What Does the Financial Crisis Tell Us About the Determinants of Municipal Bonds Yields?

Su Huang*

Ph.D. Program in Economics
City University of New York — Graduate Center

January 2, 2012

Abstract

This paper performs a comprehensive analysis on the determinants of general obligation municipal bond yields and examines the impact of the recent financial crisis on the underlying relations. Three classes of variables affect yields of municipal bonds: the economic status of the state where the bond is issued, the demographic characteristics of the state, and the financial status of the state or the local government where the bond is issued. This paper employs several new econometric techniques, including nonparametric regressions and Gradient Boosting, to capture both nonlinearity and potential random interactions among key determinants. A systematic comparison of the relations before and after the 2008 financial crisis shows that the economic and financial health of local governments has become markedly more diverse since the crisis began. The relation between the municipal bond yields and the economic and financial health of the local governments has also become stronger because of the larger differentiation among the local governments and hence larger signal to noise ratio, as well as closer scrutiny by market participants on the economic fundamentals of municipal governments. The analysis also shows that accommodating nonlinearities and random interaction effects can significantly enhance the predictive performance on the municipal bond yields.

JEL classification: C13 C14 G12

Key words: Municipal bond market, determinants of bond yields, bond yield prediction, Local Polynomial regression, Gradient Boosting method

*Address: Su Huang, PhD Program in Economics, CUNY Graduate Center, 365 Fifth Avenue, New York, NY 10016-4309, Tel: 212-817-8176, email: huangsuustcgc@gmail.com. I thank Ram Ahluwalia, Michael Grossman, Jun Jiang, Catherine Lau, Sangeeta Pratap, Thom Thurston, Xu Wang, Wei Zhang, and all participants in the City University of New York Graduate Center Applied Economics seminar for comments. I especially appreciate my academic adviser Liuren Wu's kind help all the time. I continue to welcome any comment.

1 Introduction

A municipal bond is a bond issued by a local government, or its agencies. Since World War II, both the size and the importance of the municipal bond market have increased dramatically. Several studies examine the determinants of municipal bond yields at issuance, e.g., Kessel (1971) and Grossman et al. (1993); however, less research analyzes municipal bond yields in the secondary market. Due to the liquidity of the bond and the data availability of factors affecting bond yields, most research focuses on a few variables and a small set of bonds, e.g., Badu et al. (1996), Rubinfeld (1973) and Michel (1977). With different sample periods, independent variables and econometric methods, researchers often reach different conclusions. After Hastie (1972), there is no systematic analysis about determinants of municipal bond yields by using all states' General Obligation municipal bonds.

Investors tend to believe that local governments are unlikely to default, as they can raise taxes to pay off debts. Most municipal bonds are rated as AAA, and even under the same rating class, the historic default rates of municipal bonds are much smaller than that of the corporate bonds (The Committee on Financial Services of the 110th Congress (2008)). As a result, the cross sectional variation of municipal bond yields is much smaller than that of corporate bond yields, and much of the variation is caused by factors other than credit concerns, such as tax and liquidity, which are common factors early papers consider, e.g., Grossman et al. (1993).

Since the 2008 financial crisis, many local governments have been facing financial problems. Default has been becoming a real concern. In May 2008, Vallejo, California, which had a population of about 117,000, declared bankruptcy protection. It is by far the largest municipal bankruptcy in the United States history. In October and November 2011, Harrisburg, Pennsylvania and Jefferson County, Alabama were planning to file for bankruptcy protections after long time struggles seperately. On August 5, 2011, Standard & Poors downgraded the

federal governments of the United States from AAA to AA+, reflecting increased credit risk concern even on the agency that had traditionally regarded as the most credit worthy and virtually risk free. The cross-sectional bond yields have become much larger recently. The economic and financial health of local governments has become markedly more diverse since the crisis began. The data now are ready to identify the key determinants of municipal bond yields and see how these determinants affect bond yields more accurately.

This paper first identifies the key determinants of municipal bond yields and classifies them into three classes: the economic status of the state where the bond is issued; the demographic characteristics of the state; and the financial status of the state or the local government where the bond is issued. I collect the data of municipal bond yields and these determinants from April 2006 to March 2010, which covers both normal periods and the 2008 financial crisis, to examine the impact of the recent financial crisis. The relation between the municipal bond yields and the economic and financial health of the local governments has become stronger because of the larger differentiation among the local governments and hence larger signal to noise ratio, as well as closer scrutiny by market participants on the economic fundamentals of municipal governments. In the analysis, univariate linear regressions are employed first to understand the direction of the dependence of each variable. Then, two variable linear regressions with interaction terms are discussed to investigate possible interaction effects. Finally, multiple variable regressions are summarized to confirm the findings in the univariate and bivariate regressions.

The linear regression model has its own shortcomings in the municipal bond yield analysis: First, effects of most determinants are nonlinear. For example, investors care about a community's debt ratios only when these ratios are very high. The effect of debt ratios is very small generally, but becomes much larger in bad financial times. Second, interaction effects between variables are generally not linear, and might change across time. In different

time periods, the interaction terms that take effect are even different. These factors are very hard to control in linear regression models. Third, some variables are highly correlated. In linear regression models, putting highly correlated variables in one regression might cause one coefficient to be extremly high, while another one to be extremly low; while just putting one of them will lose some useful information. Therefore, besides the linear regression, I use the Local Polynomial regression to capture possible nonlinear effects. In addition, I successfully use the Gradient Boosting method, which was first introduced by Friedman (1999) and Friedman (2001), to capture the potential random interaction effects among key determinants. These methods significantly enhance the predictive performance on the municipal bond yields.

This paper is organized as follows. Section 2 summarizes the literature by classifying determinants of municipal bond yields. Section 3 describes the data and summary statistics. Section 4 illustrates the determinants by the linear regression method. Section 5 investigates nonlinear effects of determinants and discusses the shortcomings of linear regression models. Section 6 provides the predictive method based on the Gradient Boosting framework. Section 7 concludes and gives further discussions. Classifications of municipal bonds, methodologies of the Local Polynomial regression and the Gradient Boosting method, tables, and figures are in the appendix.

2 Classify the determinants of municipal bond yields

The determinants of municipal bonds under investigation are classfied into three classes: economic variables, demographic variables, and financial accounting variables.

2.1 Economic variables

Economic variables play very important roles in municipal bonds' default rates. The better the economic status, the higher potential revenues the bond issuer can collect, and the lower chance that the bond would default. Most previous papers use at least one economic variable in their analysis.

GDP or income is a measure of an economy's overall output. This is the most important macroeconomic variable to measure the health of the economy. Badu et al. (1996), Horton (1969), Lipnick et al. (1999), Loviscek and Crowley (1990), Palumbo et al. (2006), Poterba and Rueben (1997), and Rubinfeld (1973) all use this measure. GDP itself is not a stationary variable, therefore I use the growth rate of GDP per capita as one of my predictors. Since state GDP is updated just yearly, as Boyd and Dadayan (2010) do, I also include the state's coincident index (COI), which is reported by the Federal Reserve of Philadelphia monthly, in the variable list. A state COI provides information of current economic activity in an individual state.

Like Badu et al. (1996), Lipnick et al. (1999), Palumbo et al. (2006), and Poterba and Rueben (1997), I include the unemployment rate. The unemployment rate is not only an economic variable, but also closely related to a state or county's budget. When a state experiences a high unemployment rate, it not only produces less potential goods and services and collects less tax revenues, but also has to spend more money on social welfare, which might cause financial problems. The unemployment rate also reflects the economic structure of that state. In addition, the correlations between the unemployment rate and most other variables are small, which shows that the unemployment rate contains a great deal of unique information. Finally, states update unemployment rates monthly, while most other variables are released just quarterly or yearly. Therefore, the unemployment rate can reflect very up-to-date information. Because of debates about measurement problems of the unemployment

rate, I also include the non-farm payroll employment rate in my variable list, which contains similar information to the unemployment rate and still is released monthly.

While GDP and COI are coincident indicators of the economy and the unemployment rate is a lagging indicator, we need one leading indicator of the economy as well. Since the number of building permits of each state is released every month and contains very accurate information, I use the building permits per capita as the measure of the economic leading variable. And also I include existing home sales per person to contain information of the health of the housing market, which plays a very important role for a state's fiscal status.

As Hastie (1972) states, under a very bad national or local economic condition, "municipalities that are economically well diversified will be less susceptible to declining property values and incomes than will less diversified communities; therefore, they may also be less susceptible to default". Therefore, economic diversification variables are largely used in literature, e.g., Badu et al. (1996), Hastie (1972), Lipnick et al. (1999), and Loviscek and Crowley (1990). Similar to Loviscek and Crowley (1990), I generate an index called the coefficient of specialization (CDIV) to measure economic diversification for each state. Compared to its use of the number of workers employed in each industry, I use GDP of each industry, which contains more accurate information. The state CDIV is defined as $\frac{1}{2}\sum_{i=1}^{N}|\frac{e_{ic}}{E_{ic}} - \frac{e_{in}}{E_n}|$, where e_{ic} is the GDP in industry i in area c, e_{in} is the GDP in industry i in the nation, E_c is the GDP in area c, E_n is the GDP in the nation, and N is the number of GDP components. As the state becomes more (less) diversified, the coefficient approaches 0 (1).

2.2 Demographic variables

Although demographic variables do not affect municipal bonds' default rates directly, they contain some useful information. Advantages of demographic variables are that they are

easier to access and more stable than economic and financial variables, and there are big differences across states.

The most used demographic variable in the literature is population, e.g., Badu et al. (1996), Carleton and Lerner (1969), Hastie (1972) and Horton (1969). States with big populations have more economic diversification and higher potential resources of revenues than states with small populations. However, states with big populations have large debt burdens as well. Therefore, the direction of the effect of population is not clear. In addition, since I don't have exact or even approximate numbers of states' taxable property values, I have to include some demographic variables relevant to the housing information. Because the data is available, I include persons per house and homeownership rates in the analysis.

2.3 Financial accounting variables

While economic variables predict communities in the future, current financial accounts of communities contain a great deal of information about communities' current financial status. Debts, revenues, expenditures and abilities to collect taxes are the most important of these variables.

If investors think of a community as a company, the first thing they need to consider before buying a bond is the leverage ratio. Although the amount of total liabilities is available for all communities, the amount of total assets for a community is much harder to define than that for a company. We need a variable to measure the relative debt burden. First, I use a community's taxable assets, which is largely used in the literature, e.g., Carleton and Lerner (1969), Hastie (1972), Horton (1969), Michel (1977), and Rubinfeld (1973). Theoretically speaking, a general obligation municipal bond issued by a community will not default as it can always collect taxes to pay off debts. The ability to collect taxes is determined by the

amount of its taxable assets. However, due to data availability, I use total capital assets as the alternative. Second, like Poterba and Rueben (1997), I use GDP per capita, which includes other potential resources besides capital assets to refund the debt. Third, similar to Badu et al. (1996) and Michel (1977), I use revenues, which measure the community's ability to collect money right now. Finally, I use the fund's own assets, which measure its current financial status. While the first two approximations of assets measure the potential resources for a community's refunding, the last two measure a community's current refunding ability.

Expenditures and revenues are two other important financial variables that early studies consider, e.g., Badu et al. (1996), Michel (1977), Poterba and Rueben (1997), Lipnick et al. (1999) and Palumbo et al. (2006). The bigger the government, the more money it needs to maintain its daily operation. The financial status of a government is determined by a government's ability to collect revenues as well. I define two deficit variables: the absolute deficit and the relative deficit ratio. In addition, as Lipnick et al. (1999) states, I also calculate the general fund balance as a percent of revenues and the general fund balance as a percent of total assets to measure a government's financial strength. "These ratios provide measures of the financial reserves potentially available to fund unforeseen contingencies" (Lipnick et al. (1999)).

3 Data descriptions and summary statistics

3.1 Data resources

As discussed in Appendix A, there are different types of municipal bonds traded in the market. To avoid possible different effects of determinants in different types, I only include the most typical ones: General Obligation, Non-Refunded, Tax Free, Fixed or Fixed OID Coupon, Not Bank Qualified, Non-Callable and Non-Sinking Fund bonds. I have downloaded

all Bloomberg terminal available bid and ask quotes of municipal bond yields in the secondary market from April 2006 to March 2010 for this specific type. To increase the accuracy of data, I drop the intra dealer transactions and the bonds whose time to maturities are less than one month or more than twenty years. Since the municipal bond is not traded as frequently as stocks, I sample the data monthly. If a security is quoted multiple times in a month, I take the weighted average yield based on the volume of each quote. In addition, I drop those securities whose total volume is less than 25,000 in that month. Finally, I drop the bonds issued by the District of Columbia and the Commonwealth of Puerto Rico.

Due to the data availability, all economic and demographic variables, and parts of financial variables, I only have state level data. For these variables, I use their state level information. The economic and financial status of a state will affect all local communities in that state. For example, in the process of rescuing Harrisburg, Pennsylvania in 2011, the state government of Pennsylvania played a very large role. When Jefferson County, Alabama was planning to file the bankruptcy, the state's governor said: "Bankruptcy will negatively impact not only the Birmingham Region, but also the entire state."[1] For other parts of financial variables, I use their community level information because they are available.

I have collected each state's yearly GDP and industry GDP from the Bureau of Economic Analysis; each state's monthly building permits, non-farm payroll employment, unemployment rate and 12-month moving average existing home sales from the Bureau of Labor Statistics, each state's monthly coincident index from the Federal Reserve Bank of Philadelphia, and each state's total population, persons per house and homeownership rate from the Bloomberg terminal.

For financial accounting variables, I have collected each state's total liabilities, total rev-

[1] Wall Street Journal A6, November 10, 2011

enues, total capital assets, total expenses, government fund total liabilities and government fund total revenues. For each local bond issuer, I have collected its total liabilities, total revenues, total assets, total operating expenses and total fund balance. All financial accounting data of each year from 2005 to 2008 have been downloaded from the Bloomberg terminal.

In each month from April 2006 to March 2010, I merge the determinants into the bond yields of their corresponding municipalities or states, and do cross-sectional analysis later. I assume that the market observes the monthly released data one month later, quarterly and yearly released data one quarter later.

3.2 Mature-adjusted bond yields

Figure 1 plots the correlations between the bond yield and time to maturity for both bid and ask quotes from April 2006 to March 2010. The correlations are very large in most months. The only exception is from late 2006 to early 2007, when the term structures in the treasury market were unusual as well. In addition, the correlations from bid quotes are consistently higher than those from ask quotes. The reason is that lower ask quotes result in higher bond yields, which causes larger observational errors of bond yields than bid quotes do.

We need to exclude the effects of bid-ask quote difference and term to maturity, and focus on the cross state or cross county analysis.[2] For bid or ask quotes, in each month, I first run a Local Polynomial regression of the bond yields on time to maturities by using Epanechnikov kernel with degree one.[3] To illustrate why I choose the Local Polynomial regression instead of the linear regression, Figure 2 plots the R-squared of both the Local Polynomial

[2] I acknowledge that this might affect the standard errors of coefficients in later regressions. But the term to maturity or bid-ask quote difference is uncorrelated with variables of determinants both therotically and empirically (the largest absolute value of correlations is less than 0.1), which minimizes the bias. Taking out these two factors will make my results clearer, especially in the nonlinear analysis.

[3] The methodology of the Local Polynomial regression is discussed in detail in Appendix B.1

regressions and the linear regressions from Apr 2006 to Mar 2010 by including ask and bid quotes together. It shows that the Local Polynomial regression is slightly better, especially between late 2006 and early 2007 when the term structure is not clear. Another interesting thing is that R-squared are higher in the financial crisis period than those before the financial crisis.

At each month, for both ask and bid Local Polynomial regressions, I predict the bond yield based on time to maturity and calculate the difference between the observed and the predicted yield of all securities. Then for each security, I take the average of differences in ask and bid regression predictions and use this average as the number of the dependent variable—the mature-adjusted municipal bond yield. This average can only be explained by communities' own characteristics. To avoid bound problems of Local Polynomial regression methods, I drop securities whose time to maturities are less than two months or more than fifteen years. All municipal bond yields mentioned later in this paper are actually their corresponding mature-adjusted bond yields.

3.3 Summary statistics

Table 1 lists descriptions, sample frequencies and notations of determinants of municipal bond yields investigated in this paper.

3.3.1 Bond yields

We have 4,645 municipal bonds in average each month, with the minimum 3,737 bonds in December 2007 and the maximum 6,250 bonds in October 2008. Figure 3 plots the standard deviations of the mature-adjusted municipal bond yields each month from April 2006 to March 2010. Not surprisingly, the cross community difference of bond yields becomes dramatically large during the 2008 financial crisis. Huge jumps happen in September and

October 2008, which are widely thought as the beginning of the financial crisis. The larger cross sectional bond yield difference will make our analysis more reliable and convincing.

3.3.2 State economic variables

From Figure 5, we observe that all variables except economic diversification index (CDIV) reflect the financial crisis's effects in 2008. The coincident index moves more smoothly than GDP per capita growth rates because it is reported more frequently. In addition, during the 2008 financial crisis, although high COI states decay earlier than low COI states, low COI states decay more dramatically than high COI states. Furthermore, most states' 12-month moving average existing home sales per person decrease during 2006 to 2008, but bounce back a little bit in 2009 except for those states already having sold a great deal of homes. However, the cross state difference does not change greatly. The building permits per person decrease from 2006 to 2009 in most states and the cross state difference becomes smaller.

Panel F of Figure 5 clearly shows two important effects of the financial crisis. First, unemployment rates increase across all states. Second, unemployment rates increase more for those states that already have experienced high unemployment rates than those who have experienced low ones. The second effect increases the cross state difference of unemployment rates during the financial crisis period, which is a very important phenomenon in later analysis. The non-farm payroll employment rate tells a similar story to the unemployment rate but has a little bit flatter time series difference than the unemployment rate has.

Panel G of Figure 5 shows that the difference between the state's economic diversification index becomes slightly narrower from 2005 to 2008, but the time series difference is still much smaller than the cross state difference. Since that the industry GDP is released well behind the total GDP, I use the one year lag of economic diversification index (CDIV) instead of the current one in the analysis. From the graph, we see that the state CDIV does not change

from year to year across different years.

3.3.3 State demographic variables

Figure 4 summarizes the state's demographic information under investigated from late 2005 to late 2009. We clearly observe that unlike economic and financial variables, demographic variables do not change over years or across states. These variables are very stable predictors in the analysis.

3.3.4 State financial accounting variables

Table 2 summarizes means, medians, standard deviations and skewness of all state level financial accounting variables in each year from 2005 to 2008. Based on the previous analysis and data availability, four variables are included: total liabilities over total capital assets, total liabilities over GDP per capita, total liabilities over total revenues, and government fund total liabilities over total revenues. For all years, the state of Wyoming is dropped. In 2006 and 2008, the state of South Dakota is dropped as well. The reasons are that there are no qualified municipal bonds under my criteria discussed before. But these two states have the lowest and the third lowest populations in the United States, which should not affect my analysis later.

Table 2 shows three important findings: First, all debt ratio variables are generally stable from 2005 to 2008 if measured by means and medians. Second, both the measure of skewness and the phenomenon that means are consistently higher than medians show that all variables are right skewed. Generally speaking, most states are in good financial status, but there are a few exceptions. Finally, the effect of the 2008 financial crisis is not reflected by these variables.

3.3.5 Community financial accounting variables

Table 3 summarizes the community level financial accounting variables in each year from 2005 to 2008. Panels A and B show that leverage ratios, no matter whether measured by total liabilities over total assets or total liabilities over total revenues, are stable from year to year. The measure of total liabilities over total revenues is largely right skewed, while the measure of total liabilities over total assets is more symmetric. Those outlier communities will play important roles in the later studies, especially in the nonlinear analysis.

Panels C and D show that more than half the communities run surpluses from 2005 to 2008. In addition, in 2008, because of the financial crisis, both communities' absolute and relative deficits increase on average. Furthermore, the absolute deficit is largely left skewed and the cross community difference is very large, while the relative deficit ratio is largely right skewed and the cross community difference is relatively small. These findings probably come from the definitions of these two variables. There are a few communities having either large deficits or large surpluses, while surpluses are even larger in absolute values. But when measured by the deficit ratio, most communities are in good status, while a few of them are in very bad condition.

Panels E and F show that communities' total fund balance over both total assets and total revenues are relatively flat these years, but the average of total fund balance to total revenues reduces a little bit in 2008 because of the financial crisis. In addition, the ratios of total fund balance to total revenues are highly right skewed from 2005 to 2007, which states that very few communities run large relative surpluses. But this skewness becomes much smaller in 2008, which shows that the number of large surplus communities decreases a lot in 2008. Finally, the ratios of total fund balance to total assets are more symmetric than other variables.

In summary, Table 2 and Table 3 tell us that most variables do not reflect the financial

crisis effects immediately in 2008, except the community's absolute and relative deficits and the total fund balance as a percentage of total revenues.

4 Illustrations of key determinants by linear regressions

In this section, we are going to investigate how these determinants discussed in section 2 can affect municipal bond yields. Simple linear regression models are applied in this section.

4.1 Univariate analysis

I first do univariate analysis to identify and understand the direction of the key determinants of municipal bond yields. The advantage of univariate regressions is that we do not have the potential multicollinearity problem. For each variable, I first pool all monthly samples together and then compare regression results in different months.

4.1.1 Pooling all months together

Table 4 reports the pooled linear regression result of each variable. To make the analysis comparable, I standardize all variables first: For each observation of each variable, the variable's mean is substracted and the result is divided by the variable's standard deviation to calculate the z-score. I use all z-scores as new independent variables. We observe that most variables have expected signs and are significant.

For economic variables, the coincident index (COI) plays larger roles than the growth rate of GDP per capita. The possible reason is that COI has more updated information. Both the non-farm payroll employment rate and the unemployment rate have large effects on munic-

ipal bond yields. But existing home sales per person, building permits per person and the economic diversification index don't have the expected signs.

For demographic variables, population has positive and significant effects on municipal bond yields. Based on my earlier analysis, the effects of population are mixed. The pooled linear regression result shows that the deficit problem dominates. In recent years, we notice that states having financial deficit problems are generally big population states, such as Florida, New York and California. In addition, the homeownership rate and persons per house have small but significant effects on bond yields, but the sign of the homeownership rate is not expected.

For state debt ratio variables, all variables except the amount of total liabilities over total revenues have the expected signs. But the estimation coefficients of state debt ratio variables are much smaller than those of most state economic and community financial variables. A very interesting result is that if we measure a state's total liabilities over total revenues by a state's parent government fund, we obtain a positive effect. This clearly shows that when investors buy a general obligation municipal bond, they do care about the whole state's financial status, besides a local one. The financial status of a community is indeed affected by the financial status of its state government.

For community's financial variables, all variables except total liabilities over total revenues and total fund balance over total revenues have the expected signs, and are significant. Among them, total liabilities over total assets, absolute deficits and total fund balance over total assets have relative larger effects.

We clearly see that most variables relevant to state housing information do not have the expected signs. There are two possible reasons. First, the 2008 financial crisis was largely

caused by the failure of the housing mortgage market, investors actually felt that before the crisis. When a state had a large number of housing sales or building permits, or more people owned houses, investors did not think that this was an indicator of a good economy, but a bubble instead. Second, the demand side of municipal bond markets dominates the effect. Generally speaking, investors only buy their own states' municipal bonds because of tax advantages. When investors have more investment choices or have spent more money on housing markets, naturally they have less demand to buy municipal bonds. Lower demand causes lower bond prices, which raises bond yields.

4.1.2 Time series analysis

Pooled regressions give us rough ideas about how these variables can affect municipal bond yields, but since many variables have large variations across time and the determinants' structures of dependence can change across time, we need to analyze variables month by month. Again, I use all variables' z-scores to make them comparable.

For each variable, I first run an ordinary least square regression in each month from April 2006 to March 2010. Suppose $\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_T$ are the coefficient results of any variable. To smooth the monthly effect, I define another coefficient $\tilde{\beta}_t$, where

$$\tilde{\beta}_t = \begin{cases} 0.85\tilde{\beta}_{t-1} + 0.15\hat{\beta}_t & \text{for } t = 2, 3, \cdots, T \\ \hat{\beta}_t & \text{for } t = 1 \end{cases} \tag{4.1}$$

Figure 6 draws the weighted coefficients $\tilde{\beta}_t$ of all variables. To avoid confusion of directions of variables, I adjust the signs to make all coefficients positive by expectation from the credit concern side.[4]

Panel A and B of Figure 6 show that except existing home sales per person and economic

[4] Or, we can say the supply side.

diversification index, all variables play larger roles during the financial crisis than before the financial crisis. The effect of coincident index (COI) is larger than that of the growth rate of GDP per capita during the financial crisis. Building permits per person do have the expected sign during the financial crisis, and the effect is even larger than COI or the growth rate of GDP per capita. However, existing home sales per person never have the expected sign. The demand side factor dominates the effect of home sales. In addition, the effect of the unemployment rate is the largest one among all variables. Economic diversification index only has the expected sign but small before the financial crisis, which is not consistent with Hastie (1972). But we will see that economic diversification does tell the right story after controlling other variables later.

Panel C of Figure 6 shows that the positive effect of population appears after 2007, while the positive effect of persons per house appears after 2009. Homeownership rates have expected signs just briefly during the financial crisis. Again, demand factors dominate the effect of variables that contain housing information. And the effects of all demographic variables are very small.

Panel D of Figure 6 draws the time series effects of states' debt ratio variables. No variables play roles before the financial crisis, but the effects become very strong during the financial crisis, except for the total liabilities over total revenues. In addition, the effects of all state level financial variables are smaller than most state economic variables and community level financial variables.

Panel E and F of Figure 6 show the effects of community level financial variables. Again, the effects are almost zero before the financial crisis, but gradually become stronger in the financial crisis, except the variable of the community's total liabilities over total revenues, and total fund balance over total revenues. Total liabilities or total fund balance over total

revenues are not good measures to predict municipal bond yields in the linear regressions: Revenues should be relative to expenses, while total liabilities or total fund balance should be measured relative to assets instead.

Generally speaking, the relation between the municipal bond yields and the economic and financial health of the local government has become stronger during the 2008 financial crisis because of the larger differentiation among the local governments and hence larger signal to noise ratio, as well as closer scrutiny by market participants on the economic fundamentals of municipal governments. Including the samples of both normal periods and in the financial crisis can make us understand the determinants of municipal bond yields in a better and broader way.

4.2 Bivariate analysis

Before moving to multiple variable analysis, I provide analyses of two variables for two reasons. First, the effect of a variables might change after controlling another variable. Second, there are possible interaction effects between variables. I choose three typical and interesting bivariate cases to run linear regressions with interaction terms. The pairs I choose are coincident indexes and unemployment rates, populations and economic diversification indexes, and a community's absolute deficits and a community's total fund balance over total assets. To make variables comparable, I use all regressors' z-scores as well.

Pooling all sample results will be reported upon request; I only report the time series results in Figure 7. The same weighted function of (4.1) is used to smooth the monthly effects. All variables have larger amounts of coefficients during the financial crisis than before the financial crisis, which is consistent with the conclusions of univariate regressions. In addtion, panel D of Figure 7 tells us that the predictive powers of regressions increase a lot during

the financial crisis period, and the community financial variables play larger roles than state economic and demographic variables do before the financial crisis.

From the analysis of the last section, we find that coincident indexes (COI) and unemployment rates have the clearest effects on municipal bond yields among state's economic variables. Therefore, I choose these two variables to see their interaction effects. Panel A of Figure 7 shows that, after controlling the unemployment rates, the effects of COI become positive, which is not consistent with the theory. However, the negative coefficients of interaction terms show that COI plays a role on municipal bond yields in another way: The effect of the unemployment rate is larger in a state that has a low COI than in one that has a high COI.

Economic diversification plays very important roles in determining municipal bond yields. Most papers use population as the approximation, while Loviscek and Crowley (1990) generates a new coefficient of specialization index (CDIV) to measure economic diversification. In the univariable analysis, CDIV does not have the expected signs, while the effect of population is positive because other information that population contains dominates economic diversification. Therefore, I choose these two variables to see their interaction effects. One thing worth pointing out is that both the population and economic diversification index are very stable across time in most states, while the cross state differences are very large. The different results we obtain in different time periods must reflect the change of investors' views of how these two variables can affect the municipal bond yields.

Panel B of Figure 7 shows that CDIV does have expected signs after controlling population, especially during the financial crisis. The amounts of coefficients are very large. This finding is consistent with Hastie (1972): Under a very bad economic condition, municipalities that are well diversified will be less susceptible to default than those that are less diversified, therefore have lower bond yields. In addition, the effect of population is far smaller than

CDIV, especially during the financial crisis. Furthermore, the negative coefficients of interaction terms show that both variables' effects are smaller when the value of another variable is larger: The effects of economic diversification index are smaller in large population states than that in small population states, because population itself is a measure of economic diversification. On the other hand, the effects of population are smaller in less economically diversified states than in well economically diversified states.

I choose a community's absolute deficits and a community's total fund balance over total assets to invesitigate possible interaction effects on municipal bond yields because these two variables have the largest estimation coefficients in the univariate analysis, and the low correlation between these variables makes the linear regression analysis relatively more accurate. Panel C of Figure 7 shows that the financial crisis's effects are weaker on these two variables than on the state level variables. In addition, the negative interaction terms show that both variables' effects are amplified when the community's financial status is bad if measured by another variable. However, we have to be careful of the finding here because, different from demographic and economic variables, the financial accounting data have long reporting delays, and the cross time difference of one community is not small. Investors must rely on other information instead of the published financial accounting data.

4.3 Mutiple variable analysis

From earlier analysis, we have already had a brief picture about how these determinants can affect municipal bond yields separately. In this subsection, I combine variables together. If there is a large correlation between two variables, we can only have one of them in the linear regression because of the multicollinearity problem. Under this situation, I choose the determinant that has a stronger effect in the univariate regression. In addition, if the sign of a variable is not expected in earlier analysis, I exclude that variable.

I select eleven determinants in the final linear regression analysis. Eight of them are at state levels: coincident indexes (COI), growth rates of GDP per capita, unemployment rates, economic diversification indexes (CDIV), populations, persons per house, total liabilities over total capital assets, and government fund total liabilities over total revenues. Another three determinants are at community levels: total liabilities over total assets, absolute deficits, and total fund balance over total revenues. I also include the interactions of COI and unemployment rates, and populations and CDIV.[5] I first pool all monthly samples together; then run regressions month by month to investigate possible time series effects. All determinants are standardized before all regressions as well.

4.3.1 Pooling all monthly samples together

The first column of Table 5 reports the regression result of pooling all monthly samples together. Although the signs of most variables match the results in uni- and bivariate analysis, there are a few exceptions: First, the growth rate of GDP becomes positive significant, but the amount is very small. Second, the coefficients of persons per house and community level total fund balance over total revenues are insignificant. Finally, the coefficient of state government fund total liabilities over total revenues becomes negatively significant, which is not expected. All of them are because of the large correlations between variables.

Obviously, community level variables have larger variations than state level variables have, as a result, they might dominate in the regression as well. In addition, community level information is not available for about two thirds of municipal bonds. Therefore, I exclude these community level variables and run the linear regression by using the same samples as the previous regression. The last column of Table 5 shows that the R-squared does not

[5]Including these two interactions can increase R^2 dramatically. Although we need to include more interactions between financial variables, it is hard to control in the linear regression analysis. This will be discussed later in Section 6.

decrease by a large amount, and the coefficients do not change dramatically except one thing: The effect of CDIV is larger, but the coefficient of the interaction of CDIV and population becomes larger as well, which is because of the multicollinearity between CDIV and the interaction of CDIV and population.

We compare the last column's regression results with the regression on state level variables only but including all available samples. The results are reported in the middle column. The coefficient results are very similar to the previous two regressions. However, we obtain a positive coefficient of state government fund total liabilities over total revenues, but the coefficient of state total liabilities over total capital assets decreases a lot. This again is caused by the large correlation between variables of state government fund total liabilities over total revenues and state total liabilities over total capital assets.

4.3.2 Time series analysis

Now we investigate the time series effects. At each month of April 2006 to March 2010, I run a linear regression of municipal bond yields on the eleven determinants and two interaction terms discussed above. Suppose $\hat{\beta}_{1t}, \hat{\beta}_{2t}, \cdots, \hat{\beta}_{13t}$ are the coefficient results at any month t. I still use the equation (4.1) to smooth monthly coefficients. Generally speaking, we observe from Figure 8 that most determinants have larger effects on municipal bond yields during the financial crisis than before the financial crisis as well.

Panel B of Figure 8 is very similar to Panel B of Figure 7. While, although Panel A of Figure 8 and Panel A of Figure 7 provide similar information, the amounts of coefficients are much larger in multiple regressions than those in bivariate regressions. The multicollinearity problem is amplified in the multiple regresions. If we compare Panel C and D of Figure 8 with the corresponding coefficients in Figure 6, the growth rate of GDP per capita, state government fund total liabilities over total revenues, and a community's absolute deficits

change the signs, especially during the financial crisis. The unexpected signs of the first two variables are obtained in the pooled regression as well. Besides the potential multicollinearity problem, the delays of data reporting is another possible reason. Investors might rely on other more updated variables to obtain valuable information of the municipal bond market. A very interesting result is that the coefficients of community absolute deficits never become positive in the time series analysis, which is opposite of the pooled regression. Here, multicollinearity is surely the reason.

In summary, the effects of most determinants can be found in just one or at most two variable analysis. Large correlations between variables can weaken the conclusions of the multiple linear regression analysis. A better model is needed if we need to combine all variables' information together more effectively.

5 Nonlinear effects of variables and shortcomings of the linear regression

5.1 Nonlinear effects in univariate analysis

The effects of most determinants on municipal bond yields are not linear. Investors may only care about the variable when a variable provides the information that the economy or the financial status is really bad. In the municipal bond market, most bonds have very low yields and are rated as AAA, while much fewer bonds have very high yields. This subsection investigates the nonlinear univariate effects, first by pooling all monthly samples together, then by month by month analysis.

5.1.1 Pooling all monthly samples together

Figure 9 draws the nonlinear effects of selected state economic and all financial accounting variables respectively, by using Local Polynomial regressions and pooling all monthly samples together. The graphs of the rest economic and demographic variables are not reported because the results are very close to linear regressions. All variables are standardized first for comparison as well.

Panel A tells several nonlinear effects of economic variables: First, the negative effect of the coincident index (COI) mainly exists when a state is poor. Second, different from the linear regression, we do see a negative effect of building permits per person when permits are low. Finally, the unemployment rate mainly takes effect when the rate is very high.

Panel B shows that state debt ratio variables generally tell unexpected or unclear stories when a state's financial status is very strong. However, states' total liabilities over total capital assets and total liabilities over total revenues have positive effects on municipal bond yields when states are in medium or bad financial status. But the amount of states' total liabilities over GDP per capita does not have a clear effect. The nonlinear effect of states' parent government fund total liabilities over total revenues is not very obvious.

For a community's financial variables, we observe from Panel C and D of Figure 9 that most variables tell similar stories to linear regressions, but we can obtain several extended results: First, there are big nonlinear effects of total liabilities over total assets, relative deficit ratios and total fund balance over total assets. The effects of all these three variables are very small when a community's financial status is good, but big when the status is very bad. The effect of deficit ratios is the largest one. Second, although the amount of total liabilities over total revenues has a negative and significant effect on municipal bond yields in the linear regression, we do see a positive effect within the bad financial status communities. Third,

the amount of total fund balance over total revenues has negative effects except when the amount is very big. Fourth, although the positive effects of absolute deficits are bigger than those of deficit ratios, the large and positive effects of absolute deficits appear in communities that have large surpluses. When communities are running in surplus, investors care about the absolute deficits; while when communities are running in balance or in deficits, deficit ratios play larger roles than absolute deficits do.

5.1.2 Time series analysis of selective variables

Since our real interest is the cross sectional analysis at different months, we have to investigate the nonlinear effects across all months as well. Almost every determinant has its own nonlinear effect on municipal bond yields in every month. Since listing all monthly results for all determinants is impossible, I just choose the variables of the unemployment rates and a community's total liabilities over total assets to state that the nonlinear effects not only exist in the pooled analysis, but also exist and even fluctuate in different time periods.

Suppose that $\hat{f}_1(x), \hat{f}_2(x), \cdots, \hat{f}_T(x)$ are the prediction functions each month from April 2006 to March 2010 of any independent variable x by the Local Polynomial regressions.[6] At each month t, for each variable, I define a new prediction function $\tilde{f}_t$:

$$\tilde{f}_t(\cdot) = \begin{cases} 0.85\tilde{f}_{t-1}(\cdot) + 0.15\hat{f}_t(\cdot) & \text{for } \ t = 2, 3, \cdots, T \\ \hat{f}_t(\cdot) & \text{for } \ t = 1 \end{cases} \tag{5.1}$$

I draw the new predictive municipal bond yields versus the time at some typical values of unemployment rates and a community's total liabilities over total assets in Panel A and B of Figure 10. Notice that, for some values, predictions are not available for all periods, because these values' samples do not exist in some specific months.

[6]These pictures will be almost the same if using the univariate Gradient Boosting method as long as proper boosting parameters are chosen. I use the Local Polynomial regression because it is more commonly used.

Panel A shows that the effect of the unemployment rate is very small before the financial crisis, and in those states with strong employment status during the financial crisis. The effect peaks at the beginning of the financial crisis and in those states that experience very high unemployment rates. Although the nonlinear effect of total liabilities over total assets is not very obvious, we can still observe some time series difference in Panel B. Before the financial crisis, the effect is almost flat; while the positive effect is only clear at the beginning of the financial crisis.

To avoid that the typical values of each variable I choose might be subjective, Panel C and D of Figure 10 draw the effect of determinants by using three typical monthly samples: November 2007, which is before the financial crisis; November 2008, which is at the beginning of the financial crisis; and November 2009, which is at the end of the financial crisis. We can clearly observe that the nonlinear effects and the effects change across time. The municipal bond yields increase a lot when unemployment rates or a community's total liabilities over total assets are very high, and the nonlinear effect is larger during the financial crisis than before the financial crisis. Although not reported, time series univariate nonlinear analysis shows that most determinants have clear and big effects only when a state's economy is really bad or/and a community's financial status is very weak.

5.2 Bivariate analysis

We have already investigated interaction effects between several variables in the linear regression. But as we will see in this subsection, the interaction effects are not linear and might change across time as well. I choose the variables of populations and economic diversification to illustrate the fluctuated nonlinear interaction effect in three typical months: November 2007, November 2008 and November 2009. Figure 11 summarizes the results of the Gradient

Boosting method.[7]

Figure 11 clearly shows that the nonlinear interaction effects not only exist but also fluctuate across time. The effects of both population and economic diversification are very small in 2007. At the beginning of the financial crisis—November 2008, the positive effects of population appear only in high population states, while the positive effects of economic diversification index appear only in well economically diversified states. At the end of the financial crisis—November 2009, positive effects of both population and economic diversification index exist in almost in all groups and become much stronger. In addition, in well economically diversified states, population dominates the impacts on municipal bond yields, while in medium or less diversified economies, economic diversification index plays a more important role than population plays.

The effects of population and economic diversification are summarized as followings. Both variables' effects are very small before the financial crisis. During the financial crisis, investors first investigate the financial status of the states, which can partly be reflected by population sizes. If the current financial status is too bad, the municipal bond yields will be very high; if the status is not too bad, investors will investigate potential economic problems of states, which is measured by economic diversification index. The bond yields will be higher for less economically diversified states than for well economically diversified states. The effect of economic diversification index is not constant under different economic situations during the financial crisis.

[7]The methodology of the Gradient Boosting method is discussed in detail in Appendix B.2

5.3 Multivariate analysis

Besides the nonlinearity effects that cannot be captured by linear regressions, we have already seen that multicollinearity is another potential problem of the multiple linear regression model. There are large correlations between state level debt variables, and between community financial variables. Even for economic variables, moderate correlations exist. I highlight this issue by running two simple regressions including variables that have large correlations. The first regression contains the non-farm payroll employment rate and the unemployment rate. The second regression includes all state level debt ratio variables. In both regressions, I compare the result of all samples with that of the samples where community level financial information is available. Table 6 summarizes the results.

From the first regression to the second one, the coefficient of the unemployment rate becomes positively significant from insignificant. But the absolute amount of the non-farm payroll employment rate decreases as well. Since these two variables are negatively correlated, I calculate the difference between these two coefficients. Very interesting but not surprisingly, the difference does not change dramatically. From the third regression to the fourth one, we obtain similar conclusions. Although the coefficient of government fund total liabilities over total revenues becomes positively significant from insignificant, the summation of coefficients of all debt ratio variables, which are highly positively correlated, does not change a lot either. Also the amount of most coefficients are amplified compared with the univariate regressions.

From these two examples, we clearly see how multicollinearity can affect the regression results dramatically. If variables are highly positively correlated, some coefficients might be largely positive, while others will be largely negative. It will be very challenging to explain these coefficients. In the multiple regression analysis, we cannot have all variables. However, excluding variables will lose some useful information, especially when constructing predictive

models. Another shortcoming of the linear regression is that, as we have already seen in the previous analysis, the determinants that affect municipal bond yields change from time to time. The linear regression gives the same weight to all variables all the time, which is not a good pre-assumption. Therefore, we need another framework to provide a better predictive method.

5.4 Summary of shortcomings of the linear regression

The previous examples in this section clearly indicate three important shortcomings of the linear regression method: First, it cannot capture nonlinear effects, which are universally existed in the determinants of municipal bond yields. Second, even though we can add square or interaction terms in the linear regression, large fluctuations of nonlinear and interaction effects across different time make the model specifications very challenging. Third, highly correlated variables can cause multicollinearity, which can reduce predictive powers. Excluding some variables will cause losses of useful information. Therefore, we need a more powerful predictive model to capture these effects and improve the prediction power of the municipal bond market. The new model has to be able to capture nonlinear and interaction effects, be flexible when choosing variables under different situations, and can summarize all variables' information more effectively.

6 Advantages of the Gradient Boosting method and predictive framework

Although the Local Polynomial regression method can capture most nonlinear effects, it is not very good when variables have interaction effects. The previous section shows that random interaction effects do exist. In addition, Local Polynomial regressions cannot provide

variables' weights automatically in multiple variable analysis. Therefore, I rely on the Gradient Boosting framework to offer a predictive framework in this paper.

6.1 Advantages of the Gradient Boosting method

To see how the Gradient Boosting method can provide useful information that cannot be provided in the traditional linear regression method, I simulate a few models based on the properties of municipal bond markets and report both the in-sample and out-sample R-squared of both methods.

Suppose $X_1, X_2, \cdots, X_L$ independently follow student t distributions with degrees of freedom 10, and all ϵ below independently follow a normal distribution with a mean 0 and a standard deviation 10. Four models are constructed. The first model captures only linear effects. The second model contains random nonlinear effects. The third model includes truncation effects. And the last one captures random interaction effects.

$$\text{Model 1: } Y = \sum_{l=1}^{L} X_l + \epsilon. \tag{6.1}$$

$$\text{Model 2: } Y = \sum_{l=1}^{L} |X_l|^{D_L} + \epsilon, \tag{6.2}$$

where $D_1, D_2, \cdots, D_L$ are independently uniformly distributed between 1 and 4.

$$\text{Model 3: } Y = \sum_{l=1}^{L} (|X_l| - A_l)_+^2 + \epsilon, \tag{6.3}$$

where $A_1, A_2, \cdots, A_L$ are independently uniformly distributed between 1 and 2.

$$\text{Model 4: } Y = \sum_{l=1}^{L} |X_l|^{D_L} + \sum_{k_1,k_2} I(k_1, k_2) + \epsilon, \text{where} \tag{6.4}$$

$$I(k_1, k_2) = \begin{cases} x_{k_1} x_{k_2} & \text{with probability } 0.4 \\ 0 & \text{with probability } 0.6 \end{cases} \tag{6.5}$$

I let $l = 15$. For each model, I generate 2,000 observations, then randomly choose half for estimation and another half for testing. The process is repeated 20 times and I calculate the average in-sample and out-sample R-squared of different models. Table 7 summarizes the results. The Gradient Boosting method can capture nonlinear, random truncation and random interaction effects effectively. Both the in-sample and out-sample R-squared are higher. Especially, Figure 12 draws the percentage of chosen pairs of variables that have the interaction effects in Model 4 against different boosting steps. This figure clearly shows how the Gradient Boosting method can capture random interaction effects. Those percentages of chosen pairs are much higher than 0.4, which is the ratio if we choose pairs randomly at each step, even after 200 boosting steps.[8]

6.2 Predictive framework

Thanks to the properties of the Gradient Boosting framework, we can include all determinants when predicting municipal bond yields. Suppose $\hat{f}_t(x_1, \cdots, x_K)$ is the predictive function of any month t from April 2006 to March 2010, where K is the number of determinants. As before, I still construct a new predictive function $\tilde{f}_t(x_1, \cdots, x_K)$ to smooth monthly effects:

$$\tilde{f}_t(\cdot) = \begin{cases} 0.85\tilde{f}_{t-1}(\cdot) + 0.15\hat{f}_t(\cdot) & \text{for } \ t = 2, 3, \cdots, T \\ \hat{f}_t(\cdot) & \text{for } \ t = 1 \end{cases} \tag{6.6}$$

To see how this new predictive function can improve predictive powers, at each month, I

[8] Actually, most of predictive powers come from the first few steps of boosting. Therefore, the boosting framework is even more effective.

randomly choose half samples of municipal bond yields and their characteristics to construct a predictive function $\tilde{f}_t(\cdot)$, and another half samples for an out of sample test. Then I compare this framework with the multiple linear regression model with interaction terms that is discussed in the previous section. The process is repeated 10 times and the average in-sample and out-sample R-squared for both the linear regression and this new predictive method are reported. In addition, since there are lots of missing values of financial variables at the community level, I repeat the whole process by only including all state level variables. Panel A and B of Figure 13 draw the monthly average in-sample and out-sample R-squared of simulations seperately, both from the linear regression and the Gradient Boosting framework. We can obtain several pieces of information here.

Generally speaking, both the in sample and out of sample R-squared are higher during the financial crisis than before the financial crisis, no matter which method we choose. Before the financial crisis, the R-squared of regression models are close to zero, or even negative. This finding is consistent with previous analysis. Even we can obtain several useful results from the data in normal periods, we have to be careful that these effects are very small, especially in the linear analysis. Demand side factors, tax factors and liquidities of municipal bonds determine bond yields in normal periods. But during the financial crisis, credit becomes a real concern for investors.

In addition, no matter whether using state variables only or all variables, both the in sample and the out of sample R-squared of Gradient Boosting methods are consistently higher than those of linear regression methods across all time periods. Lots of determinants have large nonlinear and random interaction effects on municipal bond yields. Especially from March 2009 to September 2009, the out of sample R-squared of the linear regression that uses all variables are largely negative. But from the Gradient Boosting method, we can still have plenty of predictive powers. In this period, large nonlinearities of determinants, uncertainties

of the market, and delays of data reporting amplify the problems of the linear regression. However, Gradient Boosting framework is very robust and can overcome most of these shortcomings. Through this framework, we can significantly enhance the predictive performance on the municipal bond yields.

Furthermore, the predictive powers of including all variables are higher than those of including state variables only, except for 2007 when both powers are very low. Before the financial crisis, the effects of all determinants are very weak. However, in the financial crisis, both state and community level financial variables can provide valuable information. Especially in the Gradient Boosting framework, the information of the 2008 financial crisis has indeed helped us identify the key determinants of municipal bond yields.

7 Conclusion and further discussion

Determinants of municipal bond yields in the secondary market do not draw large attention in the literature. This paper first shows that three classes of variables can affect general obligation municipal bond yields: the economic status of the state where the bond is issued; the demographic characteristics of the state; and the financial status of the state or the local government where the bond is issued. In addition, a systematic comparison of the relations before and after the 2008 financial crisis shows that the economic and financial health of local governments has become markedly more diverse since the crisis began. The relation between the municipal bond yields and the economic and financial health of the local governments has also become stronger because of the larger differentiation among the local governments and hence larger signal to noise ratio, as well as closer scrutiny by market participants on the economic fundamentals of municipal governments. Finally, this paper applies both the Local Polynomial regression and the Gradient Boosting method to capture the nonlinear

effects and random interactions among key determinants on bond yields, and these models can significantly enhance the predictive performance on the municipal bond yields.

There are many other factors, which are not included in this paper but can affect municipal bond yields, especially bond specific characteristics. Further research can include these factors. More local community's economic and demographic data if available can provide more information as well. In addition, how these determinants can affect bond yields after the financial crisis, we cannot know by the current data, because although officially the financial crisis was over in 2009, the effect of the crisis is far from over now, especially for the local governments' deficit problems. There are two possibilities. First, investors will learn from the financial crisis, and these determinants continue to affect municipal bond yields, but at smaller amounts. Second, investors will forget the lessons from the financial crisis, the market will go back to the characteristics before the financial crisis. We can only obtain the answers after the effect of the fiscal crisis is over. Finally, there are large missing values in the community level financial accounting data. The analysis of the Gradient Boosting method is more flexible than that of the traditional linear regression method when dealing with missing values. How we can use the Gradient Boosting method more effectively to improve the predictive powers even more will continue to be investigated as well.

References

Yaw A. Badu, Kwadwo Bawuah, and Kenneth Daniels. Race and Municipal Bond Ratings in the Commonwealth of Virginia. *The Review of Black Political Economy*, 24(4):111–121, June 1996.

Donald J Boyd and Lucy Dadayan. Revenue Declines Less Severe, But States' Fiscal Crisis Is Far From Over. *State Revenues Report*, (79), April 2010.

Willard T Carleton and Eugene M Lerner. Statistical Credit Scoring of Municipal Bonds. *Journal of Money, Credit and Banking*, 1(4):750–764, November 1969.

Jerome H Friedman. Stochastic Gradient Boosting. Technical report, Stanford University, March 1999.

Jerome H Friedman. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5):1189–1232, October 2001.

Michael Grossman, Fred Goldman, Susan Nesbitt, and Pamela Mobilia. Determinants of interest rates on tax-exempt hospital bonds. *Journal of Health Economics*, 12:385–410, 1993.

Wolfgang Hardle. *Smoothing Techniques: with implementation in S*. Springer-Verlag, 1990.

Larry K Hastie. Determinants of Municipal Bond Yields. *Journal of Financial and Quantitative Analysis*, 7(3):1729–1748, June 1972.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, second edition.

Joseph J Horton. A Statistical Rating Index for Municipal Bonds. *Financial Analysts Journal*, 25(2):72–75, March-April 1969.

Reuben Kessel. A Study of the Effects of Competition in the Tax-exempt Bond Market. *Journal of Political Economy*, 79(4):706–738, July-August 1971.

Linda H Lipnick, Yaffa Rattner, and Linda Ebrahim. The Determinants of Municipal Credit Quality. *Government Finance Review*, pages 35–41, December 1999.

Anthony L Loviscek and Frederick D Crowley. What is in a Municipal Bond Rating? *The Financial Review*, 25(1):25–53, February 1990.

Allen J Michel. Municipal Bond Ratings: A discriminant Analysis Approach. *Journal of Financial and Quantitative Analysis*, 12(4):587–598, November 1977.

Municipal Securities Rulemaking Board. Glossary of Municipal Securities Terms. URL `http://www.msrb.org/msrb1/glossary/`.

George Palumbo, Richard Schick, and Mark Zaporowski. Factors Affecting A Municipality's Bond Rating: An Empirical Study. *Journal of Business & Economics Research*, 4(11): 37–42, November 2006.

James M Poterba and Kim S Rueben. State Fiscal Institutions and the U.S. Municipal Bond Market, working paper. 1997.

Daniel Rubinfeld. Credit Ratings and the Market for General Obligation Municipal Bonds. *National Tax Journal*, XXVI(1):17–27, March 1973.

The Committee on Financial Services of the 110th Congress. Municipal Bond Fairness Act. Technical Report HR 6308, The House of Representative, September 2008.

8 Appendix

A Classification of different types of municipal bonds

Municipal bonds are generally exempt from federal taxes on their interest payments, although not on capital gains. For investors who purchase municipal bonds issued in the same state where they live, interest payments are generally exempt from state and local tax as well. There are different angles to classify municipal bonds. Based on Municipal Securities Rulemaking Board, these include issuer types, refunding types, tax status, coupon types, whether it is bank qualified, whether it is callable, and flow of funds.

Based on the issuer types, municipal bonds are classified as GENERAL OBLIGATION BONDS and REVENUE BONDS. GENERAL OBLIGATION BOND is secured by a state or local government's pledge to use legally available resources, including tax revenues, to repay bond holders. Most general obligation bonds pledge to levy a property tax to meet debt service requirements at the local government level. Because property owners are usually reluctant to risk losing their holding due to unpaid property tax bills, credit rating agencies often consider these bonds to have very strong credit qualities. REVENUE BONDS may be issued by an agency, commission, or authority created by legislation in order to construct a "facility," such as a toll bridge, turnpike, hospital, university dormitory, water, sewer, utilities and electric districts, or ports. The fees, taxes, or tolls charged for use of the facility ultimately pay off the debt. Governments with the power to tax also issue revenue bonds, but restrict the debt service funds to only those funds from the governmental enterprise that generates these revenues. The credit quality of revenue bonds are generally lower than general obligation bonds. Therefore, the bond yields of revenue bonds are generally higher.

Refunding is a procedure whereby an issuer refinances outstanding bonds by issuing new bonds. The new bonds are referred to as the "REFUNDING BONDS" and the outstanding

bonds being refinanced are referred to as the "REFUNDED BONDS" or the "PRIOR ISSUE".

TAXABLE MUNICIPAL BOND is a bond issued by a municipal issuer for which interest or other investment return is included in gross income for federal income tax purposes. A municipal security may be issued on a taxable basis because the intended use of proceeds because certain federal tax law requirements are not met. Correspondingly, TAX-EXEMPT BOND is a bond whose interests are excluded from gross income for federal income tax purposes but may or may not be exempt from state income or personal property taxes. Among tax-exempt bonds, one type is BANK QUALIFIED. It is the designation given to a public purpose bond offering by the issuer if it reasonably expects to issue in the calendar year of such offering no more than $10 million par amount of bonds of the type required to be included in making such calculation under the Internal Revenue Code. Another type is NOT BANK QUALIFIED. The banks have to pay taxes if buying these bonds, although they seldom buy them. Another special bond is called TAXABLE EXCHANGEABLE BOND. This is a bond initially issued on a taxable basis that may convert to tax-exempt status upon the occurrence of a specified condition precedent.

ZERO COUPON BOND is an original issue discount bond on which no periodic interest payments are made. Correspondingly, COUPON BOND is a bond on which period interest payments are made. Bonds bear interest at either a fixed (Called FIXED BOND) or variable rate of interest (called ADJUST CONVERT TO FIXED BOND). ORIGINAL ISSUE DISCOUNT BOND (O.I.D. BOND) is a bond that was sold at the time of issue at a price that included an original issue discount, which is an amount by which the par value of a security exceeded its public offering price at the time of its original issuance.

CALLABLE BOND is a bond that the issuer is permitted or required to redeem before

the stated maturity at a specified price, usually at or above par, by giving notice of redemption in a manner specified in the bond contract. NON-CALLABLE BOND is a bond that cannot be redeemed at the issuer's option before its stated maturity date.

The order and priority of handling, depositing and disbursing pledged revenues, as set forth in the bond contract. Generally, the revenues are deposited, as received, into a general collection account or revenue fund for disbursement into the other accounts established by the bond contract. Such other accounts generally provide for payment of the costs of debt service, debt service reserve deposits, operation and maintenance costs, renewal and replacement and other requirements. SINKING FUND (MANDATORY REDEMPTION FUND) is one of the examples. It is a fund into which the issuer makes periodic deposits of moneys to be used to pay the costs of calling bonds in accordance with the mandatory redemption schedule in the bond contract or to purchase bonds in the open market in satisfaction of such mandatory redemption requirement.

B Econometric methods

B.1 Local Polynomial regression

This section introduces a simple case of the Local Polynomial regression. The framework is as follows:

Step 1: Choose M points: $x_1, x_2, \cdots, x_M$, based on positions of the independent variable.

Step 2: Let h represent the half-width of a window encompassing the nearest neighbors. Calculate h based on Silverman's rule-of-thumb (ROT) bandwidth estimator (Hardle (1990))

$$\hat{h}_{ROT} = 1.06\hat{\sigma}n^{-1/5}, \tag{8.1}$$

where $\hat{\sigma}$ is the sample standard deviation of the independent variable.

Step 3: For any $x_m, m = 1, 2, \cdots M$, an estimation $\hat{f}(x_m)$ is obtained by the polynomial regression of y on x to minimize the weighted sum of squared residuals

$$\sum_{i=1}^{n} W_T(x_i)(y_i - a - b_1 x_i - b_2 x_i^2 - \cdots - b_k x_i^k)^2, \tag{8.2}$$

where $W_T(x)$ is a unimodal weighted function, centered on x_m and goes to 0 when at the boundaries. In this paper, I use the Epanechnikov Kernel function, which is

$$W_T(x) = \begin{cases} \frac{3}{4}(1 - (\frac{|x-x_m|}{h})^2) & \text{if } \frac{|x-x_m|}{h} < 1 \\ 0 & \text{otherwise} \end{cases} \tag{8.3}$$

To simplify, I let $k = 1$ in this paper unless stated.

B.2 Gradient Boosting method using regression trees

The Gradient Boosting method is first introduced by Friedman (1999) and Friedman (2001), and summarized in more details in Hastie et al. (2009). The method is an effective procedure that combines the outputs of many weak predictors to produce a powerful one. The purpose of boosting is to sequentially apply the weak predictor algorithm to repeatedly modified versions of the data, thereby producing a sequence of weak predictors $b(\boldsymbol{x};\gamma_j), j = 1, 2, \cdots, J$, where $b(\boldsymbol{x};\boldsymbol{\gamma}) \in \mathbb{R}$ is any function of the single or multivariate argument $\boldsymbol{x}$, characterized by a set of parameters $\boldsymbol{\gamma}$. I use the forward stagewise additive boosting by using the regression tree as the basic function. The first part of this subsection introduces the regression tree. The second part explains the forward stagewise additive modeling boosting. The final part combines them together.

Suppose in each month, the data consists of p predictors and a yield, for each of N securities: that is, $(\boldsymbol{x_i}, y_i)$ for $i = 1, 2, \cdots, N$, with $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \cdots, x_{ip})$.

B.2.1 Regression trees

First, partition the sample space into M regions: $R_1, R_2, \cdots, R_M$, and model the yield as a constant c_m in each region:

$$T(\boldsymbol{x};\Theta) = \sum_{m=1}^{M} c_m I(\boldsymbol{x} \in R_m) \tag{8.4}$$

where $I(\cdot)$ is the indicator function defined as:

$$I(\boldsymbol{x} \in R_m) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in R_m \\ 0 & \text{otherwise.} \end{cases} \tag{8.5}$$

Parameters $\Theta = \{R_m, c_m\}_{m=1}^{M}$ are needed to estimate.

Use the criterion of

$$\arg\min_{T} \sum_{i=1}^{N} (y_i - T(\boldsymbol{x_i}))^2, \tag{8.6}$$

it is obvious to see that the best c_m is just the average of y_i in each region R_m if the region is given.

The problem now is that finding the best partition is generally computationally infeasible. I process a greedy algorithm based on Hastie et al. (2009) section 9.2. Starting with all the data, consider a splitting variable k and split point s, and define the pair of half-planes

$$R_1(k,s) = \{\boldsymbol{x}|x_k \leq s\} \text{ and } R_2(k,s) = \{\boldsymbol{x}|x_k > s\}. \tag{8.7}$$

Then seek the splitting variable k and split point s that solve

$$\min_{k,s}[\min_{c_1} \sum_{x_i \in R_1(k,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(k,s)} (y_i - c_2)^2.] \tag{8.8}$$

For any choice k and s, the inner minimization is solved by

$$\hat{c}_1 = ave(y_i|x_i \in R_1(k,s)) \text{ and } \hat{c}_2 = ave(y_i|x_i \in R_2(k,s)). \tag{8.9}$$

For each splitting variable $x_1, x_2, \cdots, x_p$, I use its $2.5^{th}, 5^{th} \cdots 97.5^{th}$ percentiles as the possible split points respectively. Having found the best split, the data is partitioned into the two resulting regions and the splitting process is repeated on each of the two regions. In each step, the split is chosen by minimizing the squared loss error (8.6). Then this process is repeated M times on all of the resulting regions.

B.2.2 Forward stagewise additive modeling boosting

As Hastie et al. (2009) section 10.3 offers, I use the algorithm of the forward stagewise additive modeling boosting. I first initialize $f_0(x) = 0$. Then at each iteration j, I solve the basis function $b(\boldsymbol{x}; \gamma_j)$ and its corresponding coefficient β_j,

$$(\beta_j, \gamma_j) = \arg\min_{\beta,\gamma} \sum_{i=1}^{N} L(y_i, f_{j-1}(x_i) + \beta b(x_i; \gamma)). \tag{8.10}$$

Finally, I add this to the current expansion $f_{j-1}(x)$ by setting

$$f_j(x) = f_{j-1}(x) + \beta_j b(x; \gamma_j). \tag{8.11}$$

This process is repeated J times to obtain $f_J(x)$, which is the final predictive function.

For the squared-error loss function

$$L(y, f(x)) = (y - f(x))^2, \tag{8.12}$$

$$L(y_i, f_{j-1}(x_i) + \beta b(x_i; \gamma)) = (y_i - f_{j-1}(x_i) - \beta b(x_i; \gamma))^2 = (r_{ij} - \beta b(x_i; \gamma))^2, \tag{8.13}$$

where $r_{ij} = y_i - f_{j-1}(x_i)$ is just the residual of the current model on the ith observation. Therefore, for squared-error loss function (8.12), in each iteration, I only need to predict the residual of the current model by using the existed predictors.

B.2.3 Boosting Trees

I combine the framework of the previous two parts together. The boosted tree model is a sum of trees of (8.4),

$$f_J(x) = \sum_{j=1}^{J} T(x; \Theta_j). \tag{8.14}$$

At each step in the forward stagewise procedure, solve

$$\hat{\Theta}_j = \arg\min_{\Theta_j} \sum_{i=1}^{N} L(y_i, f_{j-1}(x_i) + T(x; \Theta_j)) \tag{8.15}$$

for the regression set and constants $\Theta_j = \{R_{jm}, c_{jm}\}_1^M$ of the next tree, given the current model $f_{j-1}(x)$. Given the regions R_{jm}, finding the optimal constants c_{jm} in each region is straightforward:

$$\hat{c}_{jm} = \arg\min_{c_{jm}} \sum_{x_i \in R_{jm}} L(y_i, f_{j-1}(x_i) + c_{jm}). \tag{8.16}$$

Again, finding the region is difficult, but the problem simplifies for squared-error loss function (8.12). It is simply the regression tree that best predicts the current residuals $y_i - f_{j-1}(x_i)$, and $\hat{c}_{jm}$ is the mean of these residuals in each corresponding region.

As Hastie et al. (2009) section 10.10 states, fast approximate algorithms for solving (8.16) with any differentiable loss criterion can be derived by analogy to numerical optimization, which is called the Gradient Boosting method. Under the loss function (8.12), the Gradient Boosting on its own is equivalent to the method stated in this section. I use the terminology "Gradient Boosting method" in this paper to refer the framework I have summarized here.

C Tables

Table 1: **Variable lists of the determinants of municipal bond yields**

Variables	Descriptions	Frequencies
State economic variables:		
COI	The coincident index	monthly
GDP	The growth rate of GDP per capita	yearly
PERMIT	The number of building permits per capita	monthly
HOMESALE	12-month existing home sales per person	quarterly
NONFARM	The non-farm payroll employment rate	monthly
UNEMPLOY	The unemployment rate	monthly
CDIV	Economic diversification index	yearly
State demographic variables:		
POP	The logarithm of total populations	yearly
PERSONS	The average number of persons per house	yearly
HOMEOWE	The homeownership rate	yearly
State financial variables:		
S_LIA_CAPASSET	Total liabilities over total capital assets	yearly
S_LIA_GDP	Total liabilities over GDP per capita	yearly
S_LIA_REV	Total liabilities over total revenues	yearly
S_LIA_REV2	Government fund total liabilities over total revenues	yearly
Community financial variables:		
C_LIA_REV	Total liabilities over total revenues	yearly
C_LIA_ASSET	Total liabilities over total assets	yearly
C_DEF	Absolute deficits	yearly
C_DEFRATIO	Relative deficit ratio	yearly
C_FUND_REV	Fund balance over revenues	yearly
C_FUND_ASSET	Fund balance over total assets	yearly

Table 2: **Summary statistics of state's financial accounting variables**

Panel A: Total liabilities over total capital assets

year	2005	2006	2007	2008
numbers of observation	49	48	49	48
mean	1.21	0.90	0.92	0.92
median	0.71	0.71	0.79	0.75
standard deviation	2.17	0.63	0.60	0.61
skewness	6.01	1.42	1.45	1.64

Panel B: Total liabilities over GDP per capita

year	2005	2006	2007	2008
numbers of observation	49	48	49	48
mean	0.46	0.47	0.45	0.48
median	0.22	0.27	0.25	0.23
standard deviation	0.61	0.59	0.59	0.62
skewness	2.82	2.86	2.79	2.80

Panel C: Total liabilities over total revenues

year	2005	2006	2007	2008
numbers of observation	49	48	49	48
mean	0.57	0.55	0.55	0.57
median	0.48	0.46	0.47	0.46
standard deviation	0.29	0.27	0.26	0.27
skewness	0.80	0.71	0.61	0.79

Panel D: Government fund total liabilities over total revenues

year	2005	2006	2007	2008
numbers of observation	49	48	49	48
mean	0.18	0.18	0.19	0.17
median	0.17	0.16	0.19	0.16
standard deviation	0.09	0.09	0.10	0.07
skewness	1.72	0.94	1.38	0.50

Table 3: **Summary statistics of community's local financial accounting variables**

Panel A: Total liabilities over total revenues

year	2005	2006	2007	2008
numbers of observation	511	510	525	489
mean	0.24	0.25	0.23	0.22
median	0.13	0.13	0.13	0.12
standard deviation	0.30	0.35	0.32	0.28
skewness	3.1	3.9	4.7	4.1

Panel B: Total liabilities over total assets

year	2005	2006	2007	2008
numbers of observation	512	510	527	491
mean	0.49	0.47	0.47	0.46
median	0.47	0.44	0.43	0.42
standard deviation	0.28	0.26	0.28	0.28
skewness	0.82	0.43	1.37	0.79

Panel C: Absolute deficits (in millions dollars)

year	2005	2006	2007	2008
numbers of observation	1094	1045	1058	1078
mean	-6.5	-8.2	-8.9	-5.4
median	-1.3	-1.6	-1.7	-1.0
standard deviation	44	41	47	47
skewness	-8.2	-9.5	-8.9	-6.2

Table 3: (continued)

Panel D: Relative deficit ratio

year	2005	2006	2007	2008
numbers of observation	1094	1045	1058	1078
mean	-0.031	-0.045	-0.041	-0.022
median	-0.028	-0.033	-0.038	-0.021
standard deviation	0.27	0.23	0.24	0.19
skewness	18	15	16	5

Panel E: Total fund balance to total revenues

year	2005	2006	2007	2008
numbers of observation	516	516	530	491
mean	0.35	0.34	0.33	0.25
median	0.15	0.16	0.17	0.17
standard deviation	2.88	2.24	1.85	0.27
skewness	22	22	22	3

Panel F: Total fund balance to total assets

year	2005	2006	2007	2008
numbers of observation	517	516	532	493
mean	0.52	0.54	0.54	0.54
median	0.53	0.57	0.57	0.58
standard deviation	0.28	0.27	0.29	0.28
skewness	-0.78	-0.41	-1.33	-0.60

Table 4: **Pooled univariate linear regression results**

Variables	Coefficients	Expected sign?
State economic variables:		
COI	-0.024***	yes
GDP	-0.007***	yes
PERMIT	0.002***	no
HOMESALE	0.028***	no
NONFARM	-0.042***	yes
UNEMPLOY	0.037***	yes
CDIV	-0.007***	no
State demographic variables:		
POP	0.012***	?
PERSONS	0.003***	yes
HOMEOWE	0.009***	no
State financial variables:		
S_LIA_CAPASSET	0.015***	yes
S_LIA_GDP	0.009***	yes
S_LIA_REV	-0.007***	no
S_LIA_REV2	0.009***	yes
Community financial variables:		
C_LIA_REV	-0.004**	no
C_LIA_ASSET	0.029***	yes
C_DEF	0.028***	yes
C_DEFRATIO	0.012***	yes
C_FUND_REV	-0.002	yes
C_FUND_ASSET	-0.029***	yes

Notes:

1. All variables are standardized first.
2. Robust standard errors are used.
3. ***: significant at 1% level; **: significant at 5% level; *: significant at 1% level.

Table 5: **Multiple variable linear regression results**

Variables	All variables (Partial Samples)	State variables only (All Samples)	State variables only (Partial Samples)
COI	0.092***	0.098***	0.096***
GDP	0.007***	0.000	0.005**
UNEMPLOY	0.342***	0.321***	0.353***
CDIV	0.107***	0.209***	0.169***
POP	0.041***	0.042***	0.050
PERSONS	0.001	-0.002*	-0.002
COI*UNEMPLOY	-0.330***	-0.288***	-0.340***
CDIV*POP	-0.103***	-0.193***	-0.168***
S_LIA_CAPASSET	0.013***	0.003**	0.014***
S_LIA_REV2	-0.007***	0.013***	-0.004**
C_LIA_ASSET	0.020***		
C_DEF	0.013***		
C_FUND_REV	-0.001		
Number of samples	64567	219451	64567
R^2	.031	.021	.028

Notes:

1. All variables are standardized first.
2. Robust standard errors are used.
3. ***: significant at 1% level; **: significant at 5% level; *: significant at 1% level.
4. Partial samples: Samples that community level financial variables are not missing.

Table 6: **Selected linear regression results of different samples**

Samples	Partial Samples	All Samples	Partial Samples	All Samples
Number of samples	64567	220320	64567	219451
Regression Number	(1)	(2)	(3)	(4)
NONFARM	-0.041***	-0.028***		
UNEMPLOY	0.003	0.020***		
S_LIA_CAPASSET			0.050***	0.054***
S_LIA_GDP			0.015***	0.011***
S_LIA_REV			-0.035***	-0.055***
S_LIA_REV2			-0.001	0.011***
Difference of coefficients	-0.044	-0.048		
Summation of coefficients			0.029	0.021
R^2	.011	.009	.010	0.007

Notes:

1. All variables are standardized first.
2. Robust standard errors are used.
3. ***: significant at 1% level; **: significant at 5% level; *: significant at 1% level.
4. Partial samples: Samples that community level financial variables are not missing.

Table 7: **Average in-sample and out-sample R^2 of simulation samples.**

	Linear Regression		Gradient Boosting	
	in-sample	out-sample	in-sample	out-sample
Pure Linear	0.70	0.69	0.75	0.59
Random Nonlinear	0.41	0.33	0.90	0.50
Random Truncation	0.49	0.47	0.76	0.63
Random Interaction	0.22	0.19	0.98	0.25

D Figures

Figure 1: **Correlations between municipal bond yields and time to maturity**
Lines plot the correlations between the bond yields and time to maturity against the time for both bid (solid line) and ask (dashed line) quotes in each month from April 2006 to March 2010.

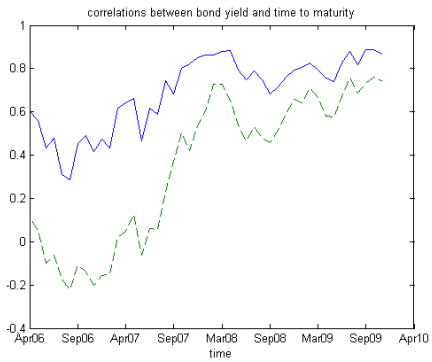


Figure 2: **The R^2 of Linear and Local Polynomial regressions of municipal bond yields on time to maturity**
Lines plot the R-squared of bond yields on time to maturity against the time for both linear regressions (solid line) and Local Polynomial regressions (dashed line) in each month from April 2006 to March 2010.

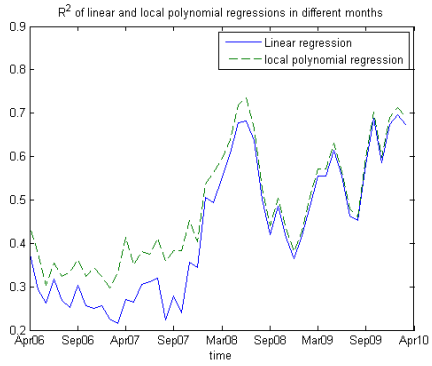


Figure 3: **The standard deviations of municipal bond yields**
The line plots the standard deviations of bond yield predictions from the Local Polynomial regressions against time in each month from April 2006 to March 2010.

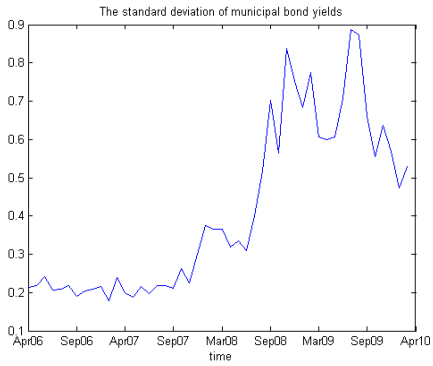


Figure 4: **Summary statistics of state's demographic variables**
Lines plot values of state's demographic variables against the time in each month from April 2006 to March 2010. The variables include logarithm of populations, persons per house and homeownership rates. For each variable, I choose the states at the fifth percentile (solid lines), the median (dashed lines) and the ninety-fifth percentile (dotted lines).

Panel A Log(population) Panel B Persons per house Panel C Homeownership rates

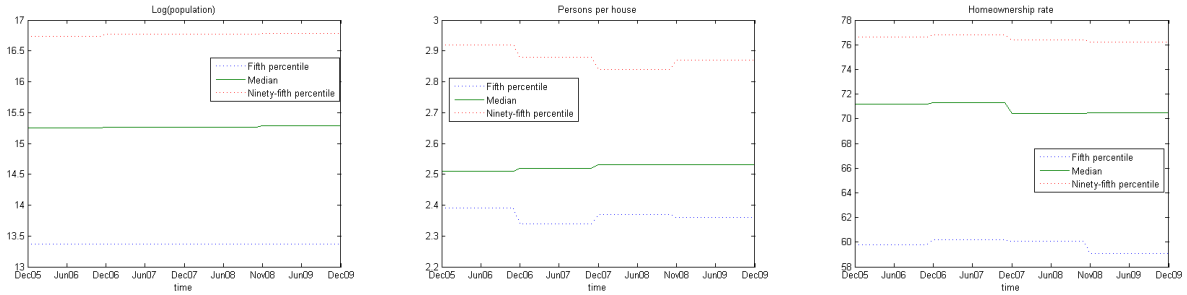
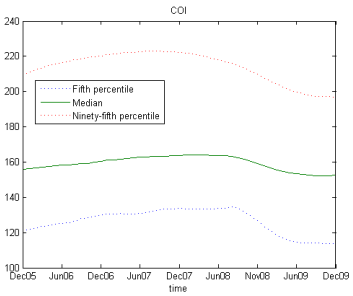


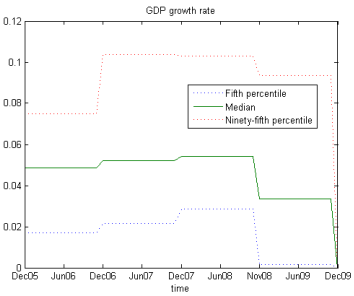
Figure 5: **Summary statistics of state's economic variables**

Lines plot values of state's economic variables against the time in each month from April 2006 to March 2010. The variables include the growth rate of GDP per capita, the coincident index, 12-month existing home sales per person, building permits per person, the unemployment rate, the non-farm payroll employment rate and economic diversification index CDIV. For each variable, I choose the states at the median (solid lines), the fifth percentile (lower dashed lines) and the ninety-fifth percentile (upper dotted lines).

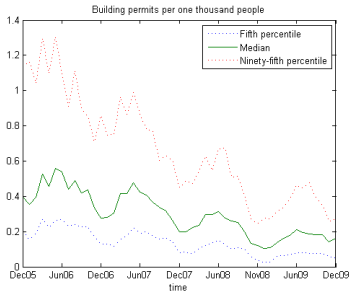
Panel A COI



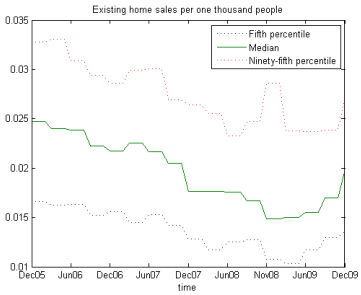
Panel B GDP growth rate



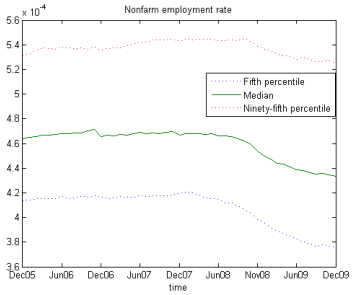
Panel C
Building permits per thousand people



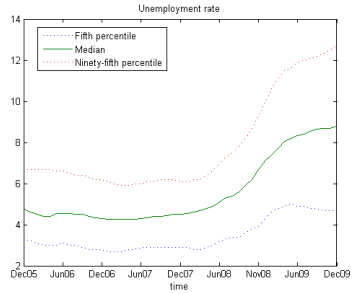
Panel D
Existing home sales per thousand people



Panel E Nonfarm employment rate



Panel F Unemployment rate



Panel G CDIV

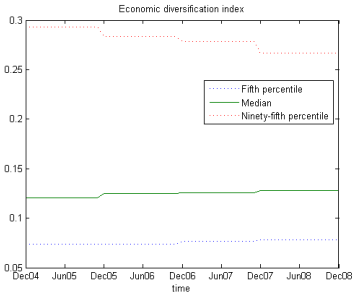


Figure 6: **Monthly univariate linear regression coefficients of variables**
Lines plot the coefficients of each standardized variable against the time. To smooth the effect, for each variable, I draw $\tilde{\beta}_t$: $\tilde{\beta}_t = 0.85\tilde{\beta}_{t-1} + 0.15\hat{\beta}_t$ for $t = 2, 3, \cdots T$, and $\tilde{\beta}_1 = \hat{\beta}_1$, where $\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_T$ are the linear regression coefficients of each month.

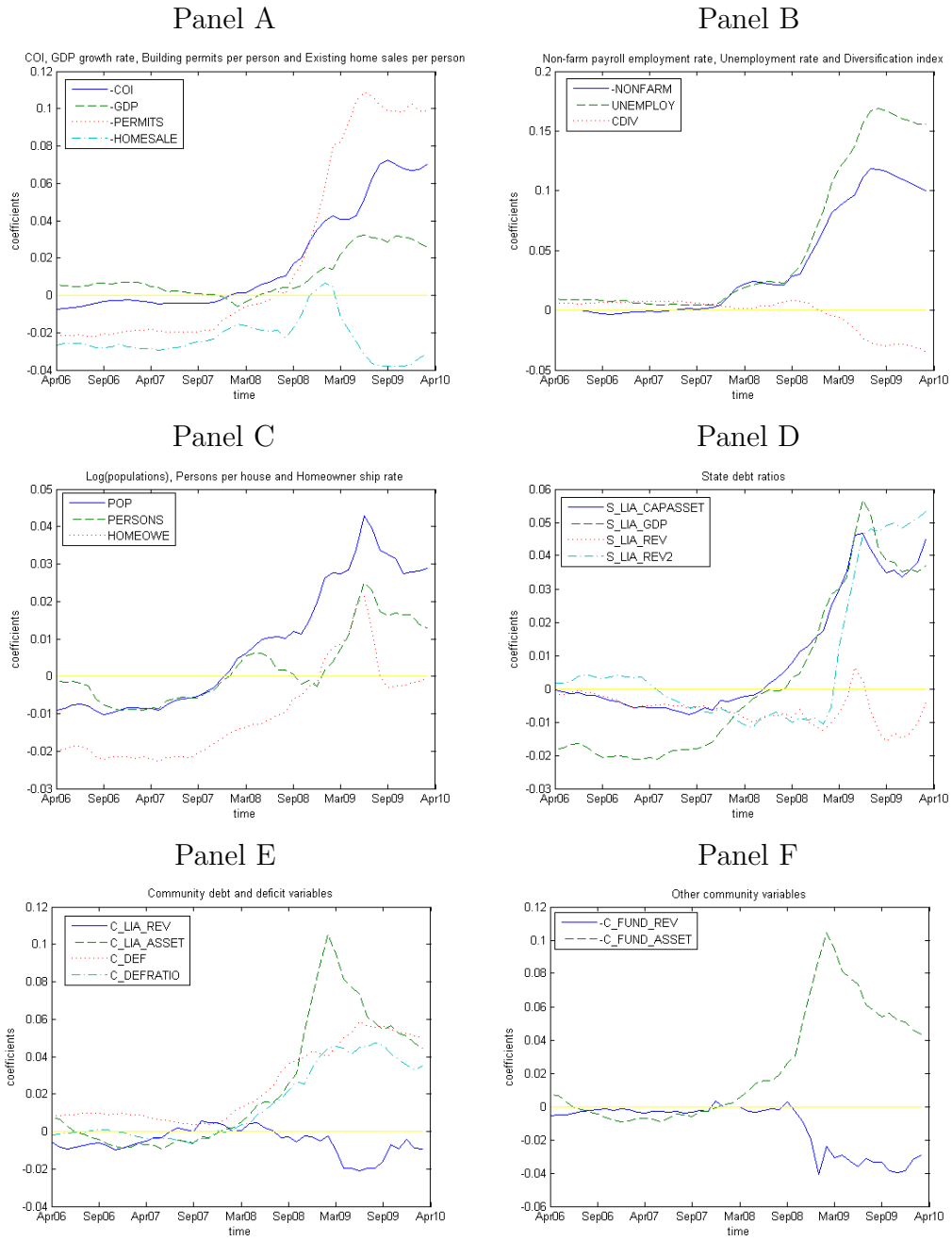
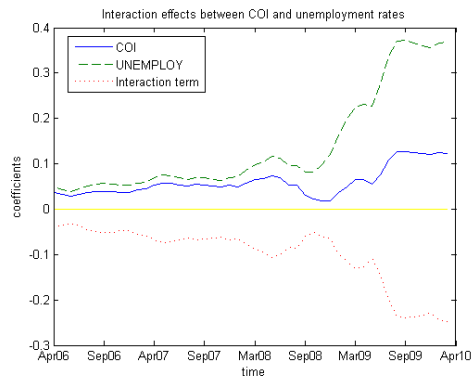


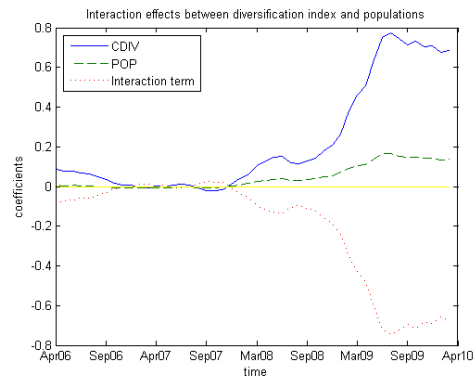
Figure 7: **Monthly bivariate linear regression results**

In the first three pictures, lines plot the coefficients of each standardized pair variables and interaction terms against the time. To smooth the effect, for each variable, I draw $\tilde{\beta}_t$: $\tilde{\beta}_t = 0.85\tilde{\beta}_{t-1} + 0.15\hat{\beta}_t$ for $t = 2, 3, \cdots T$, and $\tilde{\beta}_1 = \hat{\beta}_1$, where $\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_T$ are the linear regression coefficients of each month. The last one picture plots the R^2 of all regressions against the time.

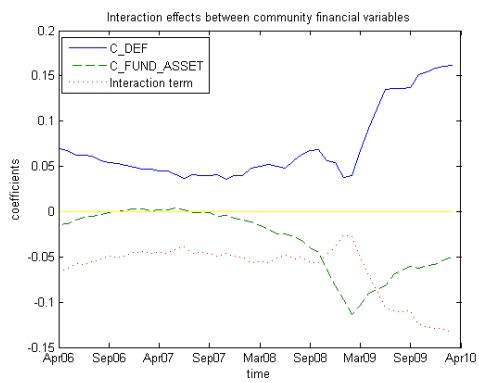
Panel A: COI and unemployment rate



Panel B: Population and CDIV



Panel C: Community financial variables



Panel D: R^2 of regressions

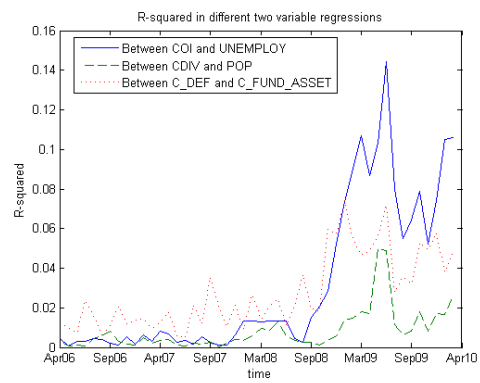


Figure 8: **Monthly linear regression coefficients of variables in multiple linear regressions**

Lines plot the coefficients for each standardized variable against the time. Each month, I run regressions of bond yields on variables list below. To smooth the effect, for each variable, I draw $\tilde{\beta}_t$: $\tilde{\beta}_t = 0.85\tilde{\beta}_{t-1} + 0.15\hat{\beta}_t$ for $t = 2, 3, \cdots T$, and $\tilde{\beta}_1 = \hat{\beta}_1$, where $\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_T$ are the linear regression coefficients of each month.

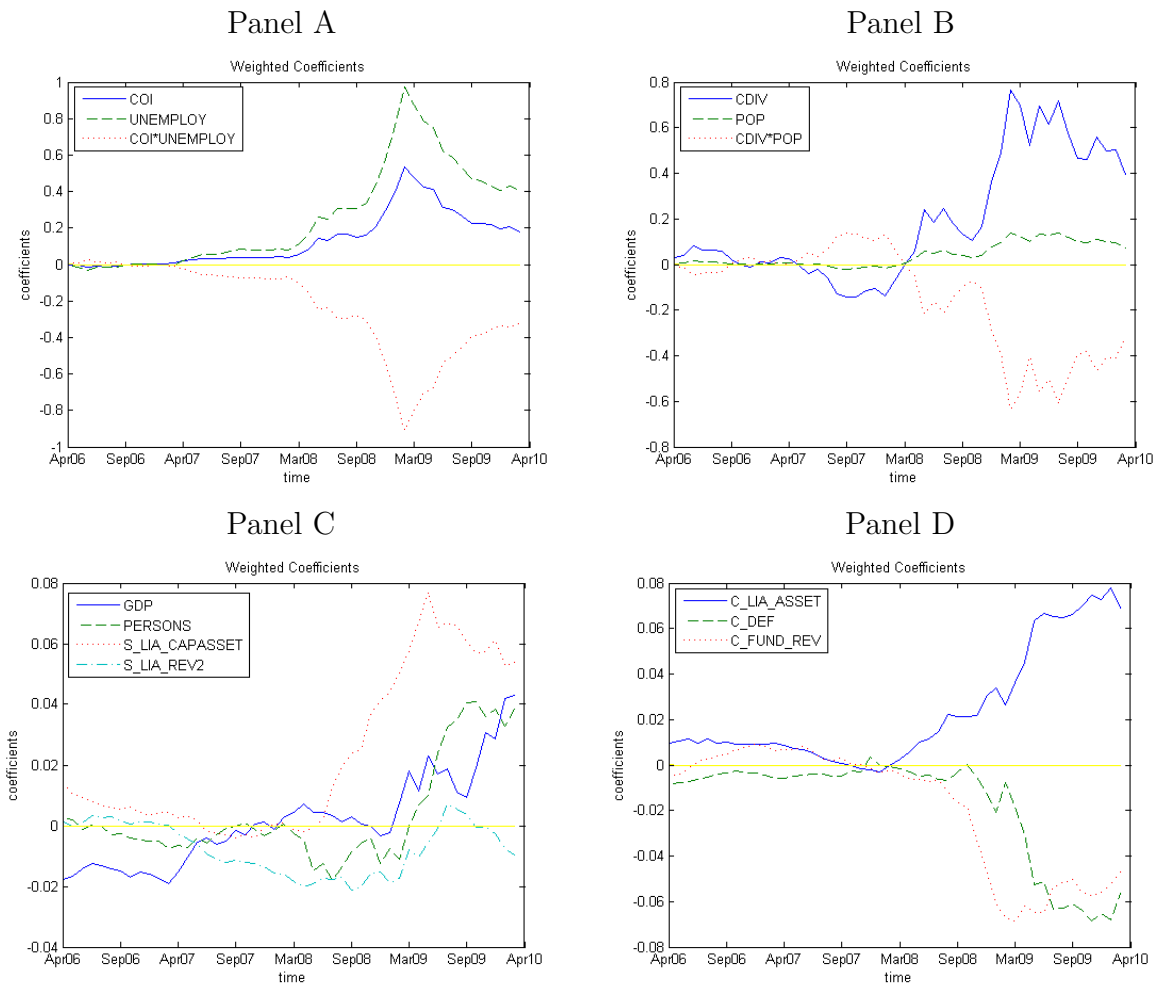


Figure 9: **Pooled Local Polynomial regression results of selected variables** Lines plot the Local Polynomial regression predictions of bond yields against selected economic and all financial variables by pooling all samples. All variables are standardized. The selected economic variables include the coincident index, building permits per person, existing home sales per person, and unemployment rates.

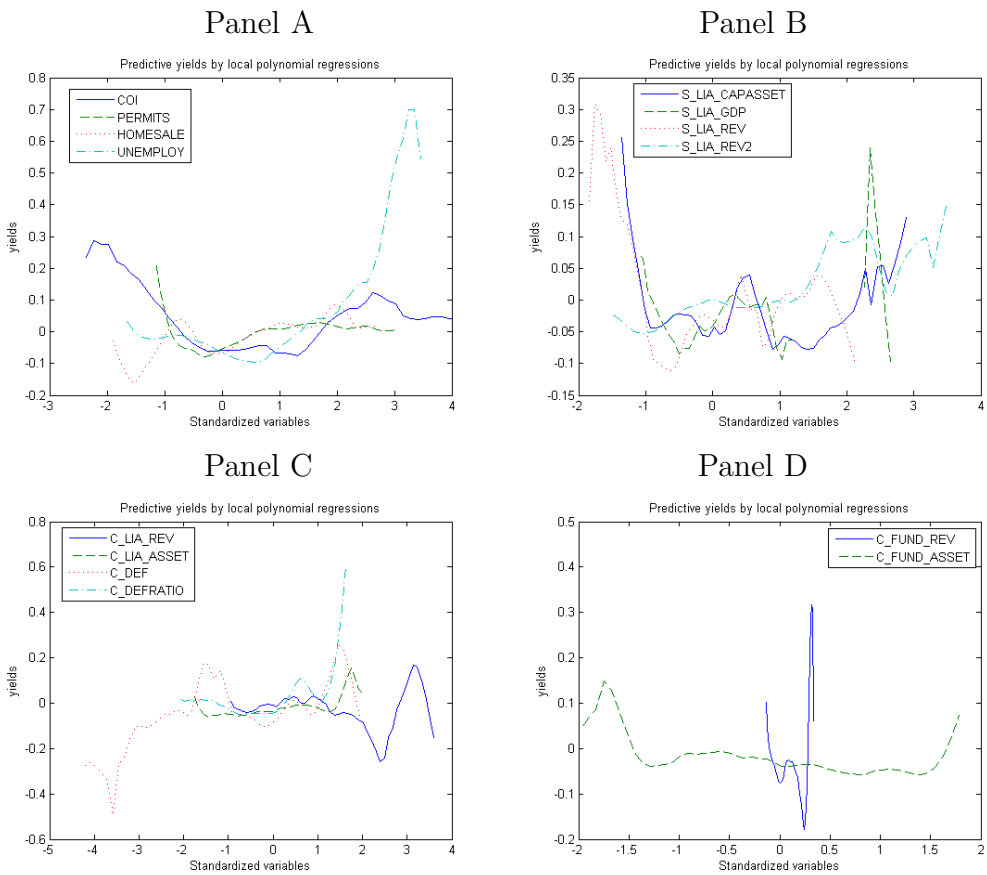


Figure 10: **Time series predictions of unemployment rates and community level total liabilities over total assets by Local Polynomial regressions**
For each month from April 2006 to March 2010 of each variable, the prediction function is $\tilde{f}_t(\cdot) = 0.85\tilde{f}_{t-1}(\cdot) + 0.15\hat{f}_t(\cdot)$ for $t = 2, 3, \cdots, T$ and $\tilde{f}_1(\cdot) = \hat{f}_1(\cdot)$, where $\hat{f}_1(\cdot), \hat{f}_2(\cdot), \cdots, \hat{f}_T(\cdot)$ are the Local Polynomial regression predictive functions of each month. In Panel A and B, lines plot the predictive yields against the time at four selected values of each variable. In Panel C and D, lines plot the predicted yields against variables at three typical months: November 2007, November 2008, and November 2009.

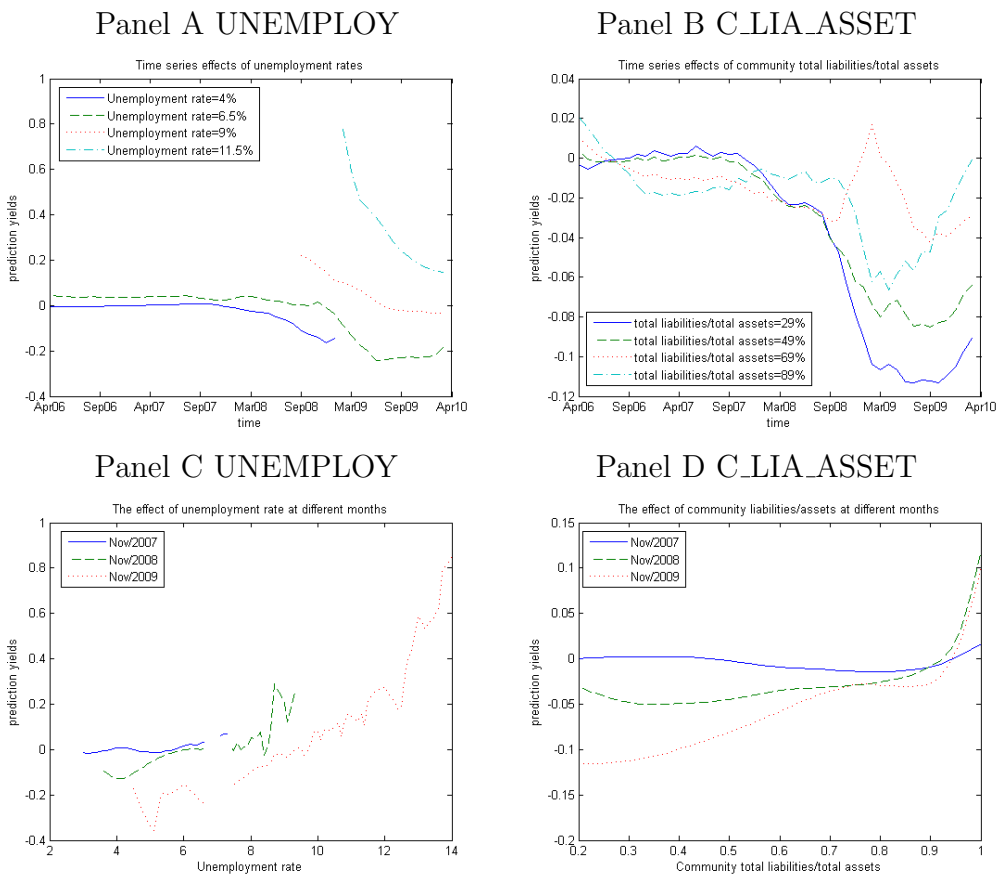
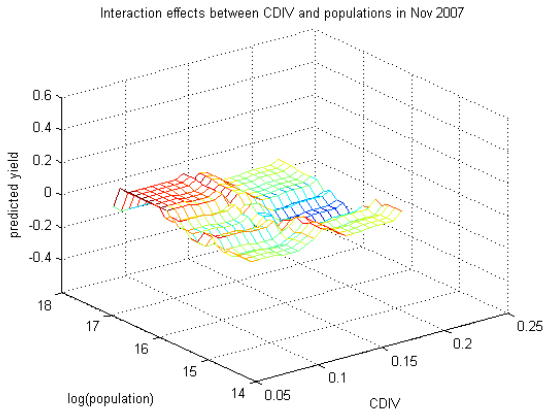
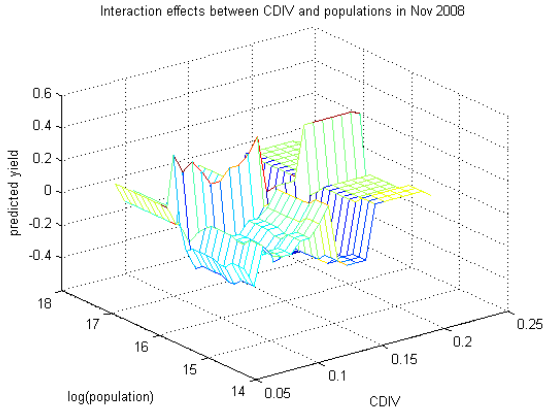


Figure 11: **Interaction effects of state's populations and economic diversification index by Gradient Boosting methods**
Surfaces plot the predicted yields of municipal bonds against state's populations and economic diversification index at November 2007, November 2008, November 2009 by Gradient Boosting methods.

Panel A November 2007



Panel B November 2008



Panel C November 2009

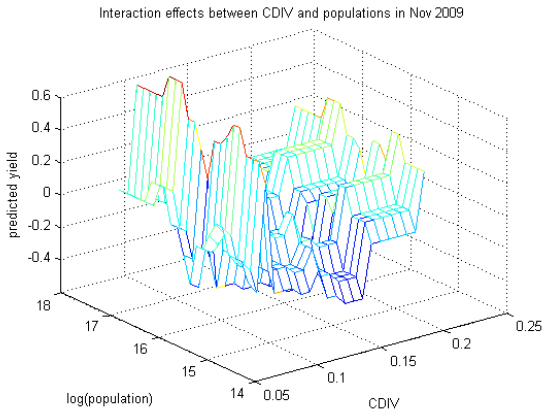


Figure 12: **Percentages of chosen pairs of variables that have the interaction effects for simulation samples**
The line draws the percentages of chosen pairs of variables that have the interaction effects against the boosting steps in the simulation samples.

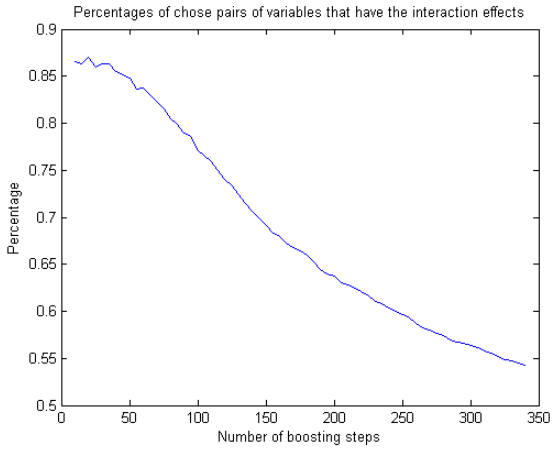
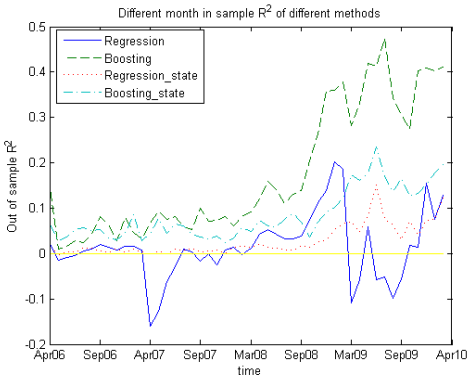
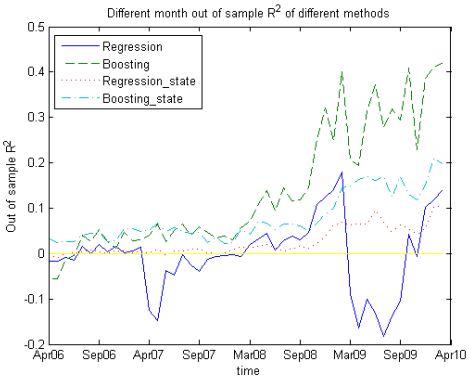


Figure 13: **In and Out of sample R^2 of different predictive methods**
Lines plot the average in sample and out of sample R^2 of different simulations in different predictive methods against the time of each month from April 2006 to March 2010. Methods include: linear regressions including state variables only (solid line); Gradient Boosting including state variables only (dashed line); linear regressions including all variables (dotted line); and Gradient Boosting including all variables (dashed dotted line).

Panel A In-sample R^2



Panel B Out-sample R^2