

where $\mathbf{Y}_i = (Y_{i,1}, Y_{i,2}, ..., Y_{i,n_i})'$ is a vector of outcomes, $\mathbf{X}_i$ is an $n_i \times (p+1)$ design matrix with full column rank and the first column is a vector of $\mathbf{1}$'s while the rest of the columns are $(X_{i,1}, X_{i,2}, ..., X_{i,n_i})'$, $\boldsymbol{\beta}_i = (\beta_{i,0}, \beta_{i,1}, ..., \beta_{i,p})$ is a vector of unknown regression coefficients and $\mathbf{e}_i = (e_{i,1}, e_{i,2}, ..., e_{i,n_i})$ is a vector of random errors where $e_{i,j} \sim N(0, \sigma^2)$, $i = T, C$, $j = 1, 2, ..., n_i$. Under this setup, $\mathbf{Y}_T$ and $\mathbf{Y}_C$ are the observed outcomes of treatment and control groups, respectively, that depends on the same p covariates. While it is important to determine if the two models are different, the magnitude of dissimilarity of these regression lines is of more interest for the estimation of the ATE.

To assess the magnitude of difference between the two models, a simultaneous confidence band

$$\mathbf{x}'\boldsymbol{\beta}_T - \mathbf{x}'\boldsymbol{\beta}_C = \left(1, x_1, x_2, ..., x_p\right)\boldsymbol{\beta}_T - \left(1, x_1, x_2, ..., x_p\right)\boldsymbol{\beta}_C \tag{3}$$

to bound the difference between the two models over the whole covariate space is generated (Liu, 2011). To estimate this band, suppose $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_T - \hat{\boldsymbol{\beta}}_C$ and $\boldsymbol{\beta} = \boldsymbol{\beta}_T - \boldsymbol{\beta}_C$. It can be verified that $\hat{\boldsymbol{\beta}} \sim N_{p+1}\left(\boldsymbol{\beta}, \sigma^2 \boldsymbol{\Delta}\right)$ where $\boldsymbol{\Delta} = \left(\mathbf{X}_T'\mathbf{X}_T\right)^{-1} + \left(\mathbf{X}_C'\mathbf{X}_C\right)^{-1}$, $\dfrac{\hat{\sigma}}{\sigma} \sim \dfrac{\chi_\nu^2}{\nu}$ with $\nu = n_T + n_C - 2\left(p+1\right)$ and $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$ are independent random variables. Hoel (1951) and Scheffe (1953, 1959) generalized an exact $1-\alpha$ simultaneous confidence band for $\mathbf{x}'\boldsymbol{\beta}_T - \mathbf{x}'\boldsymbol{\beta}_C$ over the whole covariate space as

$$\mathbf{x}'\boldsymbol{\beta}_T - \mathbf{x}'\boldsymbol{\beta}_C \in \mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm \sqrt{\left(p+1\right) f_{p+1,\nu}^{\alpha}}\, \hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}, \ \ \forall \mathbf{x}_{(0)} \in \mathbb{R}^p \tag{4}$$

where $f_{p+1,\nu}^{\alpha}$ is the upper α point of an *F*- distribution with degrees of freedom $p+1, \nu$ is as defined above and $\hat{\sigma}^2 = \dfrac{n_T - p - 1}{n_T + n_C - 2\left(p+1\right)}\hat{\sigma}_T^2 + \dfrac{n_C - p - 1}{n_T + n_C - 2\left(p+1\right)}\hat{\sigma}_C^2$ with $\hat{\sigma}_T^2$ and $\hat{\sigma}_C^2$ being the respective mean square residuals of the treatment and control regression models. If the two models are the same, then the difference in treatment outcome is the zero hyperplane $\mathbf{x}'\mathbf{0}$ which is included in the confidence band with probability $1-\alpha$. This specification results to a size α test of the hypotheses $H_0 : \boldsymbol{\beta}_T = \boldsymbol{\beta}_C$ against $H_a : \boldsymbol{\beta}_T \neq \boldsymbol{\beta}_C$ using the confidence band where the null is rejected if and only if $\mathbf{x}'\mathbf{0}$ is outside the band for at least one $\mathbf{x}_{(0)} \in \mathbb{R}^p$.

However, in many real-life applications, the values of the covariates do not span the entire $\mathbb{R}^p$ space. In such cases, estimating confidence bands as above is inefficient and inappropriate. In fact, Stewart (Stewart 1991b) discussed possible drawbacks of visualizing the bands over the entire $\mathbb{R}^p$. This led to the development of a more useful confidence band for the difference between two models over a restricted region of the covariates. A two-sided constant-width simultaneous confidence band for $\mathbf{x}'\boldsymbol{\beta}_T - \mathbf{x}'\boldsymbol{\beta}_C$ over the covariate region $\chi_r = \left\{\left(x_1, x_2, ..., x_p\right) : a_i \leq x_i \leq b_i,\ i = 1, 2, ..., p\right\}$ has the form

$$\mathbf{x}'\boldsymbol{\beta}_T - \mathbf{x}'\boldsymbol{\beta}_C \in \mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}, \ \ \forall \mathbf{x}_{(0)} \in \chi_r \tag{5}$$

where c is a critical constant chosen so that the simultaneous confidence level of the band is $1-\alpha$ (Liu 2011). The confidence level of this simultaneous confidence band is given by $P(S<c)$, where

$$S = \sup_{\mathbf{x}_{(0)}\in\chi_r} \frac{\left|\mathbf{x}'\left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_T - \hat{\boldsymbol{\beta}}_C + \boldsymbol{\beta}_C\right)\right|}{\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}}$$

$$= \sup_{\mathbf{x}_{(0)}\in\chi_r} \frac{\left|(\mathbf{Px})' \dfrac{\left(\mathbf{P}^{-1}\left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_T - \hat{\boldsymbol{\beta}}_C + \boldsymbol{\beta}_C\right)\right)}{\hat{\sigma}}\right|}{\sqrt{(\mathbf{Px})'(\mathbf{Px})}}$$

$$= \sup_{\mathbf{x}_{(0)}\in\chi_r} \frac{|(\mathbf{Px})'\mathbf{T}|}{\|\mathbf{Px}\|} = \sup_{V\in C(\mathbf{P},\chi_r)} \frac{|\mathbf{V}'\mathbf{T}|}{\|\mathbf{V}\|} \tag{6}$$

with $\mathbf{P}$ being the square root matrix of $\boldsymbol{\Delta}$, $\mathbf{T}\sim T_{p+1,\nu}$ and $C(\mathbf{P},\chi_r)=\{\lambda\mathbf{Px}:\lambda\geq 0,\mathbf{X}_{(0)}\}=$ $\{\lambda(\mathbf{p}_0+x_1\mathbf{p}_1+\cdots+x_p\mathbf{p}_p)\}$, $x_i\in[a_i,b_i]$ for $i=1,2,...,p$. $C(\mathbf{P},\chi_r)$ can be viewed as the cone spanned by these vectors. It is apparent that if $p\geq 1$, then the derivation of the distribution of S and essentially the critical constant c becomes non-trivial. Simulation-based methods have been presented to calculate c and have been shown to be close to the exact values under sufficiently large number of replications (Liu et al. 2005). It is further noted that the distribution of the pivotal quantity S is independent of the unknown parameters σ and $\boldsymbol{\beta}$ but is dependent on the bounds $[a_i,b_i]$ as well as the design matrix $\mathbf{X}$. Lui (2011) described the complicated relationship of these components in a general setting in details.

In this study, confidence bands based on a restricted covariate space are considered. These bands are used to determine the ATE based on the available subject profile information. Simulation-based methods to estimate c were also applied. With these bands, the behavior of the ATE is presented based on the remaining continuous covariates.

3. Proposed Exploratory Analysis

In general, the proposed method involves three major steps: creation of balanced subclasses, estimation of treatment effect for the treatment and control group in each subclass and construction of confidence bands that illustrate the behavior of ATE in each subclass. Through these steps, issues on balance and homogeneity of subclasses are addressed and a more extensive analysis of ATE is provided given specific ranges of the covariates. It must be noted, however, that this procedure is most applicable for data sets with manageable number of confounders. This proposed method is performed through the following mechanism:

(1) Create classes based on the categorical variables, forming a total of G subclasses. This step assures that the G subclasses are balanced with respect to the categorical variables. Hence, no balance checking procedure is necessary. Suppose the classes formed have sparse data, the levels of some categorical variables are pooled, if there is a natural way of doing it. If classes remain to be sparse even after pooling, these sparse classes are discarded.

(2) Given the remaining non-sparse subclasses that were formed, fit regression models and perform simultaneous regression inference.

(3) Construct confidence bands given a restricted region on subclasses with sufficiently large number of subjects. These bands will determine the magnitude of treatment effect difference between the treatment and control groups. If the zero hyperplane is contained in the band, then there is no

significant difference in treatment effect between the two treatment groups for a certain of subjects with similar categorical covariate features. On one hand, if the zero hyperplane is not included in the band, it is indicative of a significant treatment effect between the two treatment groups. The magnitude and direction of the difference will likewise be determined in these bands.

Through this procedure, one is able to detect for which group of subjects will the treatment effect difference be significant (either in the positive or negative direction).

For continuous covariates with a narrow range for $[a_i, b_i]$, one may investigate the subject profile by looking into the behavior of each independent variable considered in the confidence band while holding the other variables fixed. This will aid in understanding the behavior of treatment difference with respect to a single, independent, continuous variable under fixed values of the remaining predictors.

4. Data Analysis

To illustrate the proposed procedure, we apply the exploratory method on two data sets that aim to establish ATE of an intervention over some control.

A. National Supported Work (NSW) Data

The National Supported Work (NSW) Data is from a 1999 seminal study on the comparison of treatment and control groups to determine causal effects of a job training program (Lalonde 1999). This data combines the treated units from a randomized evaluation of the NSW demonstration with the control units drawn from survey data. The outcome of interest is RE78 (real earnings in 1978) with the treatment defined as the participation in the NSW job training program. The continuous covariates considered in the study are age and education quantified by the number of years in school while the categorical variables are Black (1 if black, 0 otherwise), Hispanic (1 if Hispanic, 0 otherwise), married (1 if married, 0 otherwise) and no degree (1 if no degree, 0 otherwise). The total number of observations used in the analysis is $n = 445$, $n_C = 260$ are under the control group while the remaining $n_T = 185$ observations are under the treatment group.

At the initial stage, subgroups are created based on the four categorical variables above. With each variable having two levels, $G = 2^4 = 16$ total subclasses that are balanced with respect to the categorical variables are constructed. The frequencies of the subclasses are as follows:

Table 1 Frequency of control and treatment units in balanced subclasses formed

Subclass	Characteristic	Control	Treat	Total
1	Black = 0, Hispanic = 0, Married = 0, No Degree = 0	6	7	13
2	Black = 1, Hispanic = 0, Married = 0, No Degree = 0	26	37	63
3	Black = 0, Hispanic = 1, Married = 0, No Degree = 0	1	1	2
4	Black = 1, Hispanic = 1, Married = 0, No Degree = 0	0	0	0
5	Black = 0, Hispanic = 0, Married = 1, No Degree = 0	0	2	2
6	Black = 1, Hispanic = 0, Married = 1, No Degree = 0	9	6	15
7	Black = 0, Hispanic = 1, Married = 1, No Degree = 0	1	1	2
8	Black = 1, Hispanic = 1, Married = 1, No Degree = 0	0	0	0
9	Black = 0, Hispanic = 0, Married = 0, No Degree = 1	10	8	18
10	Black = 1, Hispanic = 0, Married = 0, No Degree = 1	154	90	244
11	Black = 0, Hispanic = 1, Married = 0, No Degree = 1	23	7	30
12	Black = 1, Hispanic = 1, Married = 0, No Degree = 1	0	0	0
13	Black = 0, Hispanic = 0, Married = 1, No Degree = 1	1	1	2

Table 1 (Continued)

Subclass	Characteristic	Control	Treat	Total
14	Black = 1, Hispanic = 0, Married = 1, No Degree = 1	26	23	49
15	Black = 0, Hispanic = 1, Married = 1, No Degree = 1	3	2	5
16	Black = 1, Hispanic = 1, Married = 1, No Degree = 1	0	0	0

Since Subclasses 3, 4, 5, 7, 8, 12, 13, 15 and 16 are sparse groups, 13 observations are discarded from the analysis. This results to a total of $n^* = 432$ experimental units with $n_C^* = 254$ control units and $n_T^* = 178$ treatment units. For the seven remaining subgroups, confidence bands, illustrating the magnitude in the earning difference between subjects who participated in the NSW job training program and those who did not, were generated using the continuous predictors age and number of years in school. The NSW dataset is best modeled with these main effects and a quadratic term on age (Lalonde 1999). This same relationship is used in modeling the real earnings of the treatment and control groups for each of the 7 classes formed. The results of the regression modeling procedure are shown in Table 2. Information on the restricted covariate region of interest of the independent variables age x_1 and number of years in school x_2 are also presented in the same table.

Table 2 Regression estimates of NSW data

Subclass	Regression Model Estimates	$\hat{\beta}$ Estimates	Covariate Region
1	$\hat{y}_T = -59637 + 3622X_1 + 942.2X_2 - 54.03X_1^2$	$\hat{\beta} = (486420, -44753, 9028.2, 826.9)$	$x_1 : [20, 41]$
	$\hat{y}_C = -395310 + 40276X_1 - 10393X_2 - 757.48X_1^2$		$x_2 : [12, 14]$
2	$\hat{y}_T = -59637 + 3622X_1 + 942.2X_2 - 54.03X_1^2$	$\hat{\beta} = (-46372, 2612, 664.0, -36.2)$	$x_1 : [18, 46]$
	$\hat{y}_C = -13265 + 1010X_1 + 278.2X_2 - 17.84X_1^2$		$x_2 : [12, 16]$
6	$\hat{y}_T = -339270 + 11101X_1 + 15105X_2 - 183.5X_1^2$	$\hat{\beta} = (-237260, 4429.7, 14966, -79.7)$	$x_1 : [23, 42]$
	$\hat{y}_C = -102010 + 6671X_1 + 139.4X_2 - 103.82X_1^2$		$x_2 : [12, 14]$
9	$\hat{y}_T = -25544 + 1287X_1 + 1749.5X_2 - 22.38X_1^2$	$\hat{\beta} = (-21203, -810.1, 3434.4, 16.92)$	$x_1 : [17, 38]$
	$\hat{y}_C = -4341 + 2097X_1 + 1683.9X_2 - 39.29X_1^2$		$x_2 : [7, 11]$
10	$\hat{y}_T = -2308 + 322.4X_1 + 269.8X_2 - 5.06X_1^2$	$\hat{\beta} = (-5807, 490.3, 25.01, -8.74)$	$x_1 : [17, 55]$
	$\hat{y}_C = 3499 - 167.9X_1 + 244.7X_2 - 3.67X_1^2$		$x_2 : [3, 11]$
11	$\hat{y}_T = 141450 - 10180X_1 - 3471.8X_2 + 238.7X_1^2$	$\hat{\beta} = (163990, -11704, -4620.5, 256.7)$	$x_1 : [17, 50]$
	$\hat{y}_C = -25538 + 1524X_1 + 1148.7X_2 - 27.01X_1^2$		$x_2 : [4, 11]$
14	$\hat{y}_T = -4526 + 952X_1 - 210.7X_2 - 15.7X_1^2$	$\hat{\beta} = (-65991, 4812, -97.01, -79.4)$	$x_1 : [19, 46]$
	$\hat{y}_C = 61464 - 3860X_1 - 113.7X_2 + 63.6X_1^2$		$x_2 : [4, 11]$

Upon generation of the regression model estimates for both groups, the critical constant c was simulated using the proposed procedure of Lui et al. (1995). It is noted that the values of $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ and $\boldsymbol{\Delta}$ are based on the design matrix. These components were calculated using MATLAB. The same program was used in generating the graphs of the confidence bands for each class.

Based on the figures of the simultaneous confidence bands, shown in Figure 1 to Figure 7, it can be observed that the zero hyperplane is included in the band which implies that there is no significant treatment difference between the two groups for all subclasses except Subclass 1. For Subclass 1, the treatment difference between the two groups is shown to be negative in some portion of the surface of x_1 and x_2.

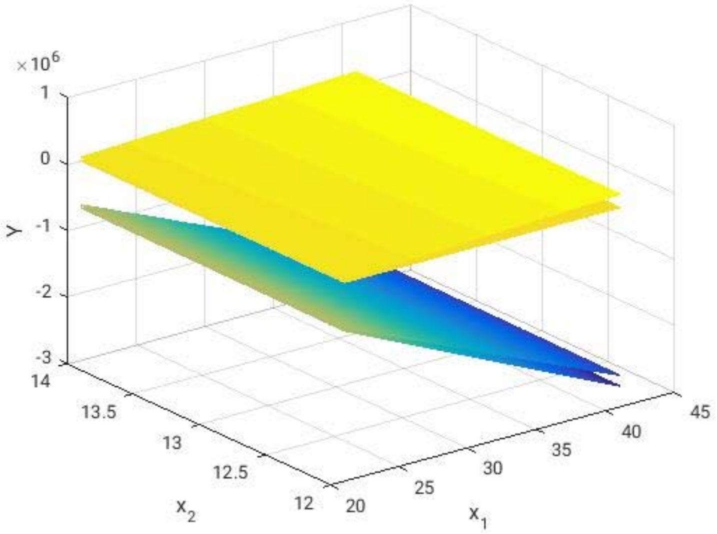

Figure 1 Simultaneous confidence band of ATE for Subclass 1

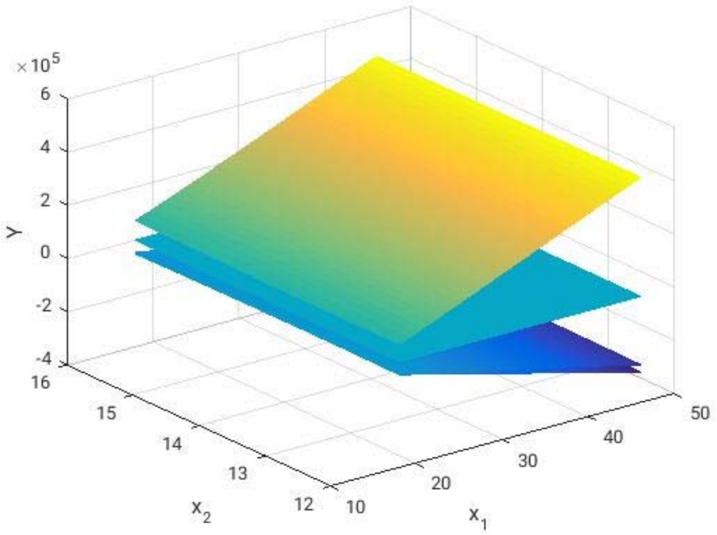

Figure 2 Simultaneous confidence band of ATE for Subclass 2

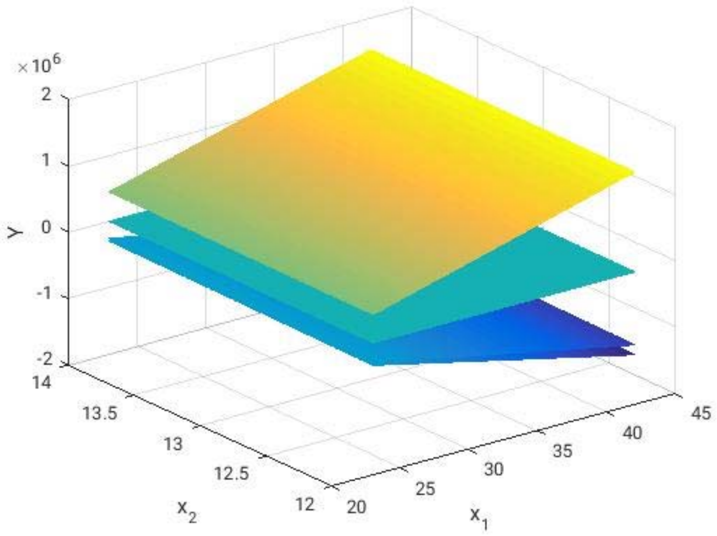

Figure 3 Simultaneous confidence band of ATE for Subclass 6

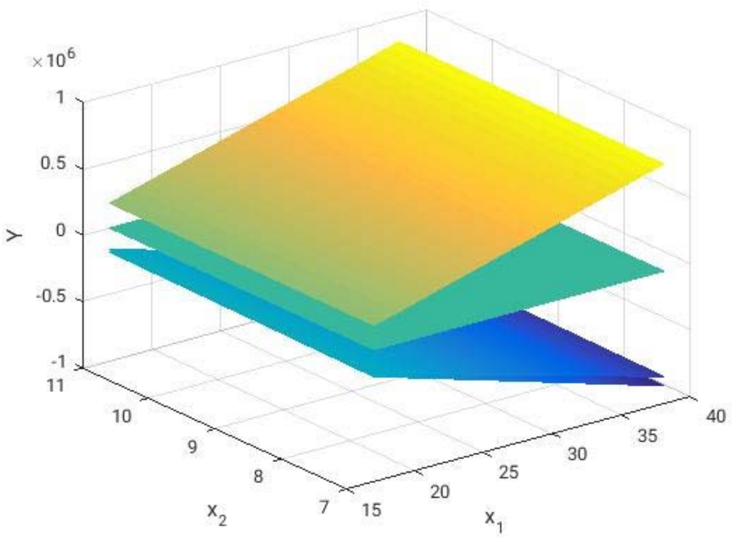

Figure 4 Simultaneous confidence band of ATE for Subclass 9

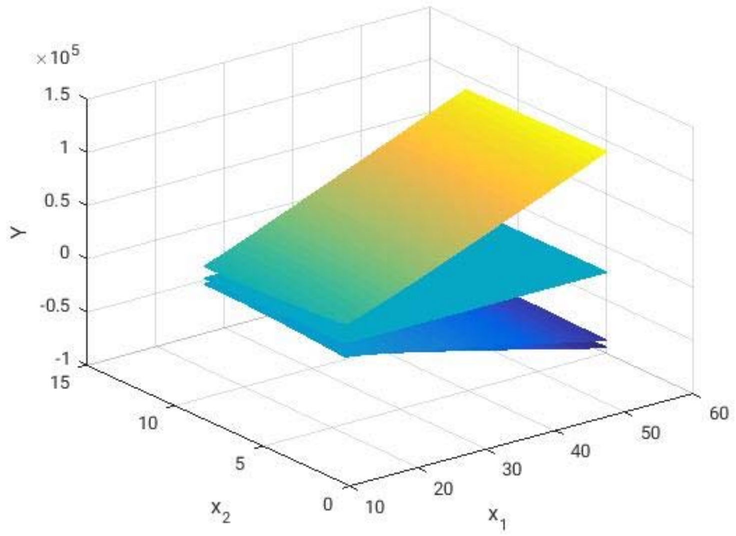

Figure 5 Simultaneous confidence band of ATE for subclass 10

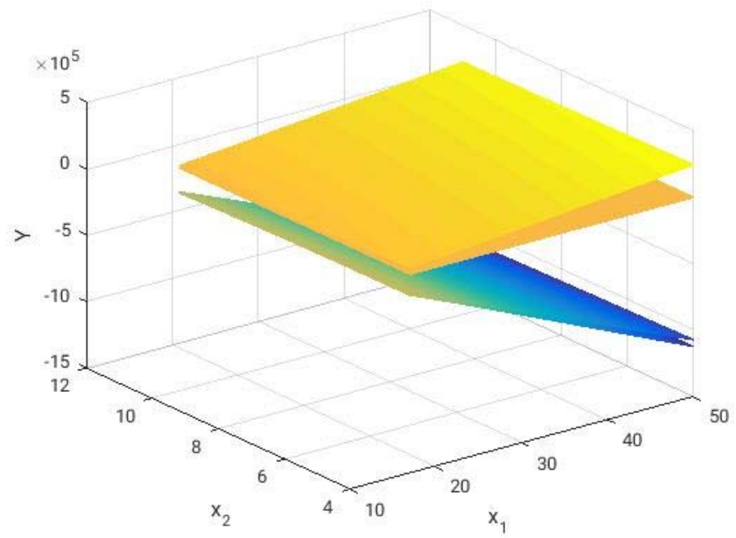

Figure 6 Simultaneous confidence band of ATE for Subclass 11

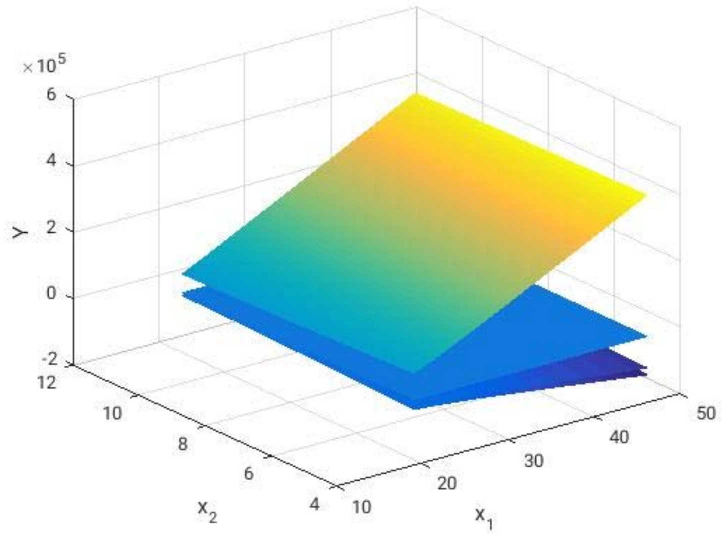

Figure 7 Simultaneous confidence band of ATE for Subclass 14

Table 3 shows the p-values of the simultaneous inference in regression, which support the graphs of the confidence band as evident by the p-values greater than $\alpha = 0.05$ except for Subclass 1. It can be deduced that for Subclass 1, the zero hyperplane $\mathbf{x}'\mathbf{0}$ is outside the band for at least one $\mathbf{x}_{(0)} \in \chi_r$. Graphically, it can be observed that this conclusion occurs on lower values of age and number of years in school. This shows that those who have attended the NSW job training still tended to have lower earnings than those who did not avail of the training for the younger respondents who have not attended much years in school. For the remaining subgroups, the p-value suggests that $\mathbf{x}_{(0)}$ is contained in all possible $\mathbf{x}_{(0)}$ in the restricted region.

Table 3 Simultaneous inference p-values on subclasses of NSW data

Subgroup	p-value	Conclusion
1	0.0308	Reject H_0
2	0.1568	Do not reject H_0
6	0.1057	Do not reject H_0
9	0.6005	Do not reject H_0
10	0.8425	Do not reject H_0
11	0.0658	Do not reject H_0
14	0.1163	Do not reject H_0

This behavior is compared to the classical z-test for testing means of two independent populations. The mean difference of Subclass 1 is shown to be negative, as likewise observed on the confidence band. However, the confidence band suggests that the difference in regression model fit between the two groups is significantly different. This is not reflected on the z-test carried out for Subclass 1 (p-value = 0.3795). For the remaining subclasses, consistent results are observed as depicted by the non-rejection of the null hypothesis for the z-test and the inclusion of the zero hyperplane in the confidence bands.

Given that the bounds for the number of years of education is narrow, one may be able to explore the ATE by looking into the behavior of the confidence bands as a function of one variable in each subclass. By finding the solutions of the confidence bands under a fixed covariate x_j, one is able to determine the value of x_i, $i \neq j$, for which the direction of the treatment effect changes. The choice of the independent variable to be held fixed is based on the covariate region, with the narrowest region. In the NSW data set, an ATE function in terms of age (x_1) is generated by holding the number of education (x_2) fixed because the x_2 -region is narrower than the x_1 -region. For a fixed covariate, the ideal scenario is for the confidence bands to illustrate a positive treatment effect over the entire covariate region of interest. This implies that the training will increase the salary of all subjects across all age and number of years of education. Although, it is noted that this result may not necessarily hold true for all subclasses. In some instances, varying directions may occur and thus, the fixed value for which the function with a positive treatment effect is not clearly identified. While this occurrence may be true about this exploratory analysis, it will, on one hand, provide a good idea of the subject profile for which the treatment is effective or not.

Tables 4-10 reflect the forms of confidence bands for fixed values of x_2, the critical constant c and regression estimates while Figures 8-14 show the corresponding quadratic function of x_1 given a constant x_2. Specifically, the minimum, mean and maximum values of x_2 were considered as fixed points.

For Subclass 1, it can be observed that given the maximum possible value of x_2, a positive treatment effect is guaranteed across all possible values of x_1, as shown in Table 4 and Figure 8. However, when the minimum and mean values are considered to be the fixed points, the ATE is either positive or negative on some interval of x_1. In some subclass, imaginary solutions are calculated which implies that for some values of x_1 under a fixed value of x_2, no treatment effect is observed. This suggests that a respondent with a subject profile defined by Subclass 1s (non-black, non-Hispanic, single and has an educational degree with 14 years spent in school, will most likely have positive earnings given the job training regardless of their age. This relationship is more clearly illustrated in Figure 8.

Table 4 Confidence band for Subclass 1 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 12$	$826x_1^2 - 44753x_1 + 594760 \pm 4.37(2.78)\sqrt{\frac{33x_1^4 - 3000x_1^3 + 152400x_1^2 - 2760000x_1 + 18850000}{5000}}$	$x_{1+} = 23.39, 30.93$ $x_{1-} = 23.46, 30.68$
$x_2 = 12.23$	$826x_1^2 - 44753x_1 + 596836.44 \pm 4.37(2.78)\sqrt{\frac{33x_1^4 - 3000x_1^3 + 152446x_1^2 - 2762300x_1 + 18878129}{5000}}$	$x_{1+} = 23.82, 30.31$ $x_{1-} = 23.68, 30.59$
$x_2 = 14$	$826x_1^2 - 44753x_1 + 612816 \pm 4.37(2.78)\sqrt{\frac{33x_1^4 - 3000x_1^3 + 152800x_1^2 - 2780000x_1 + 19130000}{5000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$

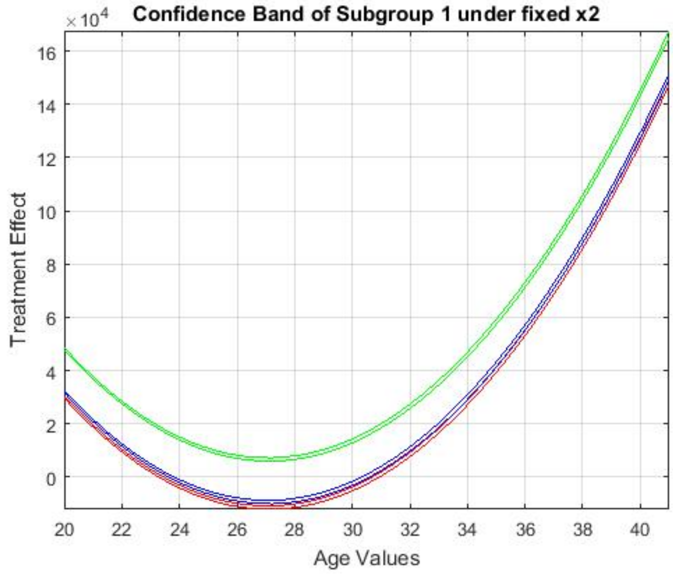

Figure 8 Simultaneous confidence band of Y for Subclass 1 on fixed minimum, mean and maximum values of x_2

For Subgroup 2, it can be observed that across all the possible range of x_2, the ATE can be non-negative or negative. This implies that there is no guarantee for which ranges of x_1 and x_2 is the treatment effective in increasing earnings, on black, non-Hispanic, single respondents with an educational degree. These results are reflected in the solutions of the confidence bands fixed at x_2, as shown in Table 4. However, Figure 9 demonstrates that the behavior of the ATE on the restricted region of x_1 and fixed values of x_2 considered tend more towards the positive direction. Hence, most of the black, non-Hispanic, single respondents regardless of degree status will gain positive effects from the job training.

Table 5 Confidence band for Subclass 2 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 12$	$-36x_1^2 + 2612x_1 + 38404 \pm 3.12(2.58)\sqrt{\frac{39x_1^4 - 4000x_1^3 + 164600x_1^2 - 2980000x_1 + 19820000}{500000}}$	$x_{1+} = 20.48, 52.07$ $x_{1-} = 20.50, 52.04$
$x_2 = 12.37$	$-36x_1^2 + 2612x_1 + 38158.32 \pm 3.12(2.58)\sqrt{\frac{78x_1^4 - 8000x_1^3 + 329792x_1^2 - 5989600x_1 + 40053549}{1000000}}$	$x_{1+} = 20.27, 52.34$ $x_{1-} = 20.28, 52.22$
$x_2 = 16$	$-36x_1^2 + 2612x_1 + 35748 \pm 3.12(2.58)\sqrt{\frac{39x_1^4 - 4000x_1^3 + 167800x_1^2 - 3140000x_1 + 23580000}{500000}}$	$x_{1+} = 18.29, 54.28$ $x_{1-} = 20.28, 52.22$

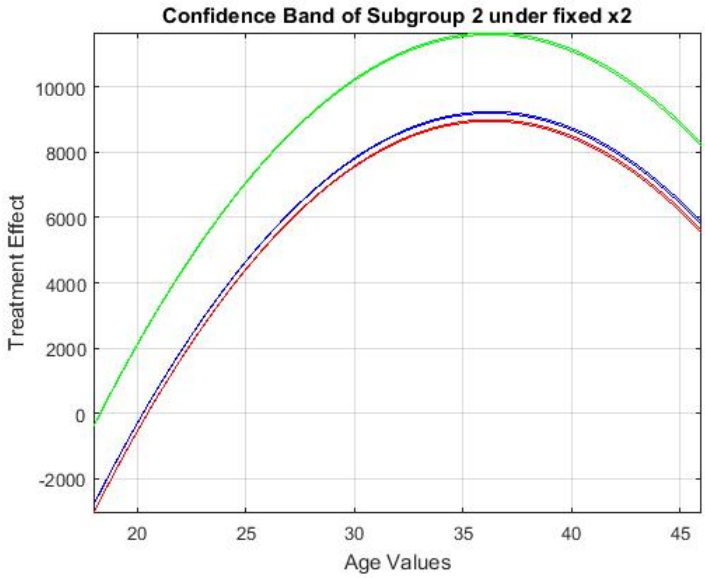

Figure 9 Simultaneous confidence band of Y for Subclass 2 on fixed minimum, mean and maximum values of x_2

Meanwhile, for Subclass 6, there is a guaranteed behavior realized when x_2 is fixed at the minimum, according to the results shown in Table 6. Figure 10 displays this relationship as well. The graph suggests that for black, non-Hispanic, married respondents with a degree, the ATE is negative over all possible domain of x_1 when x_2 is at the minimum. On one hand, the behavior of the ATE is always positive when x_2 is held fixed at the maximum for all possible values of x_1. This change in behavior may be an indicator of the effect of the number of years in school on the direction of the treatment effect for this group of respondents. It shows that positive earnings are achieved after the training for respondents who have stayed for 14 years in school while an opposite behavior is observed for those who have stayed for 12 years.

Table 6 Confidence band for Subclass 6 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 12$	$-79x_1^2 + 4430x_1 + 57665 \pm 3.96(3.29)\sqrt{\frac{3x_1^4 - 375x_1^3 + 18595x_1^2 - 3980000x_1 + 3132500}{3125}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 12.47$	$-79x_1^2 + 4430x_1 + 50630.98 \pm 3.96(3.29)\sqrt{\frac{960x_1^4 - 120000x_1^3 + 5959424x_1^2 - 127961600x_1 + 1012176329}{1000000}}$	$x_{1+} = 15.70, 40.42$ $x_{1-} = 16.04, 39.99$
$x_2 = 14$	$-79x_1^2 + 4430x_1 + 27733 \pm 3.96(3.29)\sqrt{\frac{3x_1^4 - 375x_1^3 + 18715x_1^2 - 406000x_1 + 3270250}{3125}}$	$x_{1+} = 7.10, 48.97$ $x_{1-} = 7.27, 48.82$

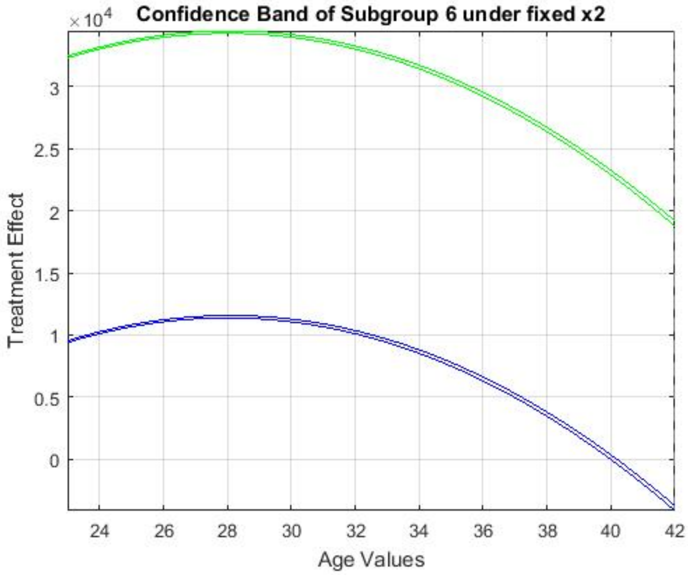

Figure 10 Simultaneous confidence band of Y for Subclass 6 on fixed minimum, mean and maximum values of x_2

On the other hand, for the confidence band of Subclass 9, which is composed of subjects who are non-black, non-Hispanic, single and non-degree holder, a positive treatment effect is guaranteed at a fixed, maximum value of x_2. This is apparent in the solutions of its corresponding band reflected in Table 7. This behavior is also true for all values of x_1 given the fixed mean value of x_2. At the minimum, however, it can be established that the training has a negative effect when the number of years in school is fixed at 7 years. This means that subjects who did not spend much time in school and did not attend the training tend to have higher earnings; thus, the training is not beneficial for this group of subjects. This behavior is illustrated also in Figure 11.

Table 7 Confidence band for Subclass 9 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 7$	$17x_1^2 - 811x_1 + 2843 \pm 3.67(1.67)\sqrt{\dfrac{9x_1^4 - 800x_1^3 + 28034x_1^2 - 418200x_1 + 2301800}{10000}}$	$x_{1+} = 3.98, 44.12$ $x_{1-} = 4.29, 43.85$
$x_2 = 9.67$	$17x_1^2 - 811x_1 + 12014.45 \pm 3.67(1.67)\sqrt{\dfrac{45x_1^4 - 40000x_1^3 + 1409977x_1^2 - 21257100x_1 + 118143412}{500000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 11$	$17x_1^2 - 811x_1 + 16583 \pm 3.67(1.67)\sqrt{\dfrac{9x_1^4 - 800x_1^3 + 28282x_1^2 - 428600x_1 + 2401800}{10000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$

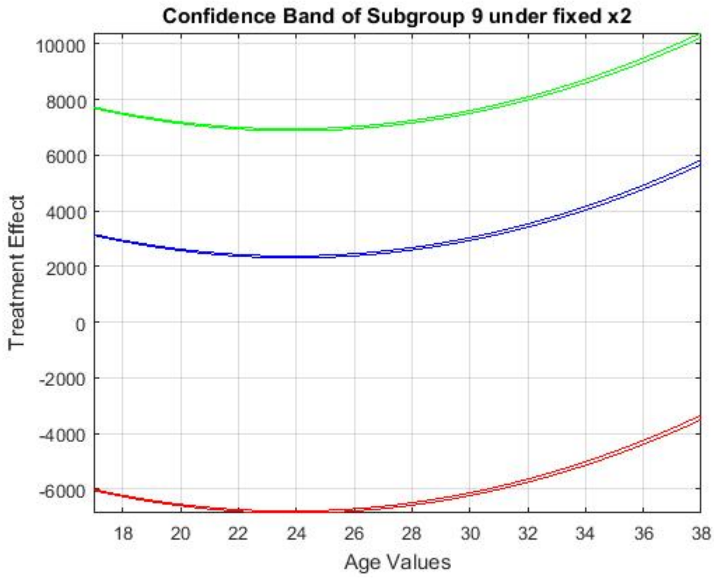

Figure 11 Simultaneous confidence band of Y for Subclass 9 on fixed minimum, mean and maximum values of x_2

For Subgroup 10, it is shown that the solutions of the confidence bands for any of the fixed values of x_2 considered is around the lower extremes of x_1, as reflected in Table 8. This suggests that the change of behavior in the effectiveness of the treatment is reflected more among younger subjects. Figure 12 shows that the ATE is dominantly positive, at least for values of x_1 for which is it defined; hence, the treatment is effective. Based on the graph, we note that observations could be made only on some values of x_1 as the confidence bands under any of the fixed values of x_2 are not defined in the entire domain of x_1.

Table 8 Confidence band for Subclass 10 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 3$	$-9x_1^2 - 490x_1 + 5733 \pm 3.06(1.17)\sqrt{\frac{3x_1^4 - 400x_1^3 + 16980x_1^2 - 309900x_1 + 2430200}{1000000}}$	$x_{1+} = 17.01$ $x_{1-} = 17.03$
$x_2 = 9.61$	$-9x_1^2 - 490x_1 + 5533 \pm 3.06(1.17)\sqrt{\frac{3x_1^4 - 400x_1^3 + 17460x_1^2 - 335500x_1 + 2503800}{1000000}}$	$x_{1+} = 16.15$ $x_{1-} = 16.16$
$x_2 = 11$	$-9x_1^2 - 490x_1 + 5533 \pm 3.07(1.17)\sqrt{\frac{3x_1^4 - 400x_1^3 + 17460x_1^2 - 335500x_1 + 2503800}{1000000}}$	$x_{1+} = 15.99$ $x_{1-} = 15.99$

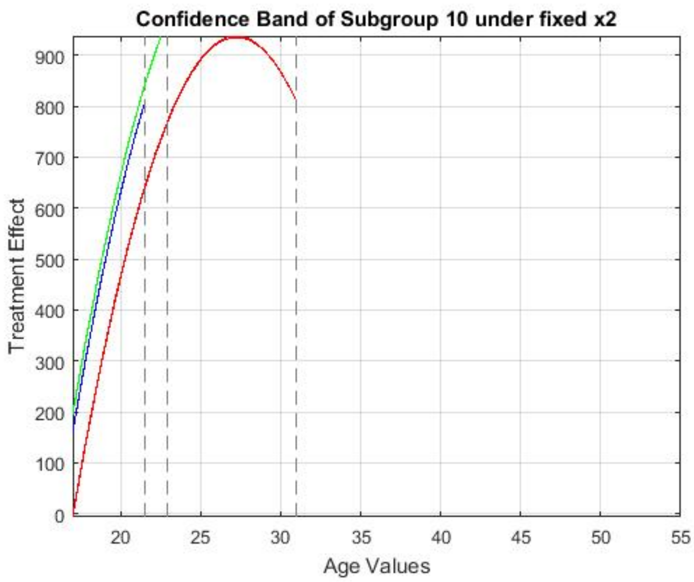

Figure 12 Simultaneous confidence band of Y for Subclass 10 on fixed minimum, mean and maximum values of x_2

Looking at subgroup 11, real solutions are calculated within the range of x_1 for the mean and maximum. This indicates that there is a change in direction of ATE for some values of x_1; hence, the effect of the training on subjects that are non-black, Hispanic, single and no degree cannot be easily determined when they have spent 9.61 and 11 years in school. It is interesting to note, on one hand, that the training yields positive effect for subjects who share the same categorical profile but has spent only 3 years in school. These observations are shown in Table 9 and Figure 13.

Table 9 Confidence band for Subclass 11 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 4$	$266x_1^2 - 11704x_1 + 145505 \pm 3.34(2.76)\sqrt{\dfrac{21x_1^4 - 2000x_1^3 + 74900x_1^2 - 1180000x_1 + 7960000}{500000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 8.9$	$266x_1^2 - 11704x_1 + 122862.1 \pm 3.34(2.76)\sqrt{\dfrac{21x_1^4 - 2000x_1^3 + 74165x_1^2 - 1155500x_1 + 7422225}{50000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 11$	$266x_1^2 - 11704x_1 + 113158 \pm 3.34(2.76)\sqrt{\dfrac{21x_1^4 - 2000x_1^3 + 73850x_1^2 - 1145000x_1 + 7522500}{50000}}$	$x_{1+} = 15.22, 29.11$ $x_{1-} = 15.56, 29.05$

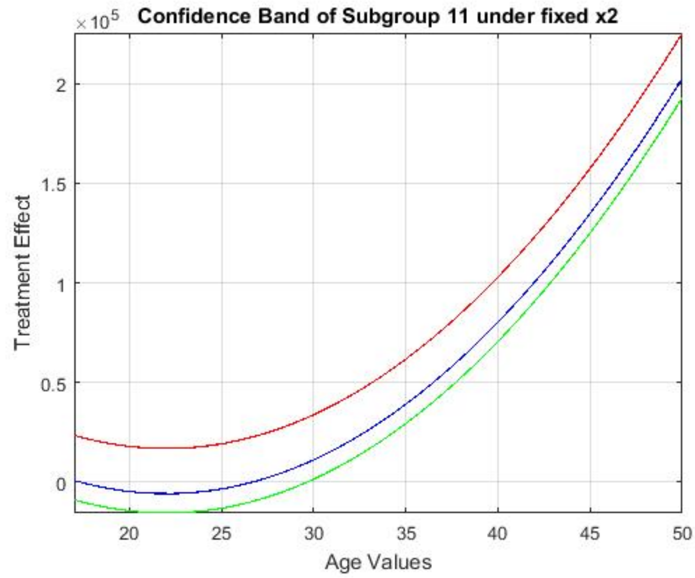

Figure 13 Simultaneous confidence band of Y for Subclass 11 on fixed minimum, mean and maximum values of x_2

For the last class, Subgroup 14, the solutions of the confidence bands on a fixed x_2 at the minimum, mean and maximum suggest that there is no guaranteed ATE behavior detected as shown in Table 10 because the solutions fall in the range of x_1. Similar to other subclasses, this is indicative of a change in treatment effect within the domain; thus, no guaranteed observation on the behavior of ATE may be established. Although, Figure 14 suggests that majority of the ATE behavior realized is towards the negative direction. It shows that respondents who are black, Hispanic, married and has no degree tend to have a negative treatment effect, at least for those with ages that are at the extremes and regardless of the number of years in school since this relationship holds true for all three fixed points of x_2 considered.

Table 10 Confidence band for Subclass 14 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 4$	$-80x_1^2 - 4812x_1 + 66374 \pm 3.19(2.78)\sqrt{\dfrac{x_1^4 - 140x_1^3 + 6859x_1^2 - 119000x_1 + 1025000}{25000}}$	$x_{1+} = 21.42, 38.74$ $x_{1-} = 21.44, 38.71$
$x_2 = 9.9$	$-80x_1^2 - 4812x_1 + 66940.4 \pm 3.19(2.78)\sqrt{\dfrac{2x_1^4 - 280x_1^3 + 13995x_1^2 - 255700x_1 + 2180685}{50000}}$	$x_{1+} = 21.81, 38.36$ $x_{1-} = 21.88, 28.26$
$x_2 = 11$	$-80x_1^2 - 4812x_1 + 67046 \pm 3.19(2.78)\sqrt{\dfrac{x_1^4 - 140x_1^3 + 7025x_1^2 - 129500x_1 + 1116000}{25000}}$	$x_{1+} = 21.92, 38.23$ $x_{1-} = 21.94, 38.21$

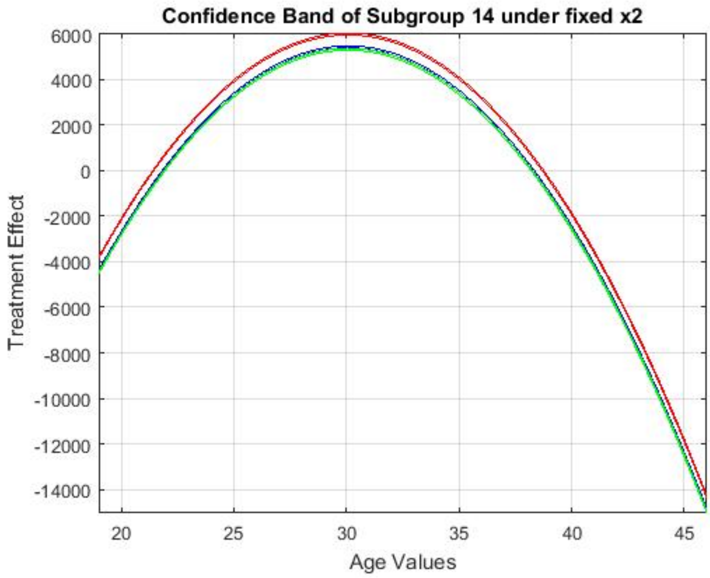

Figure 14 Simultaneous confidence band of Y for Subclass 14 on fixed minimum, mean and maximum values of x_2

Although there is no unifying behavior of the average treatment effect across the various classes on a specified covariate region, these analyses remain to be helpful in identifying the behavior of the treatment effect given certain information on the covariates. In this specific example, the expected treatment effect of a subject is identifiable based on information of their gender, marital status, ethnicity, age and number of years in school.

B. *Lilly* Data Set

We also applied the proposed data analysis to estimate the ATE of a randomized, open-label clinical trial performed by *Eli Lilly* (SAS Clinical Trial Data Portal 2018). The primary objective of the study was to demonstrate that *dulaglutide* (treatment A) was non-inferior to *liraglutide* (treatment B) in controlling glycosylated hemoglobin A1C in patients with type 2 diabetes. The outcome of interest is the hemoglobin levels of patients and the treatments are *dulaglutide* and *liraglutide*. The covariates are age, BMI, ethnicity (Hispanic/Latino or otherwise), region of residence (North America, South America and Europe), and gender (Male and Female). The total number of observations is $n = 551$, with $n_A = 275$ subjects treated with *dulaglutide* and $n_B = 276$ subjects with *luraglutide*. Subclasses are created based on the three categorical variables, with two levels for gender and ethnicity and three levels for region, resulting to $G = 12$ subclasses. The frequencies of the subclasses are shown in Table 11.

Table 11 Frequency of *Dulaglutide* and *Luraglutide* units in balanced subclasses formed

Subclass	Characteristic	*Dulaglutide* (Trt A)	*Luraglutide* (Trt B)	Total
1	Region = North America, Gender = Female, Ethnicity = Hispanic	22	21	43
2	Region = South America, Gender = Female, Ethnicity = Hispanic	15	16	31
3	Region = Europe, Gender = Female, Ethnicity = Hispanic	2	1	3
4	Region = North America, Gender = Male, Ethnicity = Hispanic	20	19	39
5	Region = South America, Gender = Male, Ethnicity = Hispanic	5	7	12
6	Region = Europe, Gender = Male, Ethnicity = Hispanic	4	2	6
7	Region = North America, Gender = Female, Ethnicity = Not Hispanic	29	24	53
8	Region = South America, Gender = Female, Ethnicity = Not Hispanic	0	0	0
9	Region = Europe, Gender = Female, Ethnicity = Not Hispanic	82	77	159
10	Region = North America, Gender = Male, Ethnicity = Not Hispanic	17	23	40
11	Region = South America, Gender = Male, Ethnicity = Not Hispanic	0	0	0
12	Region = Europe, Gender = Male, Ethnicity = Not Hispanic	79	86	165
	TOTAL	275	276	551

Since Subclasses 3, 6, 8, and 11 are sparse classes, 9 observations from these groups are discarded for further analysis. This results to a total of $n^* = 542$ subjects $n_A^* = 269$ units for treatment A and $n_B^* = 273$ for treatment B. Consequenlty, 8 non-sparse subclasses are formed using the categorical variables. The magnitudes of difference in hemoglobin A1C levels between *Dulaglutide* and *Liraglutide* are illustrated using confidence bands. As an initial step, regression estimates are generated. A linear regression model was used to estimate the mean hemoglobin A1C level for each of the treatment groups. The results are shown in Table 12. Also, the covariate regions of the independent variables, age (x_1) and body mass index (BMI) (x_2), are reflected in the same table.

Table 12 Regression estimates of *Lilly* data

Subclass	Regression Model Estimates	$\hat{\beta}$ Estimates	Covariate Region
1	$\hat{y}_T = 5.3096 - 0.0005X_1 + 0.0601X_2$ $\hat{y}_C = 6.8927 + 0.0045X_1 + 0.0087X_2$	$\ddot{\beta} = (-1.5830, -0.0050, 0.0514)$	$x_1 : [28, 72]$ $x_2 : [19.8, 44.2]$
2	$\hat{y}_T = 3.3269 + 0.0016X_1 + 0.0078X_2$ $\hat{y}_C = 1.9487 - 0.0102X_1 + 0.1578X_2$	$\ddot{\beta} = (1.3782, 0.0218, -0.0800)$	$x_1 : [23, 67]$ $x_2 : [26.3, 41.7]$
4	$\hat{y}_T = 5.7410 + 0.0361X_1 - 0.0207X_2$ $\hat{y}_C = 3.5169 + 0.0149X_1 + 0.0995X_2$	$\ddot{\beta} = (2.2240, 0.0212, -0.1202)$	$x_1 : [35, 74]$ $x_2 : [20.8, 44.3]$
5	$\hat{y}_T = 9.1803 - 0.0807X_1 + 0.0487X_2$ $\hat{y}_C = 3.2242 - 0.0030X_1 + 0.1292X_2$	$\ddot{\beta} = (5.9561, -0.0776, -0.0806)$	$x_1 : [41, 67]$ $x_2 : [24.5, 34.5]$
7	$\hat{y}_T = 6.0661 - 0.0132X_1 + 0.0349X_2$ $\hat{y}_C = 6.6703 - 0.0049X_1 - 0.0023X_2$	$\ddot{\beta} = (-0.6042, -0.0083, 0.0372)$	$x_1 : [33, 75]$ $x_2 : [24.6, 47.2]$
9	$\hat{y}_T = 5.0199 + 0.0101X_1 + 0.0270X_2$ $\hat{y}_C = 6.8766 + 0.0127X_1 - 0.0300X_2$	$\ddot{\beta} = (-1.8567, -0.0026, 0.0569)$	$x_1 : [35, 77]$ $x_2 : [21.8, 44.8]$
10	$\hat{y}_T = 7.8321 - 0.0050X_1 - 0.0342X_2$ $\hat{y}_C = 4.5625 + 0.0238X_1 + 0.0142X_2$	$\ddot{\beta} = (3.2696, -0.0288, -0.0484)$	$x_1 : [38, 74]$ $x_2 : [24.1, 43.6]$
12	$\hat{y}_T = 6.8136 - 0.0106X_1 + 0.0094X_2$ $\hat{y}_C = 9.0976 - 0.0198X_1 - 0.0363X_2$	$\ddot{\beta} = (-2.2840, 0.0092, 0.0454)$	$x_1 : [35, 73]$ $x_2 : [21.4, 45.7]$

Based on the linear regression estimates in Table 12, confidence bands are created for the specified covariate regions in each subclass. To do so, the critical constant c is simulated and the data-based components $\hat{\sigma}$ and $\mathbf{\Delta}$ are calculated. Figures 15-22 show the generated confidence bands. Looking at the surfaces of each subgroup, we observe that the zero hyperplane is inside the band for all $\mathbf{x}_{(0)}$ in the covariate regions for all 8 subgroups. This suggests that the average treatment effect is insignificant for all groups.

The conclusions drawn from the confidence bands can be further established using the p-value of the simultaneous inference in regression models, shown in Table 13. For all subclasses considered, the p-values are significantly greater than $\alpha = 0.05$ which lead to the non-rejection of the null hypothesis that $\boldsymbol{\beta}_A - \boldsymbol{\beta}_B = 0$. This establishes further the insignificant ATE between the two drugs; thus, *Dulaglutide* is non-inferior to *Liraglutide* in stabilizing hemoglobin A1C levels among type II diabetes patients.

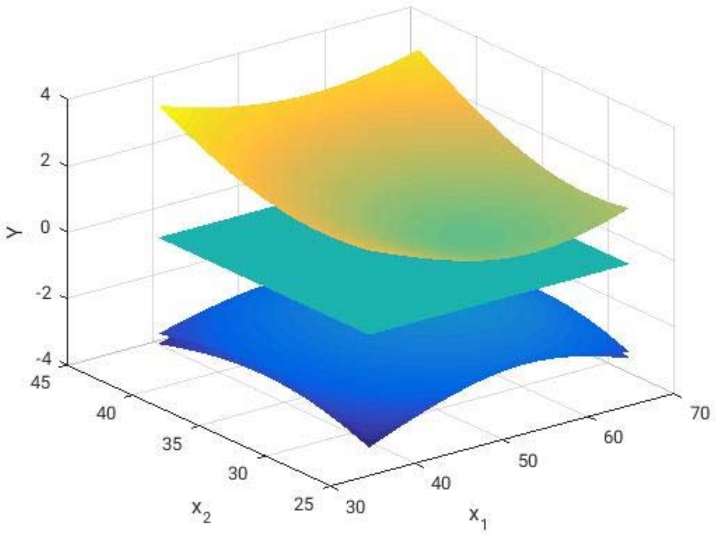

Figure 15 Simultaneous confidence band of ATE for Subclass 1

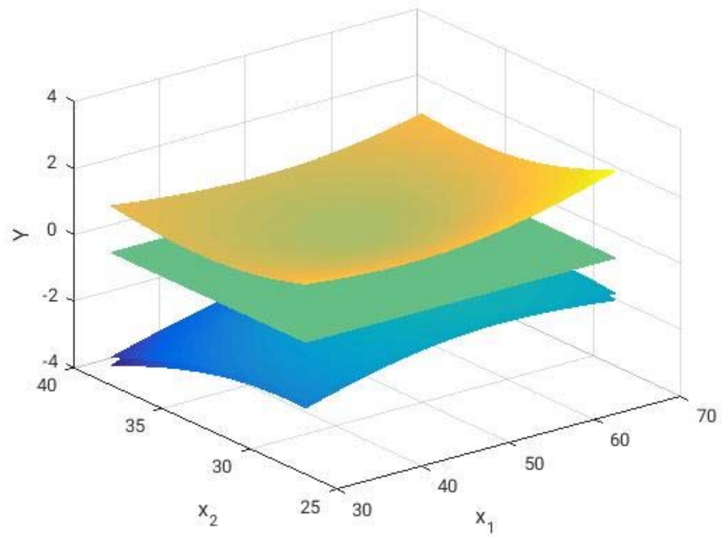

Figure 16 Simultaneous confidence band of ATE for Subclass 2

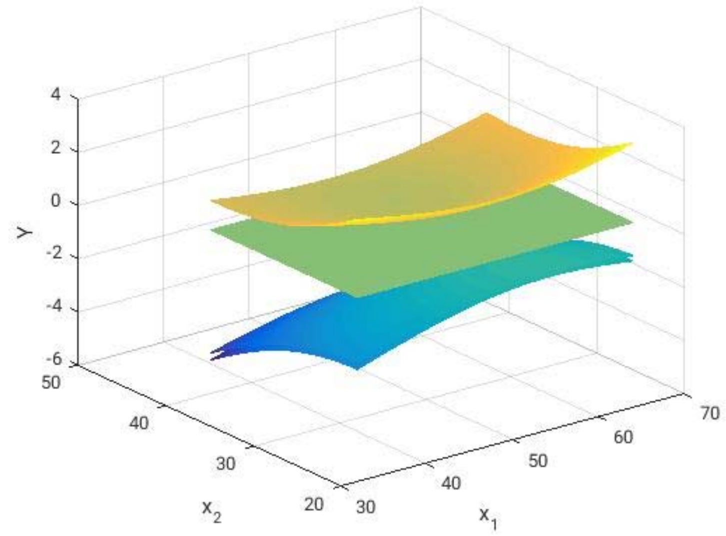

Figure 17 Simultaneous confidence band of ATE for Subclass 4

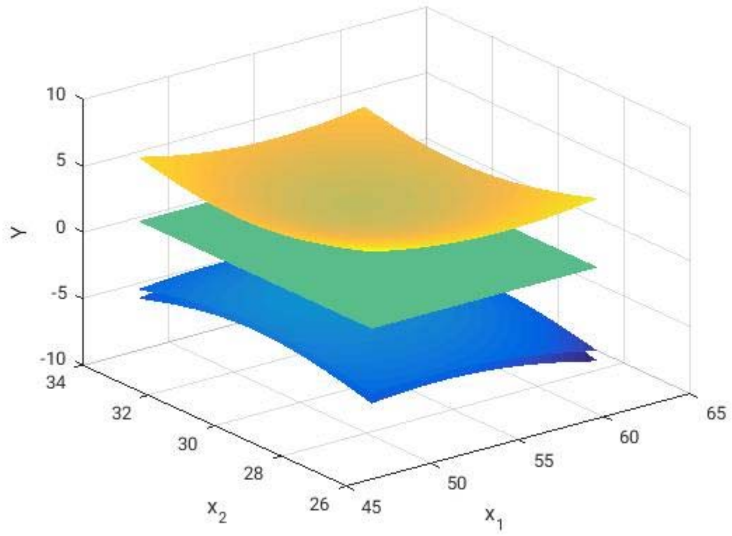

Figure 18 Simultaneous confidence band of ATE for Subclass 5

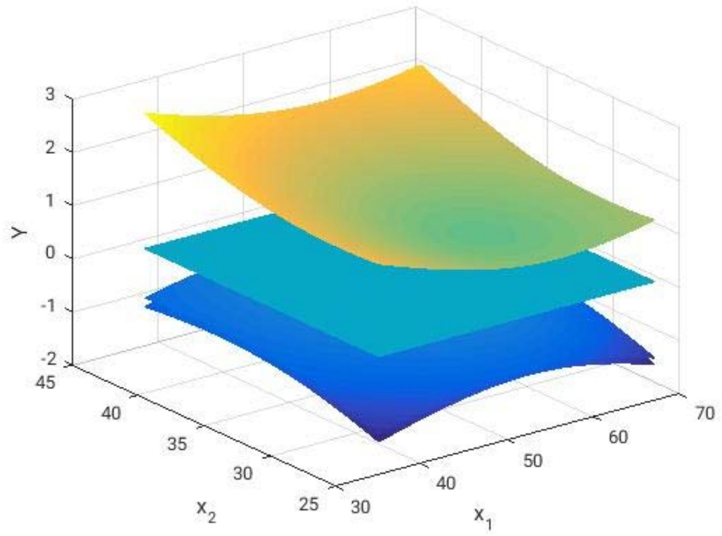

Figure 19 Simultaneous confidence band of ATE for Subclass 7

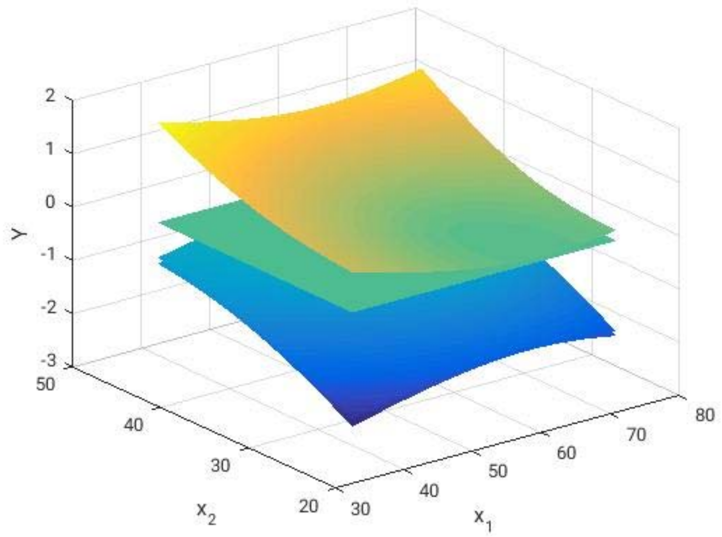

Figure 20 Simultaneous confidence band of ATE for Subclass 9

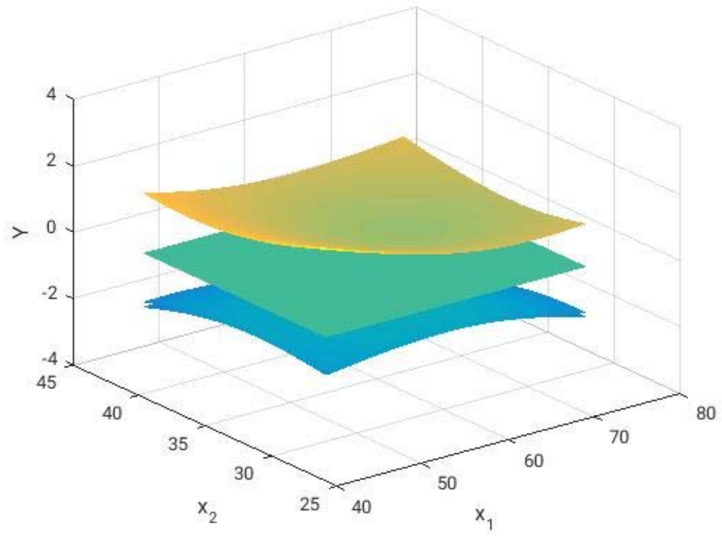

Figure 21 Simultaneous confidence band of ATE for Subclass 10

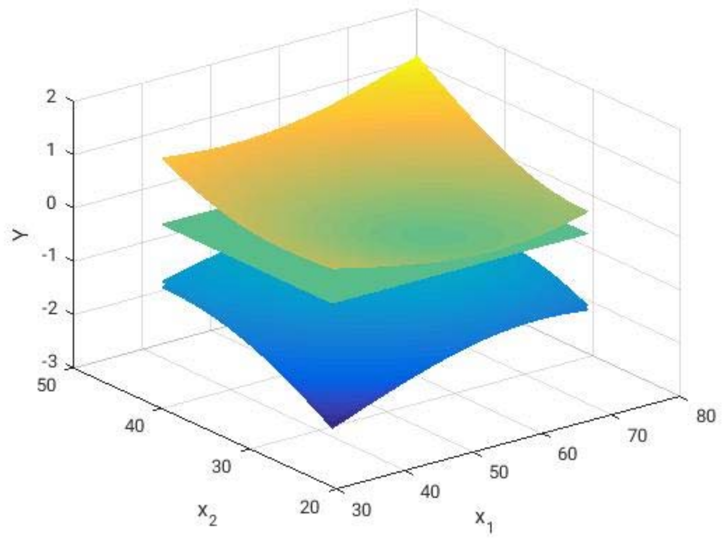

Figure 22 Simultaneous confidence band of ATE for Subclass 12

Table 13 Simultaneous inference p-values on subclasses of NSW data

Subgroup	p-value	Conclusion
1	0.8282	Do not reject H_0
2	0.6638	Do not reject H_0
4	0.2584	Do not reject H_0
5	0.7311	Do not reject H_0
7	0.6018	Do not reject H_0
9	0.0940	Do not reject H_0
10	0.7061	Do not reject H_0
12	0.1028	Do not reject H_0

Although the regions of the covariates considered are relatively wider than the NSW data in Section IV.A., the behavior of the ATE was still explored by holding one independent variable fixed and formulating the difference in treatment effect as a function of the other variable. Since BMI (x_2) has a narrower range than age (x_1), a treatment effect function in terms of age (x_1) by holding x_2 fixed was derived. The fixed values of x_2 were also set at the minimum, mean and maximum BMI for each class. These results are presented in Tables 14-21. In these tables, one may also be able to identify the simulated critical constant *c* and corresponding $\hat{\sigma}$ of the data set in each subclass.

Looking at the solutions of these bands at a fixed x_1, it can be deduced that for any value of x_1 on the given fixed values of x_2, the ATE is non-zero for subgroups 1, 4, 5, 7, 9 and 12. For the remaining subgroups 2 and 10, there are real solutions calculated that for specific values of x_1. This indicates that there is a change in ATE behavior across some interval of x_1 given the fixed values of x_2 considered. Hence, no absolute behavior of the treatment effect can be determined for these subgroups. Figures 15-22 reflect the behavior of the response as a function of x_2. The results show that the relationship of x_1 and the treatment effect is depicted by hyperbolic bands that include the x_1-axis in most subgroups.

For Subgroup 1, no solution for the confidence band as a function of x_1 is calculated, as shown in Table 14. Given the hyperbolic relationship between Y and x_1, this suggests a consistent behavior in the ATE for majority of the groups, as shown in Figure 15. This implies that hyperbolic bands generated for each fixed value of x_2 envelop the x_1-axis; thus, one treatment does not outperform the other. It is noticeable, however, that the corresponding hyperbolic band for the fixed mean value of x_2 provides the narrowest band among the 3 fixed values of x_2 considered. These hyperbolic bands also tend to be narrower towards the mean of the observed x_1 but wider near the ends. This indicates that a wider range of disparity in hemoglobin levels is observed for younger or older patients between the two groups. However, with the inclusion of $y = 0$ in the bands, the superiority of a drug still cannot be established.

Table 14 Confidence band for Subclass 1 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 19.8$	$\frac{-250x_1 - 28269}{50000} \pm 2.26(1.14)\sqrt{\frac{200x_1^2 - 21785x_1 + 701094}{125000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 31.4$	$\frac{-250x_1 - 1543}{50000} \pm 2.26(1.14)\sqrt{\frac{200x_1^2 - 21205x_1 + 583876}{125000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 44.2$	$\frac{-250x_1 - 34439}{50000} \pm 2.26(1.14)\sqrt{\frac{200x_1^2 - 20565x_1 + 661444}{125000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$

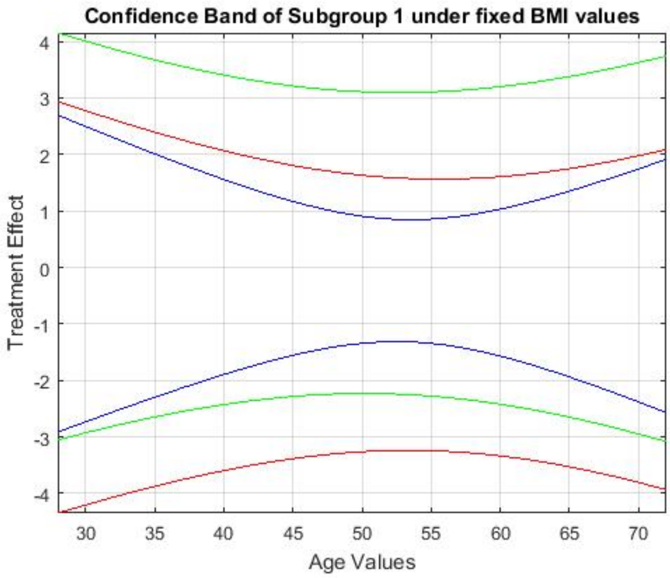

Figure 15 Simultaneous confidence band of Y for Subclass 1 on fixed minimum, mean and maximum values of x_2

Meanwhile, Subgroup 2 displays a different behavior. From the results shown in Table 15, real solutions for the upper band are calculated. This suggests that for some interval of x_1, consistent behavior of ATE is observed. Based on its corresponding figure, Figure 16, it can be observed that given a fixed x_2 at the minimum, the ATE is negative for $x_1 \in [23, 56.26]$. This means that for female subjects from South America with Hispanic descent and has a BMI = 26.8, *Liraglutide* produces increased hemoglobin A1C levels. For the mean and maximum values of x_2, although calculations show that the hyperbolic band has real solutions, these values are outside the covariate region of interest of x_1. As shown in Figure 16, the ATE is likewise negative. Therefore, under the same categorical subject profile mentioned and BMI equal to 31.7 and 41.7, *Liraglutide* provides higher hemoglobin levels. Generally for this subgroup, we can observe the inferiority of *Dulaglutide* as a treatment. It is also interesting to note that among the three fixed values of x_2, the mean generates the narrowest hyperbolic band.

Table 15 Confidence band for Subclass 2 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 26.8$	$\frac{109x_1 - 13209}{50000} \pm 2.67(0.74)\sqrt{\frac{300x_1^2 - 28080x_1 + 759951}{250000}}$	$x_{1+} = 20.32, 56.25$ $x_{1-} = \phi$
$x_2 = 31.7$	$\frac{109x_1 - 16884}{50000} \pm 2.67(0.74)\sqrt{\frac{1200x_1^2 - 119180x_1 + 3128739}{1000000}}$	$x_{1+} = 2.21, 73.03$ $x_{1-} = \phi$
$x_2 = 41.7$	$\frac{109x_1 - 24834}{50000} \pm 2.67(0.74)\sqrt{\frac{1200x_1^2 - 133180x_1 + 4964139}{1000000}}$	$x_{1+} = -13.02, 85.53$ $x_{1-} = \phi$

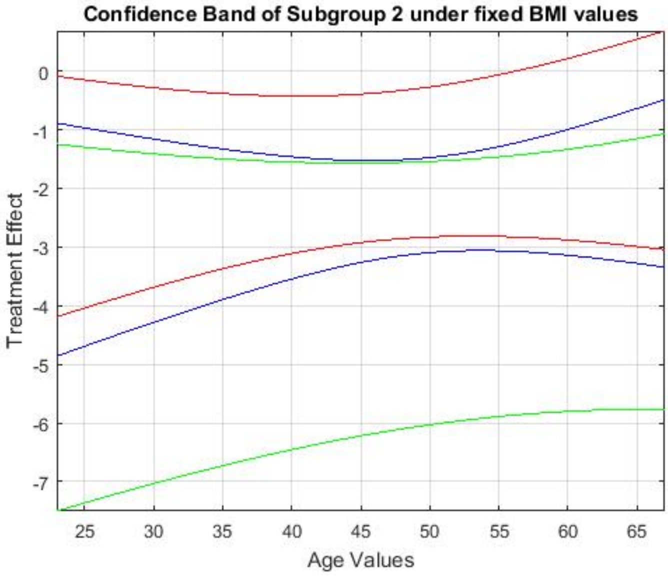

Figure 16 Simultaneous confidence band of Y for Subclass 2 on fixed minimum, mean and maximum values of x_2

Table 16 and Figure 17 illustrate the behavior of the covariates for Subclass 4 which consists of male, Hispanic respondents from North America. Based on the solutions of the hyperbolic bands, it can be concluded that the confidence bands for the ATE estimates contain 0. This suggests that one drug is non-inferior over the other. The widths of the bands for the minimum and maximum values of x_2 are comparable, while the confidence band on a fixed maximum value of x_2 is considerably wide. In all three fixed values though, the band is narrower towards the mean of the observed x_1-values. The wide band may be an indicator that for subjects with this subject profile and has high BMI, *Liraglutide* gives higher hemoglobin levels as shown in the majority of band being negative. This behavior is not quite apparent in the fixed minimum and mean values of x_2, because the band also contains comparable range in the positive side.

Table 16 Confidence band for Subclass 4 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 20.8$	$\frac{-33x_1 - 3511}{10000} \pm 2.66(1.13)\sqrt{\frac{6x_1^2 - 646x_1 + 19615}{5000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 30.5$	$\frac{-66x_1 - 20971}{20000} \pm 2.66(1.13)\sqrt{\frac{24x_1^2 - 2532x_1 + 72727}{20000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 44.3$	$\frac{-330x_1 - 252929}{100000} \pm 2.66(1.13)\sqrt{\frac{600x_1^2 - 60540x_1 + 2046151}{500000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$

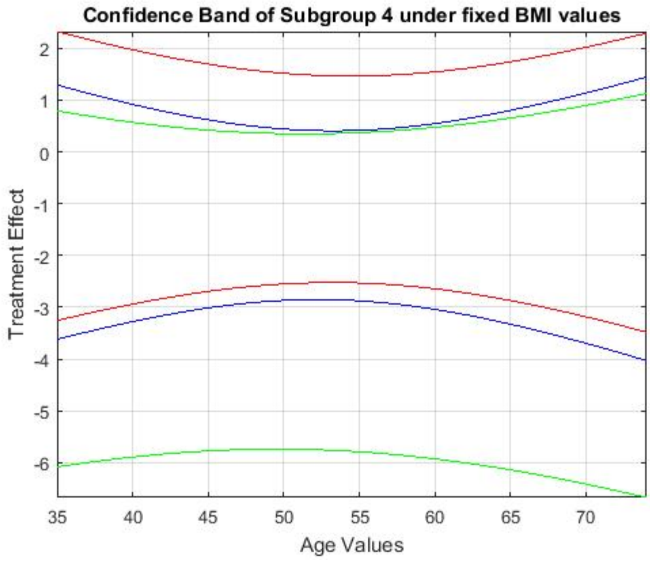

Figure 17 Simultaneous confidence band of Y for Subclass 4 on fixed minimum, mean and maximum values of x_2

For Subgroup 5, no real solution for the hyperbolic bands is calculated, as shown in Table 17. This means that *Dulaglutide* and *Liraglutide* are not significantly different in increasing hemoglobin levels. Figure 18 shows this behavior as well. For this subgroup, the band is widest for a fixed minimum value of x_2 while the fixed mean is narrowest. All three hyperbolic bands are narrowest at the middle values of x_1. However, in all cases, superiority of one drug is not established.

Table 17 Confidence band for Subclass 5 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 24.5$	$\frac{-1554x_1 + 79677}{20000} \pm 3.28(1.22)\sqrt{\frac{416x_1^2 - 44792x_1 + 1300293}{40000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 30.2$	$\frac{-777x_1 + 35250}{10000} \pm 3.28(1.22)\sqrt{\frac{2600x_1^2 - 281660x_1 + 7671387}{250000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 34.5$	$\frac{-1554x_1 + 63577}{20000} \pm 3.28(1.22)\sqrt{\frac{416x_1^2 - 45272x_1 + 1271693}{40000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$

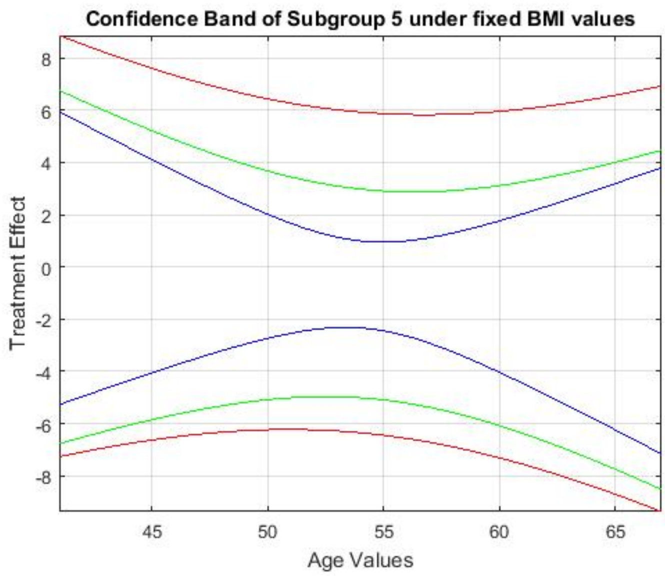

Figure 18 Simultaneous confidence band of Y for Subclass 5 on fixed minimum, mean and maximum values of x_2

Table 18 and Figure 19 show the result for Subgroup 7, comprising of female, non-hispanic subjects from North America. With the non-real solutions of the hyperbolic bands and their corresponding behavior reflected in Figure 19 it is observed that the two drugs considered have a non-inferior effect for this type of subject profile. Like the previous subgroups, the bands are narrowest at the extreme ends of the observed values of x_1 and tend to have narrower effect towards the center, given any of the fixed effect of x_2 .

Table 18 Confidence band for Subclass 7 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 24.6$	$\frac{-415x_1+15546}{50000} \pm 2.62(0.76)\sqrt{\frac{250x_1^2-28181x_1+867449}{250000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 33.8$	$\frac{-415x_1+32658}{50000} \pm 2.62(0.76)\sqrt{\frac{250x_1^2-28043x_1+806821}{250000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 47.2$	$\frac{-415x_1+57582}{50000} \pm 2.62(0.76)\sqrt{\frac{250x_1^2-27842x_1+951316}{250000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$

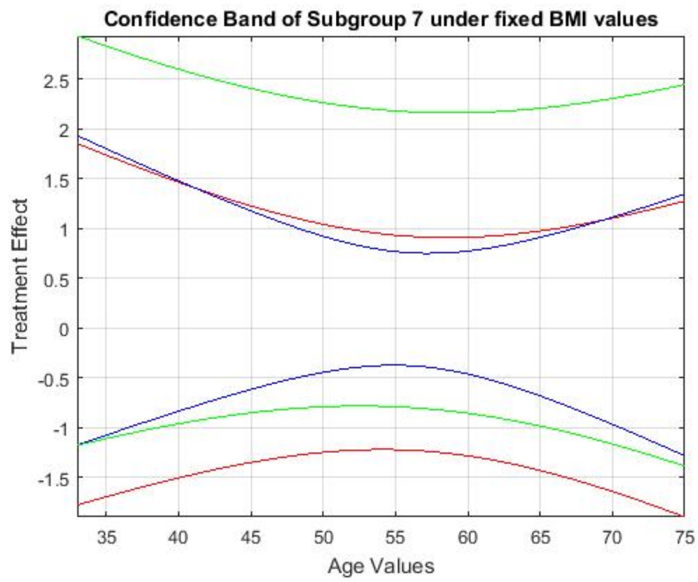

Figure 19 Simultaneous confidence band of Y for Subclass 7 on fixed minimum, mean and maximum values of x_2

Also, *Dulaglutide* and *Liraglutide* are shown to be non-inferior for female, non-hispanic Europeans as shown in Table 19 and Figure 20. However, a closer look at the hyperbolic bands show that for a fixed value of x_2 at the minimum, majority of the band lies on the negative treatment effect. Meanwhile, when x_2 is fixed at the maximum, this band tends to move upward, causing majority of the difference in the treatment groups to be positive. This suggests that under this subject profile, patients with lower BMI will have increased hemoglobin levels when taking *Liraglutide* while those who have higher BMI will have higher hemoglobin levels when taking *Dulaglutide*. At the mean value of x_2, the hyperbolic band does not tend to favor the positive nor the negative direction.

Table 19 Confidence band for Subclass 9 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 21.8$	$\frac{-26x_1+6141}{10000} \pm 2.58(0.71)\sqrt{\frac{125x_1^2-14919x_1+512988}{250000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 32.8$	$\frac{-26x_1+129}{10000} \pm 2.58(0.71)\sqrt{\frac{125x_1^2-14424x_1+451883}{250000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 44.8$	$\frac{-26x_1+6969}{10000} \pm 2.58(0.71)\sqrt{\frac{125x_1^2-13884x_1+474923}{250000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$

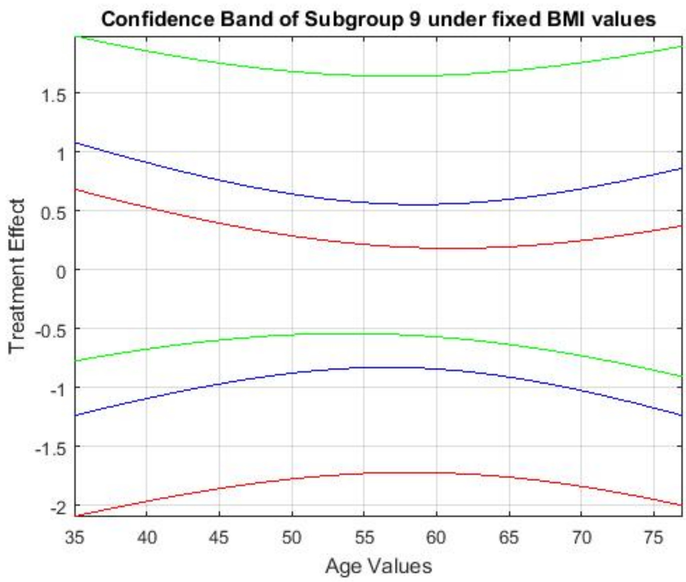

Figure 20 Simultaneous confidence band of Y for Subclass 9 on fixed minimum, mean and maximum values of x_2

Table 20 and Figure 21 show the results derived for Subgroup 10. This subgroup is comprised of male, non-Hispanic respondents from North America. The superiority of one drug is still not established for this group. However, it is interesting to note that x_1 is undefined for a fixed minimum x_2. Therefore, for subjects with this profile and has BMI = 24.1, we are not able to provide any observation of the behavior of Y in terms of x_1. On the available hyperbolic bands, the mean continues to produce narrower bands than the maximum.

Table 20 Confidence band for Subclass 10 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 24.1$	$\frac{-720x_1 + 52579}{25000} \pm 2.64(0.73)\sqrt{\frac{700x_1^2 - 77865x_1 + 1830141}{500000}}$	$x_{1+} = 77.58$ $x_{1-} = 27.13$
$x_2 = 32.7$	$\frac{-720x_1 + 42173}{25000} \pm 2.64(0.73)\sqrt{\frac{1400x_1^2 - 161940x_1 + 481569}{1000000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 43.6$	$\frac{-720x_1 + 3623}{3125} \pm 2.64(0.73)\sqrt{\frac{350x_1^2 - 35580x_1 + 1115039}{250000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$

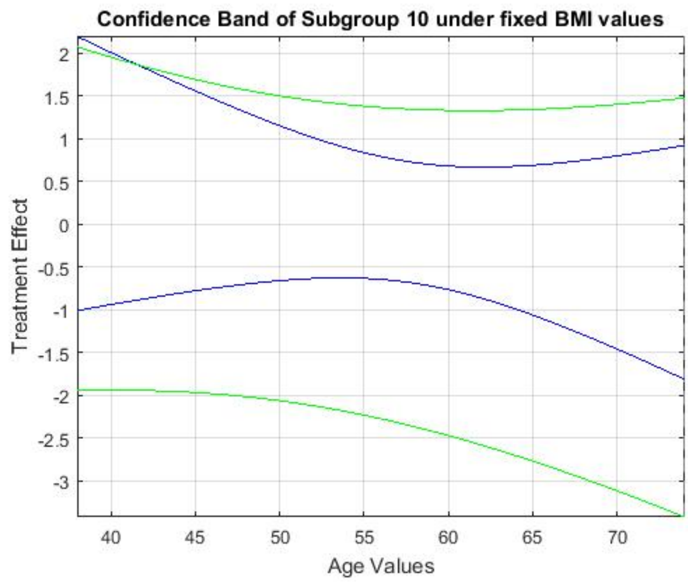

Figure 21 Simultaneous confidence band of Y for Subclass 10 on fixed minimum, mean and maximum values of x_2

For the last subgroup, the same observation is made about the non-inferiority of the two drugs being studied. Results are shown in Table 21 and Figure 22. While no treatment effect is realized on all three bands, we note that majority of the band generated under fixed minimum and mean values of x_2 tend towards the negative direction. Thus, it can be observed that for male, non-Hispanic Europeans with low to moderate BMI, the drug *Liraglutide* may be a more effective treatment in increasing hemoglobin levels among diabetic patients.

Table 21 Confidence band for Subclass 12 under fixed x_2

x_2	Confidence band: $\mathbf{x}'\hat{\boldsymbol{\beta}}_T - \mathbf{x}'\hat{\boldsymbol{\beta}}_C \pm c\hat{\sigma}\sqrt{\mathbf{x}'\boldsymbol{\Delta}\mathbf{x}}$	Solution on χ_r
$x_2 = 24.1$	$\frac{460x_1 - 65301}{50000} \pm 2.58(0.74)\sqrt{\frac{25x_1^2 - 2918x_1 + 102006}{50000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 32.6$	$\frac{460x_1 - 39709}{50000} \pm 2.58(0.74)\sqrt{\frac{25x_1^2 - 2582x_1 + 72438}{1000000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$
$x_2 = 45.7$	$\frac{920x_1 - 19551}{100000} \pm 2.58(0.74)\sqrt{\frac{350x_1^2 - 8756x_1 + 246915}{200000}}$	$x_{1+} = \phi$ $x_{1-} = \phi$

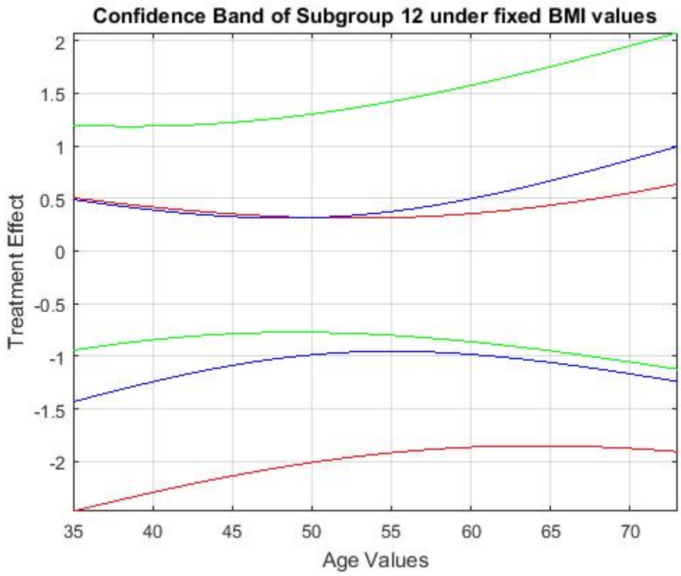

Figure 22 Simultaneous confidence band of Y for Subclass 12 on fixed minimum, mean and maximum values of x_2

In general, as illustrated in Figures 15-22, there is an insignificant treatment effect given the range of x_1 of the different subgroups under a fixed x_2. This supports the surfaces generated as confidence bands over the entire space of interest of x_1 and x_2.

As a summary, the results of the exploratory analysis of the *Lilly* data set suggest that there is no significant treatment effect between the two drugs. This means that the drug *dulaglutide* is non-inferior to *liraglutide* in stabilizing the hemoglobin A1C levels of type II diabetes patients. However, with this analysis, we are able to identify the treatment effect behavior given some information on a subject's region of origin, civil status, gender, ethnicity, BMI and age.

5. Conclusion

In this paper, an alternative method for analyzing average treatment effect (ATE) is exploited. The proposed method first requires balancing among the categorical variables resulting in several homogeneous subgroups and then performing a detailed statistical analysis via the novel method of simultaneous inference in regression. Through this procedure, we are able to explain analysis of ATE based on subject covariate profile. Thus, we are able to identify for which characteristics of the covariates the treatment is effective. This consequently allows for providing more personalized, effective treatments to subjects. Two real data sets are utilized as illustrations of our proposed method.

Acknowledgements

I thank my adviser Dr. Bimal K. Sinha from University of Maryland Baltimore County who provided insights and expertise that greatly assisted the research and highly improved the manuscript.

References

Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivar Behav Res. 2011; 46(3): 399-424.

Hoel PG. Confidence regions for linear regression. Proceedings of the Second Berkeley Symposium University of California Press, 1951.

Lalonde R. Causal effects in non-experimental studies: reevaluating the evaluation of training programs. J Am Stat Assoc. 1999; 94(448): 1053-1062.

Liu W. Simultaneous inference in regression. 1st ed. Florida: CRC Press Taylor & Francis Group; 2010.

Liu W. Jamshidian, M. Zhang, Y. and Donnelly, J. Simulation-based simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals. J Comput Graph Stat. 2005; 14(2): 459-484.

Rosenbaum PR, Rubin, DB. The central role of the propensity score in observational studies for causal effects. J Royal Stat Soc. 1983; 70(1): 41-45.

Stewart PW. The graphical advantages of finite interval confidence band procedures. Commun Stat Theory Methods. 1991; 20(12): 3975-3993.

Scheffe H. A method for judging all constraints in analysis of variance. Biometrika. 1953; 40(1/2): 87-104.

Scheffe H. The analysis of variance. New York: Wiley; 1959.

SAS Clinical Trial Data Portal, Eli Lilly Study. USA; 2018.