Exam Code: DA-04

A Statistical Analysis of Education Expenditure Practices Across Both Legacy and Core Cities in Europe

Executive Summary

This analysis attempts to describe how local cities in Europe spend their revenue on education expenditures. Potential predictors of education expenditures were examined including CPI, direct total expenditures, city types, revenue generated by the city, revenue generated from utility, total revenue, cash holdings, and outstanding debt. The natural log of education expenditures was positively associated with revenue generated by the city ($p < 0.0001$) and had a nonsignificant negative association with revenue generated from utility ($p = 0.412$). Cash holdings was found to have a negative interaction with revenue generated by the city ($p < 0.0001$) and a positive interaction with revenue generated from utility ($p <0.0001$). While outstanding debt, had a negative interaction with revenue generated from utility ($p <0.0001$). The average difference in the natural log of education expenditures in dollars per capita between legacy and core cities is 0.3260 ($p < 0.0001$). Most of the data was collected from core cities so the inferences based on this dataset for legacy cities may not be as accurate considering the smaller sample size. Future work on this subject may included the analysis of cities outside of Europe and the inclusion of other interesting variables such as average income and demographic data.

Introduction

Education expenditures are a large part of government spending and an investment in the citizens of local cities (2). The subject of funding for education is relevant to society because access to education is associated with better jobs, higher earnings, and better access to resources for good health (2). The primary focus of this analysis is to describe how legacy and core cities in Europe spend their revenue on education expenditures. The dataset that was analyzed includes information about city level expenditures and revenues from 1998 to 2017. All the quantitative variables in this analysis were converted to real per capita dollars.

Study Design

This project addresses three specific research questions. The first research question focuses on the association between the level of revenue and the level of education expenditures in local cities. I expect that revenue will be positively associated with education expenditures. To test this hypothesis, multiple linear regression with education expenditures as the response variable and revenue generated by the city as well as revenue generated from utility, as two of the predictors, will be used. The parameter estimates for those two predictors will indicate whether there is a positive or negative relationship, and a t-test will determine whether the relationship is statistically significant.

The second research question involves exploring the effect that unit increases in cash holdings and outstanding debt had on education expenditures for larger values of revenue. I expect that a unit increase in cash holdings would result in a greater increase in education expenditures for larger values of revenue generated by the city or revenue generated from utility. I expect that a unit increase in outstanding debt would result in a greater decrease in education expenditures for larger values of revenue generated by the city or revenue generated from utility. Multiple linear regression with education expenditures as a response variable will also be used to approach this question. The interpretations of the interactions of cash holdings and outstanding debt with revenue generated by the city and revenue generated from utility will be useful to answer this question. A t-test will determine if the interactions are considered significant.

The third research question involves comparing legacy and core cities to see if there is a difference between the two city types. More specifically, exactly how much they differ in education expenditures on average will be examined. Again, multiple linear regression is used with education expenditures as the response variable and the city type as a qualitative predictor. The interpretation of the parameter estimate for this qualitative predictor will help answer what the average difference is in education expenditures between legacy and core cities and a t-test will determine whether it is a statistically significant difference.

Exploratory Data Analysis

Participants

All the cities included in this study are European cities and their data was collected from 1998 to 2017. There were several different quantitative variables in the dataset, and all were converted to real per capita dollars. The variables that were used in this analysis include those that are associated with revenue such as total revenue, revenue generated by the city, and revenue generated from utility. Variables that represent spending were also included such as direct total expenditures and education expenditures. Other monetary information such as the cash holdings, outstanding debt, and the consumer price index (CPI) of the cities were included as well. Finally, the qualitative variable that represents the type of city (legacy or core) was also used. Data from a total of n = 4360 cities was utilized in the analyses.

Visualizing the Variables of Interest

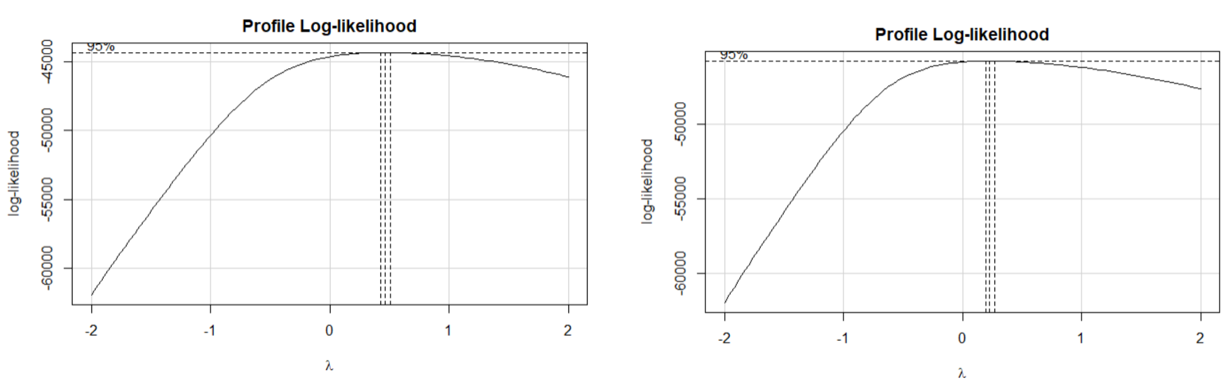
The first step was producing a pairs plot to visualize the variables of interest and the relationships between them. This plot is shown in Figure 1. Several of the variables appear to be right skewed. When looking at the categorical variable called city types there is a noticeable difference in the amount of data collected from core cities in comparison to legacy cities. The boxplots for city types do not seem to show any signs of averages that differ significantly. Both debt holdings and cash holdings do not seem to have very strong correlations with education expenditures. When it comes to the relationship between cash holdings and direct total expenditures as well as the relationship between cash holdings and total revenue, there is a large difference between core and legacy cities. The core cities have much larger correlations in these two relationships in comparison to the legacy cities. The relationship with the strongest correlation is the one between direct total expenditures and total revenue. The relationships that both direct total expenditures and total revenue have with revenue generated by the city are also strong as well. These three variables will have to be investigated further to test for multicollinearity in the statistical analyses portion of this project. Revenue generated by the city appears to have a positive correlation with education expenditures. In contrast, revenue generated from utility seems to have a slightly negative correlation with education expenditures.

Figure 1: Pairs Plot

Visualizing Functional Relationships

The next part of exploring the data was determining if a box-cox transformation was necessary. Two box-cox plots were produced, one with the direct total expenditures and total revenue variables included and one without. Since those two variables may be removed later due to multicollinearity, I wanted to check if removing them made a difference in the box-cox transformation. Figure 2a shows the box-cox plot with the variables included and Figure 2b shows the plot without those variables. Figure 2a suggests a square root transformation of the response variable ($\lambda \approx 0.5$). While Figure 2b suggests that λ is approximately equal to 0.2, which suggests a natural log transformation if λ is rounded down to zero. Which transformation I use depends on the results of the tests for multicollinearity.

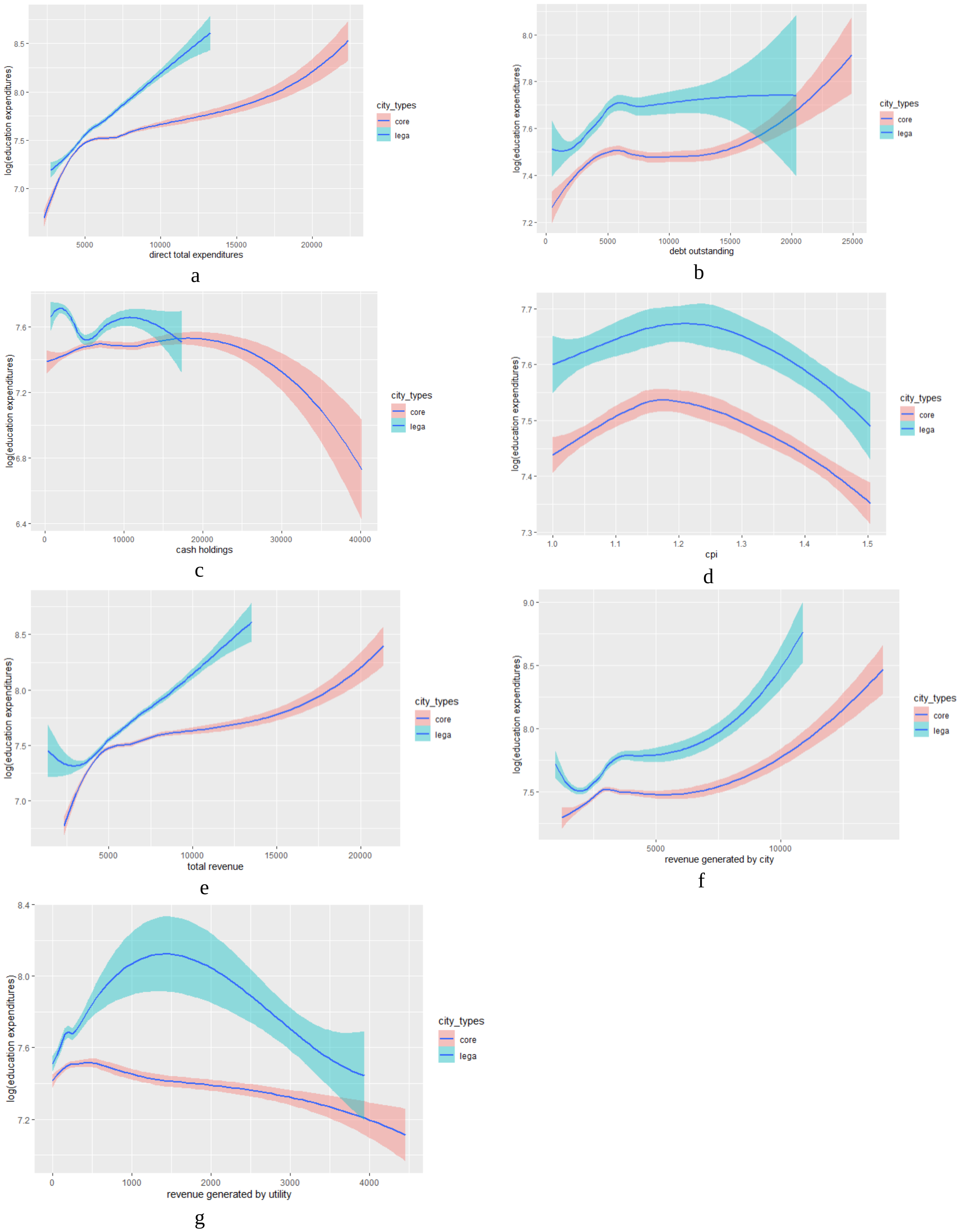
Figure 2: Box-Cox Transformation



a b

Finally, the relationships between the natural log of education expenditures and all the quantitative variables were plotted by city type. Lowess smoothing was utilized in Figure 3 to help identify possible functional relationships (Figure 3). For legacy cities Figure 3 a, b, and e visually appear to be linear while Figure 3 c, d, f, and g may be quadratic. For core cities Figure 3 a, b, e, and g show a possible linear relationship while Figure 3 c, d, and f show a possible quadratic relationship.

Figure 3: Functional relationship visualizations using lowess smoothing



Statistical Analyses

Detecting Multicollinearity

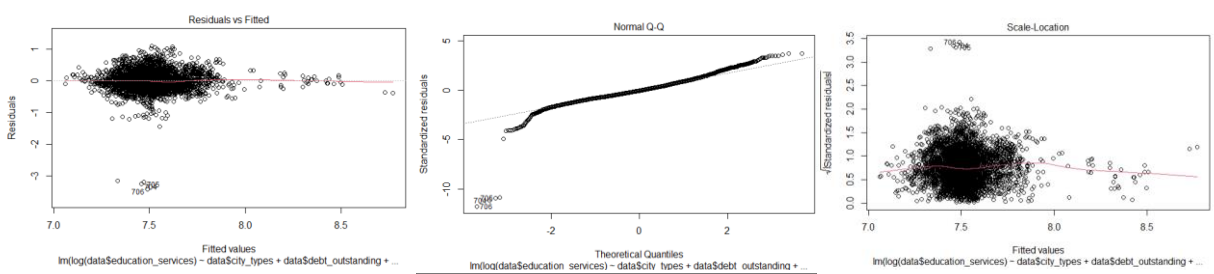
The statistical analyses portion of this project began with attempting to detect multicollinearity. Variance Inflation Factor (VIF) was used as a formal diagnostic for multicollinearity. VIF was calculated for each predictor variable (CPI, direct total expenditures, city types, revenue generated by the city, revenue generated from utility, total revenue, cash holdings, and outstanding debt). The max VIF was 13.22 and the mean VIF was 4.79 which indicates serious multicollinearity problems that may affect the coefficients and inflate standard errors in our model. The predictors with the largest VIF values were direct total expenditures (VIF = 13.16) and total revenue (VIF = 13.22). As previously mentioned in the exploratory data analysis, these two variables were very strongly correlated. After removing these two variables the max VIF was reduced to 2.32, the mean VIF was 1.52, and none of the variables had a VIF value above five. So, direct total expenditures and total revenue will not be included in the final regression model to avoid multicollinearity that may mislead inference.

Model Selection, Inference, and Interpretation

To determine the functional relationship of the final model an ANOVA F-test was used. A full quadratic model and a reduced linear model were fit. The results of the test were significant which suggests that the full quadratic regression model explains significantly more variability than the reduced model. The quadratic regression includes the qualitative variable named city types, and its' interactions with the predictor variables named CPI, revenue generated by the city, revenue generated from utility, cash holdings, and outstanding debt. I also decided to include interactions between certain quantitative variables such as the interactions of cash holdings and outstanding debt with revenue generated by the city and revenue generated from utility. These quantitative interactions were included to address the second research question outlined previously in the introduction. The response variable is the natural log of education expenditures. Using the step function in R, a model was chosen by AIC in a stepwise algorithm.

Figure 4 contains diagnostic plots that were used to evaluate the error and normality assumptions of the final regression model. The shape on the normal Q-Q plot looks relatively linear although the tails are slightly heavy (Figure 4). The residuals are centered around zero throughout the residual plot (Figure 4). The variance visually appears constant on the residual plot and scale-location plot (Figure 4). The covariance assumption also is satisfied since the residuals do not seem to be correlated (Figure 4).

Figure 4: Residuals, Normal Q-Q and Scale location plot of the final regression model



The next part of the analyses involves testing the entire model fit using the F-test. The null and alternative hypotheses are as follows…

$$H_0: All\ \beta_0 = \beta_1 = \cdots = \beta_i = 0$$

$$H_1: Not\ All\ \beta_0 = \beta_1 = \cdots = \beta_i = 0$$

$$i = 24$$

According to the F-test at an α-level of 0.05 I found that the model was significant (F-statistic: 39.81 on 23 and 4336 df, $p < 0.0001$). Since the null hypothesis is rejected, I can conclude that at least one of the coefficients does not equal zero and can proceed further and run inference on the parameter estimates.

To perform further inference on the estimates of each predictor variable in our final regression model, t-tests were performed at an α-level of 0.05.

The null and alternative hypotheses are as follows…

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

$$i = 0 \ldots 24$$

The parameter estimates of the regression are shown in the table below with the results of the t-tests and Bonferroni confidence intervals for each estimate (α = 0.05) (Table 1).

Table 1: Parameter estimates and Bonferroni Confidence Intervals of the final regression model

Coefficient	Estimate	Std. Error	t value	p-value	0.05%	99.95%
(Intercept)	4.33E+00	3.35E-01	12.922	< 0.0001	3.23E+00	5.43E+00
city types (legacy)	3.26E-01	3.24E-02	10.07	< 0.0001	2.19E-01	4.33E-01
outstanding debt	3.46E-05	6.07E-06	5.703	<0.0001	1.46E-05	5.46E-05
(outstanding debt)^2	-1.79E-09	3.30E-10	-5.41	< 0.0001	-2.87516E-09	-6.9958E-10
cash holdings	2.39E-05	6.43E-06	3.721	2.01E-04	2.75E-06	4.51E-05
(cash holdings)^2	-2.33E-10	2.94E-10	-0.792	4.28E-01	-1.20E-09	7.36E-10
CPI	4.78E+00	5.43E-01	8.789	< 0.0001	2.99E+00	6.56E+00
(CPI)^2	-1.97E+00	2.19E-01	-8.987	< 0.0001	-2.688723	-1.246779
revenue generated by the city	7.25E-05	1.72E-05	4.217	< 0.0001	1.59E-05	1.29E-04
(revenure generated by the city)^2	-3.09E-09	2.06E-09	-1.499	1.34E-01	-9.87E-09	3.69E-09
revenue generated from utility	-3.78E-05	4.61E-05	-0.82	4.12E-01	-1.90E-04	1.14E-04
(revenue generated from utility) ^2	-1.56E-08	8.61E-09	-1.818	6.91E-02	-4.40E-08	1.27E-08
city types(legacy):outstanding debt	1.61E-05	4.34E-06	3.697	2.21E-04	1.75E-06	3.04E-05
city types(legacy):cash holdings	-9.54E-05	1.07E-05	-8.917	< 0.0001	-1.31E-04	-6.02E-05
city types(legacy):(cash holdings)^2	4.45E-09	6.95E-10	6.405	< 0.0001	2.16E-09	6.74E-09
city types (legacy):(revenue generated by the city)^2	5.08E-09	1.15E-09	4.41	< 0.0001	1.29E-09	8.86E-09
city types (legacy):revenue generated from utility	7.38E-05	1.76E-05	4.191	< 0.0001	1.58E-05	1.32E-04
cash holdings: revenue generated by the city	-7.31E-09	1.51E-09	-4.847	< 0.0001	-1.23E-08	-2.35E-09
(cash holdings)^2:(revenue generated by the city)^2	9.72E-18	4.36E-18	2.231	2.57E-02	-4.62E-18	2.41E-17
(outstanding debt)^2:(revenue generated by the city)^2	2.21E-17	3.88E-18	5.703	< 0.0001	9.35E-18	3.49E-17
cash holdings : revenue generated from utility	2.98E-08	6.14E-09	4.856	< 0.0001	9.60E-09	5.00E-08
(cash holdings)^2:(revenue generated from utility)^2	-5.45E-16	1.29E-16	-4.233	< 0.0001	-9.68E-16	-1.21E-16
outstanding debt : revenue generated from utility	-2.97E-08	6.61E-09	-4.496	< 0.0001	-5.15E-08	-7.96E-09
(outstanding debt)^2:(revenue generated from utility)^2	6.43E-16	1.47E-16	4.364	< 0.0001	1.58E-16	1.13E-15

Several of the parameter estimates were found to be significant according to t-test (Table 1). CPI had a positive and significant association with the natural log of education expenditures ($p < 0.0001$). The natural log of education expenditures was positively associated with revenue generated by the city (p

<0.0001) and non-significantly, negatively associated with revenue generated from utility ($p = 0.412$). Cash holdings was found to have a negative quantitative interaction with revenue generated by the city ($p < 0.0001$) and a positive quantitative interaction with revenue generated from utility ($p < 0.0001$). While outstanding debt had a negative interaction with revenue generated from utility ($p < 0.0001$). There was also a significant difference in the natural log of education expenditures in dollars per capita between legacy and core cities ($p < 0.0001$).

<u>Conclusions</u>

The entire model was found to be significant according to an ANOVA F-test ($p < 0.0001$). To address the first research question, the parameter estimate for revenue generated by the city is positive suggesting that as this type of revenue increases, the natural log of education expenditures increases as well. However, the parameter estimate for revenue generated from utility is negative but the p-value for this estimate shows that it is not statistically significant. So, there is not sufficient evidence that the parameter estimate for revenue generated from utility is different from zero. These results may indicate that funding for education usually comes from revenue generated by the city itself.

For the second research question we to look to the interactions of cash holdings and outstanding debt with revenue generated by the city and revenue generated from utility. The parameter estimate for the interaction between cash holdings and revenue generated by the city indicate that a unit increase in cash holdings results in a greater decrease in the natural log of education expenditures for larger values of revenue generated by the city. However, a unit increase in cash holdings results in a greater increase in the natural log of education expenditures for larger values of revenue generated from utility. The nature of these two interactions is interesting when compared to our previous results where it was revenue generated by the city that had a positive association with the natural log of education expenditures and revenue generated from utility that had a negative one. The negative interaction between outstanding debts and revenue generated from utility indicate that a unit increase in outstanding debt results in a greater decrease in the natural log of education expenditures for larger values of revenue generated from utility.

Lastly, for the final research question, the parameter estimate for city types was found to be significant. This indicates that the difference in the natural log of education expenditures between legacy and core cities is significant. The parameter estimate tells us that the average difference in the natural log of education expenditures between legacy and core cities is 0.3260.

There were some limitations in the analysis. Firstly, most of the data was collected from core cities so the inferences based on the data for legacy cities may not be as accurate considering the smaller sample size. Also, only a few quantitative interactions were included in the final regression model based on my limited domain knowledge on the subject and the research questions that were formulated. There may be other significant quantitative interactions that were not covered in this project. Future work may include similar analyses of other continents outside of Europe such as North America. Also, other information on the citizens of these cities such as demographics and average income would be interesting to study further in relation to this issue. Interactions including those types of predictors and revenue may prove to be significant.

References

1. Kutner, Michael H., et al. *Applied Linear Regression Models*. 4th ed., McGraw-Hill Education/Asia, 2018.
2. “Why Education Matters to Health: Exploring the Causes.” *Center on Society and Health*, 13 Feb. 2015, https://societyhealth.vcu.edu/work/the-projects/why-education-matters-to-health-exploring-the-causes.html.

Appendix

Annotated Code

```r
#read in data
data1 <- read.csv("dae_23.csv")
head(data1)

#subset data with variables of interest
data <- data.frame(data1[,c("city_types", "cpi", "rev_total", "own_source_rev",
                            "rev_utility", "spending_direct",
                            "education_services", "debt_outstanding",
                            "cash_holdings")])

#summarize variables of interest
summary(data)


# pairs plot with log transformation
plot_frame1 <- data.frame("log(Education Expenditures)" = log(data$education_services),
                          "direct total expenditures" = data$spending_direct,
                          "outstanding debt" = data$debt_outstanding,
                          "cash holdings" = data$cash_holdings,
                          "CPI" = data$cpi,
                          "total revenue" = data$rev_total,
                          "Rev Gen by City" = data$own_source_rev,
                          "Rev Gen from Utility" = data$rev_utility,
                          "City Types" = data$city_types)
ggpairs(plot_frame1, aes(color=data$city_types, alpha=100))
```

```r
#functional relationship lowess smooths plots

plot1 <- (ggplot(data, aes(x= spending_direct, y=log(education_services) ), color=city_types) +
  #geom_point() +
  geom_smooth(method=loess, aes(fill=city_types)) +
 labs(x="direct total expenditures", y="log(education expenditures)"))
plot1
```

```r
plot2 <- (ggplot(data, aes(x=debt_outstanding, y=log(education_services) ), color=city_types) +
  #geom_point() +
  geom_smooth(method=loess, aes(fill=city_types)) +
 labs(x="debt outstanding", y="log(education expenditures)"))
plot2
```

```r
plot3 <- (ggplot(data, aes(x= cash_holdings, y=log(education_services) ), color=city_types) +
  #geom_point() +
  geom_smooth(method=loess, aes(fill=city_types)) +
 labs(x="cash holdings", y="log(education expenditures)"))
plot3
```

```r
plot4 <- (ggplot(data, aes(x=cpi, y=log(education_services) ), color=city_types) +
  #geom_point() +
  geom_smooth(method=loess, aes(fill=city_types)) +
 labs(x="cpi", y="log(education expenditures)"))
plot4
```

```r
plot5 <- (ggplot(data, aes(x= rev_total, y=log(education_services) ), color=city_types) +
  #geom_point() +
  geom_smooth(method=loess, aes(fill=city_types)) +
 labs(x="total revenue", y="log(education expenditures)"))
plot5
```

```r
plot6 <- (ggplot(data, aes(x= own_source_rev, y=log(education_services) ), color=city_types) +
  #geom_point() +
  geom_smooth(method=loess, aes(fill=city_types)) +
 labs(x="revenue generated by city", y="log(education expenditures)"))
plot6
```

```r
plot7 <- (ggplot(data, aes(x= rev_utility, y=log(education_services) ), color=city_types) +
  #geom_point() +
  geom_smooth(method=loess, aes(fill=city_types)) +
 labs(x="revenue generated by utility", y="log(education expenditures)"))
plot7
```

```r
#box-cox transformation

#with correlated variables (direct total expenditures and total revenue)
options(repr.plot.width=5, repr.plot.height=5)
boxCox(lm(education_services ~ cpi + own_source_rev + rev_utility + debt_outstanding + cash_holdings
          + spending_direct + rev_total+ city_types, data=data))
```

```r
#without correlated variables (direct total expenditures and total revenue)
options(repr.plot.width=5, repr.plot.height=5)
boxCox(lm(education_services ~ cpi + own_source_rev + rev_utility + debt_outstanding + cash_holdings
          + city_types, data=data))
```

```r
#VIF with direct total expenditures and total revenue
fit_mc <- lm(log(education_services) ~ spending_direct +
                                        own_source_rev +
                                        rev_total +
                                        cash_holdings +
                                        rev_utility +
                                        cpi +
                                        debt_outstanding +
                                        city_types,
                                        data = data)

print(vif(fit_mc))
print(paste("max VIF:", max(vif(fit_mc))))

print(paste("mean VIF:", mean(vif(fit_mc))))
```

```r
#VIF without direct total expenditures and total revenue
fit_mc <- lm(log(education_services) ~ #spending_direct +
                                        own_source_rev +
                                        #rev_total +
                                        cash_holdings +
                                        rev_utility +
                                        cpi +
                                        debt_outstanding +
                                        city_types,
                                        data = data)

print(vif(fit_mc))
print(paste("max VIF:", max(vif(fit_mc))))

print(paste("mean VIF:", mean(vif(fit_mc))))

#fit full quadratic model and reduced linear model
fit <- lm(log(data$education_services) ~ data$city_types +
                    data$debt_outstanding +
                    I(data$debt_outstanding^2) +
                    data$city_types:data$debt_outstanding +
                    data$city_types:I(data$debt_outstanding^2) +

                    data$cash_holdings +
                    I(data$cash_holdings^2)+
                    data$city_types:data$cash_holdings +
                    data$city_types:I(data$cash_holdings^2)+

                    data$cpi +
                    I(data$cpi^2) +
                    data$city_types:data$cpi +
                    data$city_types:I(data$cpi^2) +

                    data$own_source_rev +
                    I(data$own_source_rev^2) +
                    data$city_types:data$own_source_rev +
                    data$city_types:I(data$own_source_rev^2) +

                    data$rev_utility +
                    I(data$rev_utility^2)+
                    data$city_types:data$rev_utility +
                    data$city_types:I(data$rev_utility^2)+


                    data$cash_holdings:data$own_source_rev +
                    I(data$cash_holdings^2):I(data$own_source_rev^2) +

                     data$debt_outstanding:data$own_source_rev +
                    I(data$debt_outstanding^2):I(data$own_source_rev^2) +


                     data$cash_holdings:data$rev_utility +
                    I(data$cash_holdings^2):I(data$rev_utility^2) +
```

```r
                     data$debt_outstanding:data$rev_utility +
                    I(data$debt_outstanding^2):I(data$rev_utility^2)

                    )

fit_reduced <- lm(log(data$education_services) ~ data$city_types +
                    data$debt_outstanding +
                    data$city_types:data$debt_outstanding +

                    data$cash_holdings +
                    data$city_types:data$cash_holdings +

                    data$cpi +
                    data$city_types:data$cpi +

                    data$own_source_rev +
                    data$city_types:data$own_source_rev +

                    data$rev_utility +
                    data$city_types:data$rev_utility +


                    data$cash_holdings:data$own_source_rev +
                    data$debt_outstanding:data$own_source_rev +

                    data$cash_holdings:data$rev_utility +
                    data$debt_outstanding:data$rev_utility
              )

#determine if quadratic or linear by F-test
anova(fit_reduced, fit)

#select model by AIC
step_fit <- step(fit)

#check if assumptions are true
plot(step_fit, which=c(1,2,3))
```

```r
#F-test on entire model and t-tests on parameter coefficients
summary(step_fit)

#Bonferoni confidence intervals (alpha = 0.05)
g = 24
x <- 0.05 / (2*g)
x
#x = 0.001
confint(step_fit, level = (1-0.001))
```

Detecting Multicollinearity VIF values were used as a formal diagnostic of multicollinearity. I assessed multicollinearity using the VIF values of a fit without the interactions and powers so I could detect multicollinearity between my main variables. I later proceeded with the quadratic model that included the powers and interactions.

```
##  spending_direct   own_source_rev        rev_total    cash_holdings 
##        13.162825         3.655520        13.222347         2.243873 
##      rev_utility              cpi debt_outstanding       city_types 
##         1.466576         1.035452         2.407234         1.154549 

## [1] "max VIF: 13.2223469793017"

## [1] "mean VIF: 4.7935472081355"

##   own_source_rev    cash_holdings      rev_utility              cpi 
##         1.608374         1.969932         1.115484         1.030599 
## debt_outstanding       city_types 
##         2.319078         1.086975 

## [1] "max VIF: 2.319077761495"

## [1] "mean VIF: 1.52174004827819"
```

Reduced vs Full Model An ANOVA F-test was used to determine if a quadratic or linear functional relationship was necessary. A reduced linear model and a full quadratic model were fit. The results of the F-tests were significant which suggests that the quadratic functional relationship was necessary.

```
## Analysis of Variance Table
## 
## Model 1: log(data$education_services) ~ data$city_types + data$debt_outstanding + 
##     data$city_types:data$debt_outstanding + data$cash_holdings + 
##     data$city_types:data$cash_holdings + data$cpi + data$city_types:data$cpi + 
##     data$own_source_rev + data$city_types:data$own_source_rev + 
##     data$rev_utility + data$city_types:data$rev_utility + data$cash_holdings:data$own_source_rev + 
##     data$debt_outstanding:data$own_source_rev + data$cash_holdings:data$rev_utility + 
##     data$debt_outstanding:data$rev_utility
## Model 2: log(data$education_services) ~ data$city_types + data$debt_outstanding + 
##     I(data$debt_outstanding^2) + data$city_types:data$debt_outstanding + 
##     data$city_types:I(data$debt_outstanding^2) + data$cash_holdings + 
##     I(data$cash_holdings^2) + data$city_types:data$cash_holdings + 
##     data$city_types:I(data$cash_holdings^2) + data$cpi + I(data$cpi^2) + 
##     data$city_types:data$cpi + data$city_types:I(data$cpi^2) + 
##     data$own_source_rev + I(data$own_source_rev^2) + data$city_types:data$own_source_rev + 
##     data$city_types:I(data$own_source_rev^2) + data$rev_utility + 
##     I(data$rev_utility^2) + data$city_types:data$rev_utility + 
##     data$city_types:I(data$rev_utility^2) + data$cash_holdings:data$own_source_rev + 
##     I(data$cash_holdings^2):I(data$own_source_rev^2) + data$debt_outstanding:data$own_source_rev + 
##     I(data$debt_outstanding^2):I(data$own_source_rev^2) + data$cash_holdings:data$rev_utility + 
##     I(data$cash_holdings^2):I(data$rev_utility^2) + data$debt_outstanding:data$rev_utility + 
##     I(data$debt_outstanding^2):I(data$rev_utility^2)
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)    
## 1   4344 390.28                                 
## 2   4330 370.95 14    19.334 16.12 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Description of Variables

year

id_city

city_population

city_types: Core or legacy city

cpi: consumer price index

rev_total: total revenue

own_source_rev: revenue generated by the city itself

rev_utility: revenue generated from utility

spending_direct: direct total expenditures

education_services: education expenditures

social_services: social services expenditures

transportation: transportation expenditures

public_safety: public_safety expenditures

envir_housing: environment and housing expenditures

spending_utility: utility expenditures

sp_empl_retire_tr: employee retirement trust expenditures

debt_outstanding: Outstanding debt

cash_holdings: Cash holdings