Quantile Regression Analysis for Statin effects on glucose level

Contents

1 introduction

Approximately, 18 million deaths annually per year are caused by cardiovascular diseases(CVD) and similar to this number of nonfatal cardiovascular events (Hay et al. 2017). For 2011, annual costs for CVD and stroke were $320.1 billion, which is more than cancer cost. This cost include \$195.6 billion in direct costs (healthcare costs), and cost of future productivity loss is \$124.5 (Mozaffarian et al. 2015). Abnormal lipid ratio is a risk factor for CVD (Yusuf et al. 2004). Lowering low-density lipoprotein (LDL) using statin reduces risk of cardiovascular diseases even with population with no CVD (Yusuf et al. 2016). It is known that diabetes is a risk factor for CVD, however, it is shown that elevated glucose, prediabetes,is also a risk factor for CVD (Balkau et al. 2004).

Statin use associated with high risk of a new onset diabetes in normoglcemic (HR 1.19,95% CI , 1.05 to 1.35) and in prediabetes (HR 1.24 95% CI , 1.11 to 1.38). On the other hand, overall mortality risks went down in both normoglycemic (HR 0.70; 95 % CI, 0.66 to 0.80) and IFG patients (HR 0.77, 95 % CI, 0.64 to 0.91) with statin use (Castro et al. 2016). For population with one or more risk factor for diabetes, statin use associate with 39% reduction in the primary end points (the hazard ratio (HR) 0.61), and a 28% (HR 1.28)increase in diabetes(Ridker et al. 2012).Pandya at el.[pandya2015] suggested the dis-utility associated with taking a pill daily, statin price, and the risk of statin-induced diabetes negatively impacted the cost effectiveness of statin use recommendation.

Several studies esimated the association of different factors like waist circumference, Cholestrol with the conditional mean of glucose using OLS or logistic regression model. The drawback in these study is not considering the dispersion of the association of covariates with conditional distribution of the dependent variables, glucose levele.

2 Quantile Regression

Quantile regression is an important tool used to regress the dependent variable with high variance over the independent variables. QR is developed to study the relationships between variables that have week or no-relationships between their means. One of the advantage of using QR over OLS is robust for outliers.

For a random variable X, the cumulative distribution function (CDF) is

$$F(X) = P(X \leq x)$$

, and the τth quantile of X is defined by

$$F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}$$

where $0 < \tau < 1$. Let the loss function is defined by

$$\rho_\tau(u) = u(\tau - I_{(u<0)})$$

where I is the indicator function (Koenker 2005). The quantile estimator is the value that minimizes the expected loss function

$E\rho_\tau(X - \hat{x}) = (\tau - 1)\int_{-\infty}^{\hat{x}}(x - \hat{x})dF(x) + \tau\int_{\hat{x}}^{-\infty}(x - \hat{x})dF(x)$. Differentiating with respect to $\hat{x}$, we get

$$0 = (\tau - 1)\int_{-\infty}^{\hat{x}} dF(x) + \tau\int_{\hat{x}}^{-\infty} dF(x) = F(\hat{x}) - \tau.$$

Due to monotnicity of the cumulative distribution function, any solution that satisfies $\{x : F(x) = \tau\}$ is a minimizing for the expected loss function.

Least square method expresses conditional mean of y given x as $\mu(x) = x^T\beta$ and it solves

$$\min_{\beta\in\mathcal{R}^p}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2.$$

Quantile regression expresses conditional quantile function $Q_y(\tau|x) = x^T\beta(\tau)$ and solve

$$\min_{\beta\in\mathcal{R}^p}\sum_{i=1}^{n}\rho_\tau(y_i - x_i^T\beta)^2.$$

This minimization problem can be reformulated to a linear programming problem

$$\min_{\beta \in \mathcal{R}^p}$$

2.1 Quantile Regression Technique

3 Methods

A multivariate quantile regression model is used to assess the characteristics of the association variability in different quantiles of the conditional distribution of the glucose levels.

The dependent variables in our model are gender, race, age, BMI, total cholesterol, Choesterol drug uses (yes or no), and waist circumference. All type of cholesterol drug are included including statins. The included races are non-Hispanic white, non-Hispanic black, Hispanic, or other.

4 Numerical Example

4.1 Data

The data used in this study is National Health and Nutrition Examination Survey data (NHANES)(Disease Control and (CDC) 2018). The survey examines a nationally representative sample of U.S. population.It focuses on variety of health and nutrition measurements. In this study, we cumulated 6 cycles of NHANES data (2007–2018). There are around 12,000 records. We selected population age between 20 and 80. BMI are classified into different categories according to underweight, 18.5 kg/m2; normal weight, 18.5 to 25 kg/m2; overweight, 25 to 30 kg/m2; obese, 30 to 35 kg/m2; and very obese more than 35 kg/m2.

Syntax	Male	Femal
count	5990	6416

Syntax	Male	Femal
Mean of Age	49.9	49.73
glucose	112.618	106.47
TC	Title	Here’s this
Statin use (ratio)	0.198	0.181

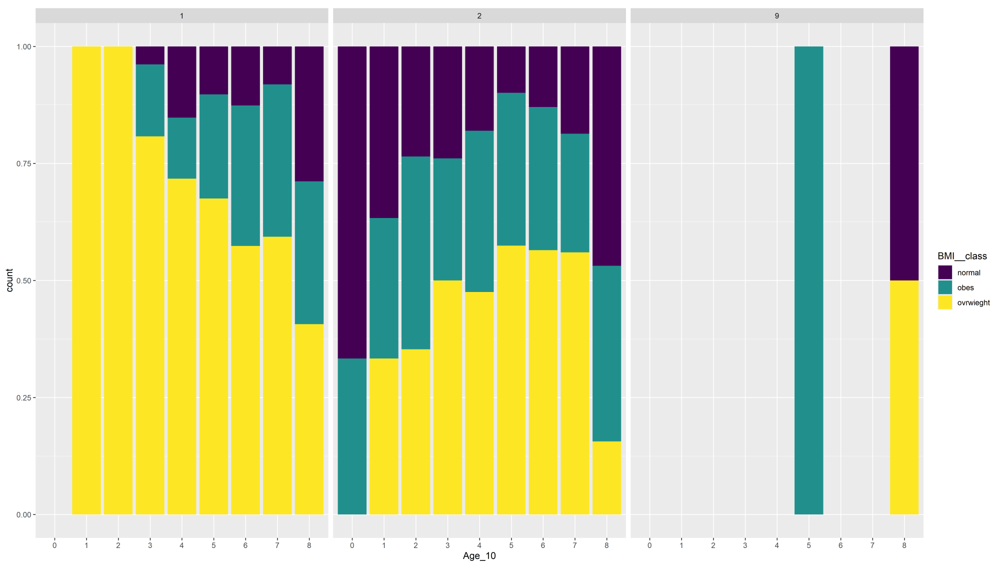

Figure 1: A better figure caption

to cite figure #See Figure 1 However,.

So,

5 Results

The resulting estimate of effects on conditional mean of glucose level may not reflect the size and nature of theses effects on lower or upper quantile. For example, in the Figure 4, the conditional mean shows the conditional mean effects of gender on glucose level is about 7. However, the disparity of the gender effects on lower tails is much smaller which is about 3 but the disparity is higher for the upper tail of the distribution somewhat larger than 10

Figure 2: A better figure caption

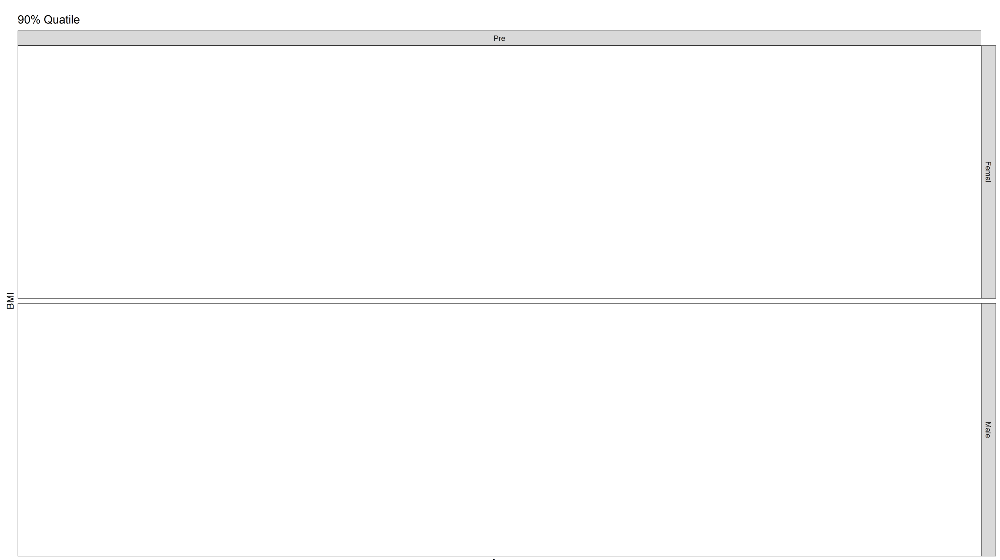

Figure 3: A better figure caption

mg/dL.

From the OLS it is obvious statin users have on average higher glucose levels if compared to non-statin users which is around 10 mg/dL. The disparity in glucose level for statin users vs non-statin users is negligible in the left tail, however, statin uses seems to be associated with a rather large effects on glucose levels somewhat larger than 55 mg/dL for the right quantile.

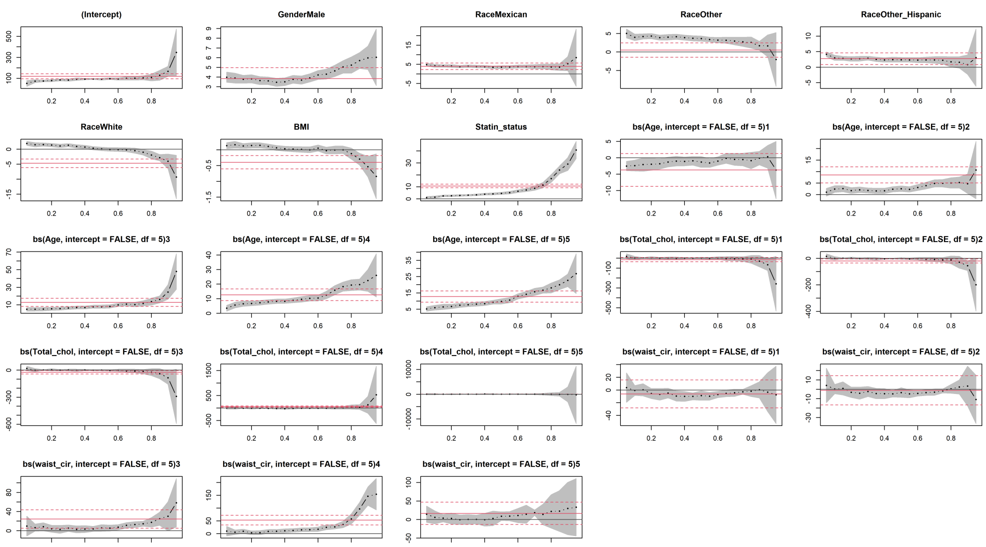

Figure 4: A better figure caption

So,

therefore

So,

Age effect modeled as a quadratic factor. The age effect is linear in general, see Figure 8. The slope for the left quantile is lower than the slope of the right quantile. This can be interpreted as age effect tends to increase over the entire range of conditional distribution of glucose in faster rate for the higher quantile than in the left quantile.In other word, diabetes population have higher rate of change with respect to age if compared to non diabetes population. It was found that there is a significantly correlation between glucose and age(Tsaousis 2014).

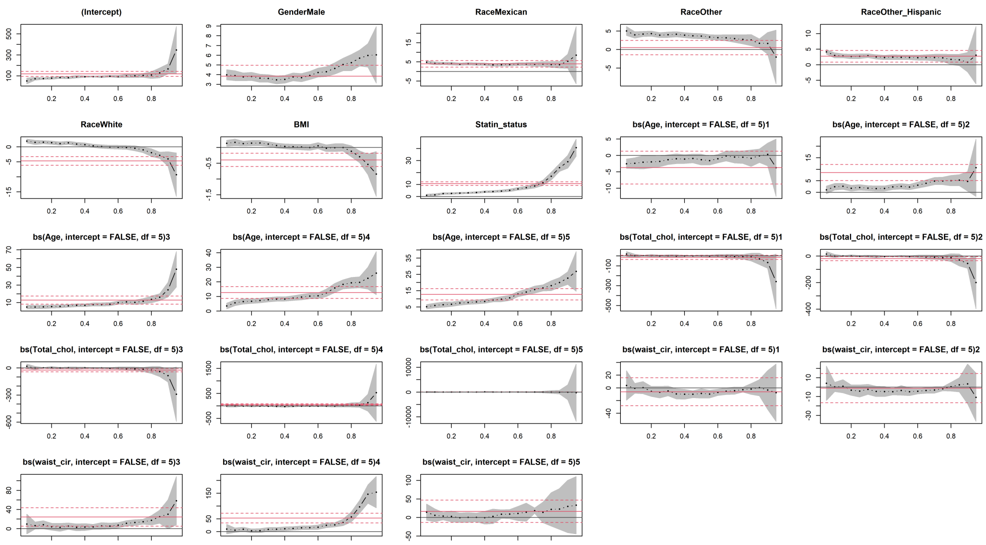

Figure 5: A better figure caption

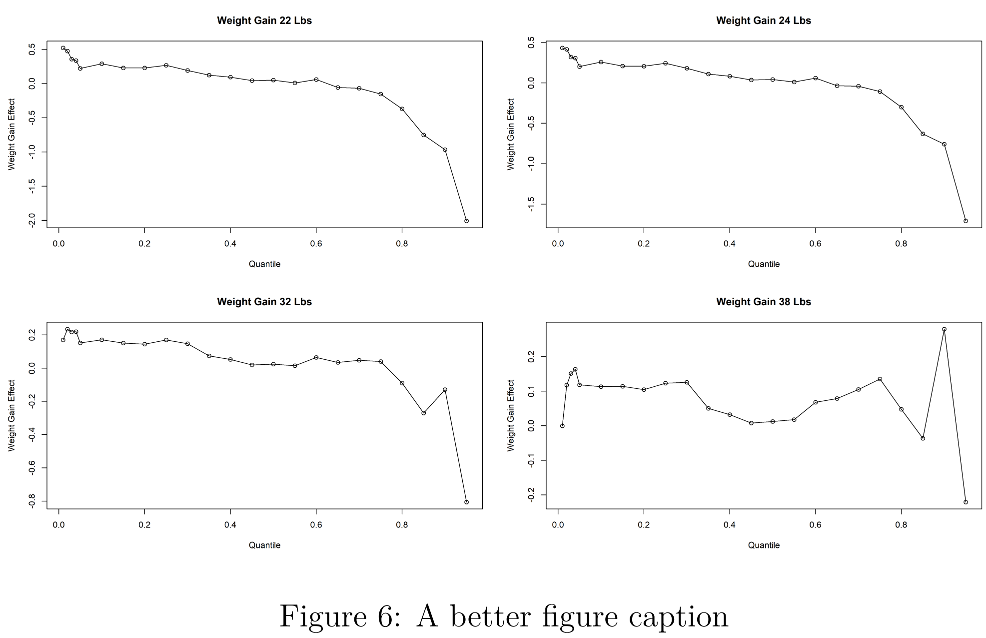

Figure 6: A better figure caption

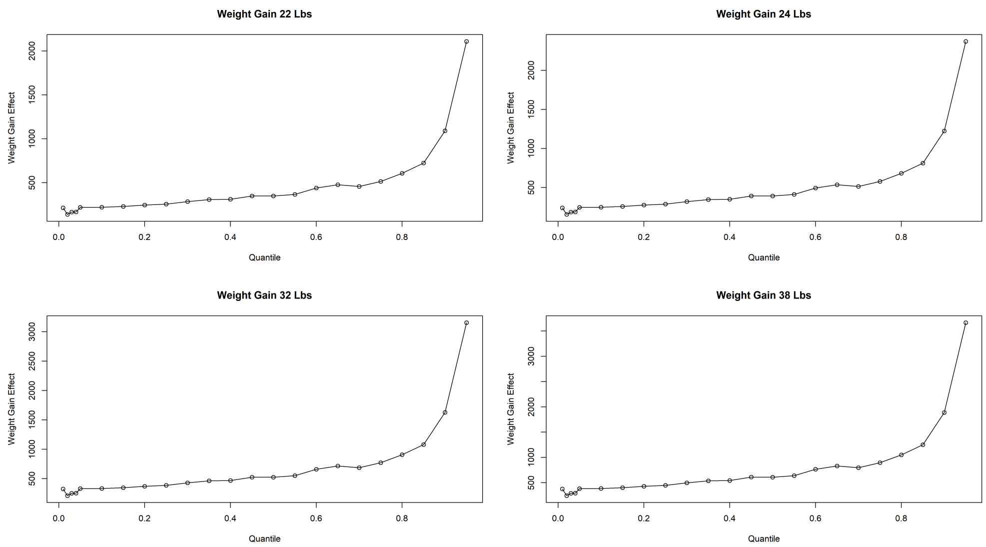

Figure 7: A better figure caption

The relation between total cholesterol and fasting glucose has been studied in (Tsaousis 2014),(Chang et al. 2011). They found that there is a positive correlation between the two groups. However, we found that

So,

Quadratic effect of total cholesterol on the conditional distribution of glucose is convex. At the lower tail, decreasing glucose level by somewhat around 1.4 up to around 180 Cholesterol level. Then, started to increase glucose levels after TC beyond 180. At higher quantile, the reverse in the cholesterol effects is shifted to higher cholesterol levels which around 200, Figure 10. The relation between total cholesterol and fasting glucose has been studied in (Tsaousis 2014),(Chang et al. 2011). They found that there is a positive correlation between the two groups on average.

So,

So,

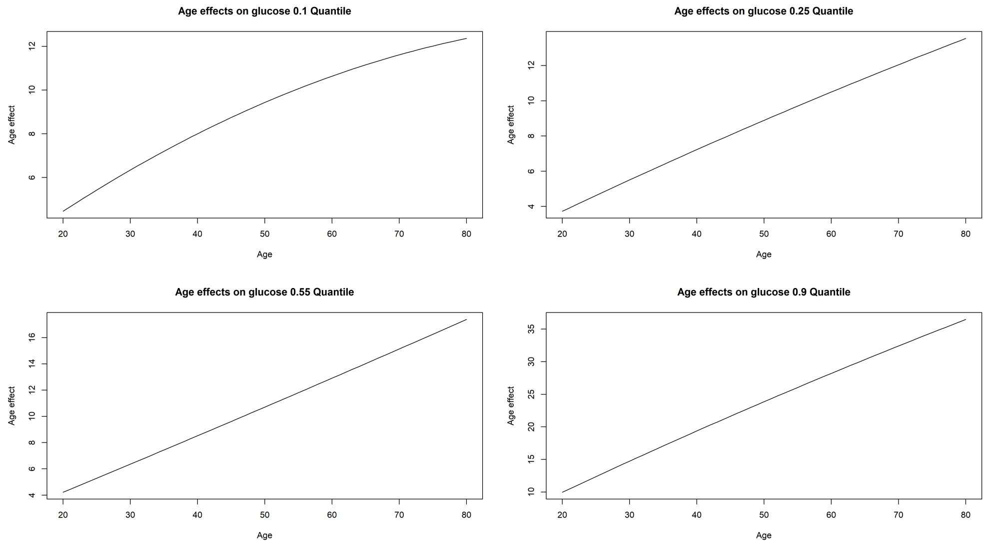

Figure 8: Illustration of the quadratic age effect on glucose leveles for four different quantiles of the conditional glucose distribution. The

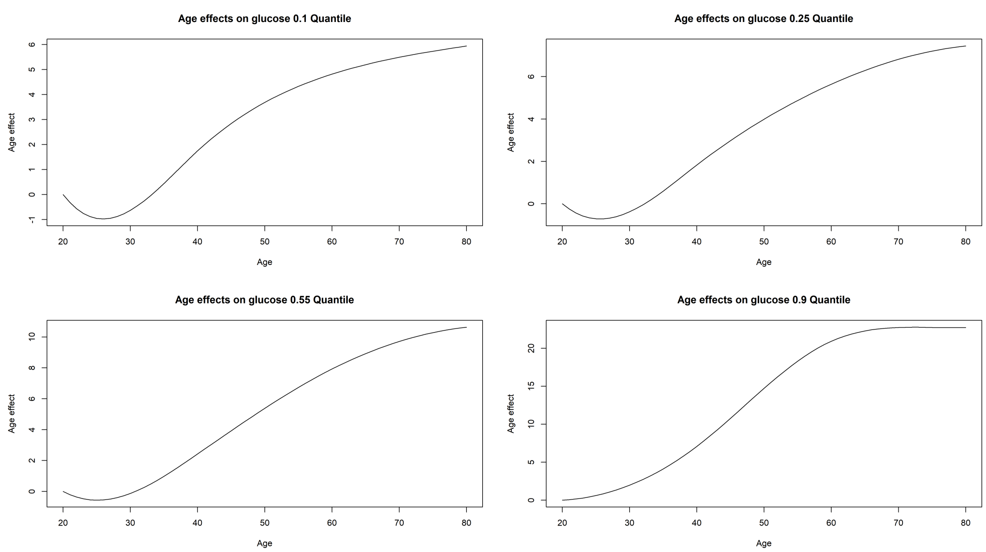

Figure 9: Illustration of the quadratic age effect on glucose leveles for four different quantiles of the conditional glucose distribution. The

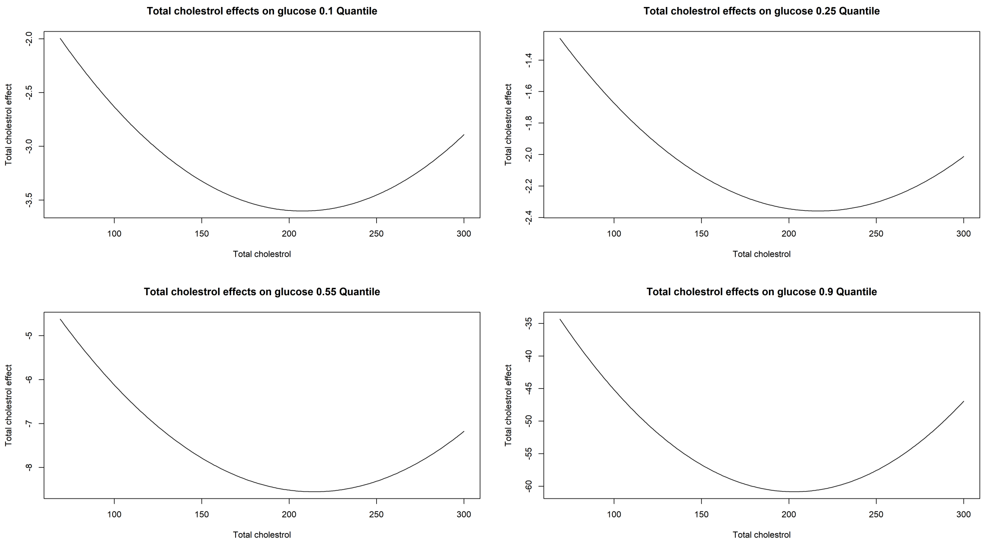

Figure 10: Illustration of the quadratic cholestrol effect on glucose leveles for four different quantiles of the conditional glucose distribution. The

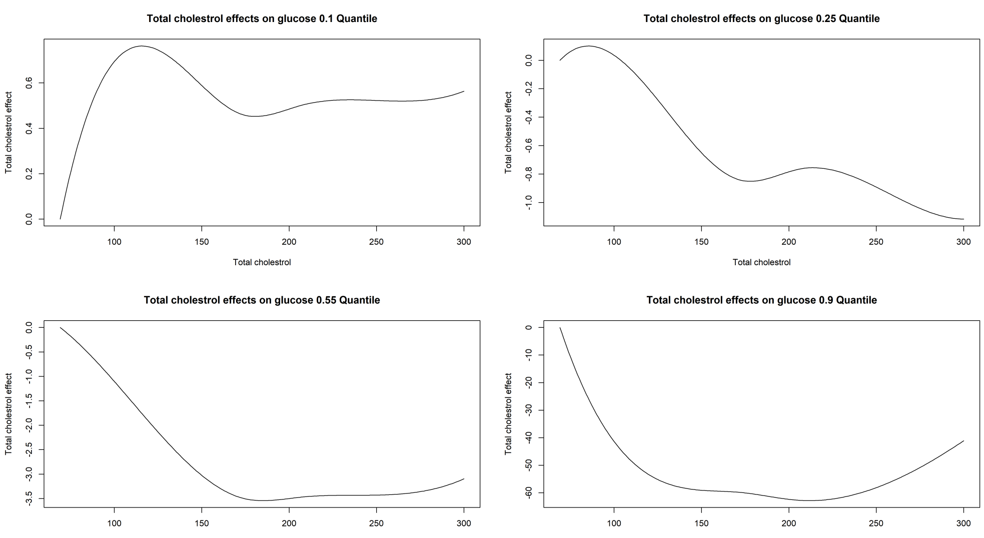

Figure 11: Illustration of the quadratic cholestrol effect on glucose leveles for four different quantiles of the conditional glucose distribution. The

6 Conclusion

Multivariate quantile regression is used to study effects of different variables on fasting glucose level. Our study showed that people who have TC levels around 190 mg/dL have lowest fasting glucose levels for the lowest quantile, for the second quantile optimal cholesterol level is around 220mg/dL, and for the upper quantile the optimal cholesterol level is around 200 mg/dL. Moreover, Statin effects on glucose at lower quantile is very small if compared to upper quantile.

7 References

Balkau, Beverley, Gang Hu, Qing Qiao, Jaakko Tuomilehto, Knut Borch-Johnsen, K Pyorala, DECODE Study Group, European Diabetes Epidemiology Group, and others. 2004. "Prediction of the Risk of Cardiovascular Mortality Using a Score That Includes Glucose as a Risk Factor. the Decode Study." *Diabetologia* 47 (12). Springer Science & Business Media: 2118.

Castro, M Regina, Gyorgy Simon, Stephen S Cha, Barbara P Yawn, L Joseph Melton, and Pedro J Caraballo. 2016. "Statin Use, Diabetes Incidence and Overall Mortality in Normoglycemic and Impaired Fasting Glucose Patients." *Journal of General Internal Medicine* 31 (5). Springer: 502–8.

Chang, Jin-Biou, Nain-Feng Chu, Jhu-Ting Syu, An-Tsz Hsieh, and Yi-Ren Hung. 2011. "Advanced Glycation End Products (Ages) in Relation to Atherosclerotic Lipid Profiles in Middle-Aged and Elderly Diabetic Patients." *Lipids in Health and Disease* 10 (1). Springer: 228.

Disease Control, Centers for, and Prevention (CDC). 2018. "National Health and Nutrition Examination Survey Data (Nhanes."

Hay, Simon I, Sudha P Jayaraman, Alejandra G Contreras Manzano, Anoushka Millear, Laura Kemmer, Brent Bell, Juan Jesus Carrero, et al. 2017. "GBD 2015 Risk Factors Collaborators. Global, Regional, and National Comparative Risk Assessment of 79 Behavioural, Environmental and Occupational, and Metabolic Risks or Clusters of Risks, 1990-2015: A Systematic Analysis for the Global Burden of Disease Study 2015 (Vol 388, Pg 1659, 2016)." *Lancet* 389 (10064). ELSEVIER SCIENCE INC 360 PARK AVE SOUTH, NEW YORK, NY 10010-1710 USA: E1–E1.

Koenker, Roger. 2005. "Quantile Regression, Volume 38 of." *Econometric Society Monographs.*

Mozaffarian, Dariush, Emelia J Benjamin, Alan S Go, Donna K Arnett, Michael J Blaha, Mary Cushman, Sarah De Ferranti, et al. 2015. "Executive Summary: Heart Disease and Stroke Statistics—2015 Update: A Report from the American Heart Association." *Circulation*

131 (4). Am Heart Assoc: 434–41.

Ridker, Paul M, Aruna Pradhan, Jean G MacFadyen, Peter Libby, and Robert J Glynn. 2012. "Cardiovascular Benefits and Diabetes Risks of Statin Therapy in Primary Prevention: An Analysis from the Jupiter Trial." *The Lancet* 380 (9841). Elsevier: 565–71.

Tsaousis, Konstantinos T. 2014. "Blood Glucose and Cholesterol Concentrations in a Mediterranean Rural Population of Andros Island, Greece." *International Journal of Preventive Medicine* 5 (11). Wolters Kluwer–Medknow Publications: 1464.

Yusuf, Salim, Jackie Bosch, Gilles Dagenais, Jun Zhu, Denis Xavier, Lisheng Liu, Prem Pais, et al. 2016. "Cholesterol Lowering in Intermediate-Risk Persons Without Cardiovascular Disease." *New England Journal of Medicine* 374 (21). Mass Medical Soc: 2021–31.

Yusuf, Salim, Steven Hawken, Stephanie Ôunpuu, Tony Dans, Alvaro Avezum, Fernando Lanas, Matthew McQueen, et al. 2004. "Effect of Potentially Modifiable Risk Factors Associated with Myocardial Infarction in 52 Countries (the Interheart Study): Case-Control Study." *The Lancet* 364 (9438). Elsevier: 937–52.