Determinants of the High School Dropout Rate: Is Cigarette Usage a Factor?
**Introduction:**

What influences high school dropout rates in the United States? Recent economic literature suggests the answer lies with factors, including student mobility, as well as the level of schooling of the parents. However, these factors cannot fully encompass a sound reason for why dropout rates exist, as many other determinants may have more influence on the decision to drop out by students. The following study will analyze other variables, including cigarette usage.

This study will take into consideration common factors prominent economists have concluded as important in determining what affects dropout rates. These issues include the percentage of children under 18 mired in poverty, the youth unemployment rate, real median family income, as well as the influence the percentage of mothers in the labor force may have on the decision to drop out. Though this information has been documented earlier, new information and data sets may lend more insight to this problem.

Although these factors detailed in existing literature may help end the persistent problem of dropping out, this study will put forth a new factor that has yet to be explored in economics. By presenting cigarette usage as an additional factor, this study will be able to determine if dropout rates in high school can be linked to this behavior. As the American Lung Association website mentions, 90% of all smokers begin smoking before the age of 21. Furthermore, each day, 4,000 children under the age of 18 smoke their first cigarette. The numbers of high school children who smoke are intriguing as well. Though the rate has decreased from 36.4% in 1997 to 28.4% in 2007, the number of high school students smoking still is a problem. As the legal age for tobacco consumption has increased to 19, high school students who partake in this behavior are, in fact, committing a crime. Perhaps the penchant to break laws and social codes
would not end at just smoking cigarettes for high school students - the same students may be more susceptible to performing other illegal activities.

Furthermore, cigarette smoking may be a symptom of two underlying causes of behavioral tendencies. That is to say, those who smoke cigarettes have a high time discount rate; they are not forward-looking and will not consider the adverse effects of dropping out of school or continuing to smoke. As a result, those students who smoke may not actually consider future repercussions of dropping out of school or the unfavorable effects cigarette smoking may have on personal health.

As a result, this study will attempt to provide yet another reason why cigarette usage may have more implications than leading an unhealthy life; it could seriously hinder the ability to graduate high school, as students who break the law by smoking may indulge in other illegal activities, as if to find a haven from the rigors of school. Another reason to utilize cigarette usage as an explanatory variable of the high school dropout rate is that perchance these students who smoke are actually not forward-thinking; cigarette usage may simply be an indication of students who do not consider the effects of dropping out of high school.

**Literature Review:**

A seminal paper in the field of educational economics, Coleman (1988) first introduces the concept of social capital and its influence on human capital. He finds a direct correlation between social capital and dropout rates. To illustrate what social capital is, he presents several examples and then juxtaposes this concept with human capital. He defines human capital as skill sets acquired by some to adjust and act in new ways. Social capital, he maintains, occurs through changes in relations among people. It is not a tangible trait, but rather identifies aspects
of society. Within the family, social capital includes relationships between the parent and the child.

When analyzing the data, Coleman (1988) finds Catholic schools and other private schools have lower dropout rates. He attributes this to more adult supervision within the school, which he interprets as reflective of a stronger teacher-parent relationship.

In addition, there is evidence that mobile students graduate at a lower rate than those who do not move. Coleman (1988) concludes weaker ties are a result of higher turnover and mobility. The papers by Teachman (1996) and Coleman (1988) confirm that the interactions between parent and child negatively affect dropout rates. Student mobility and where they go to school has a strong influence as well.

Teachman, Paasch, and Carver (1996) further explore the importance of social capital and the effect it has on dropout rates. The authors define social capital as the interactions between parents and children, as well as the mobility of these students. They posit that a more close-knit relationship will provide positive role models to children, encouraging further educational achievement. Taking data from the National Educational Longitudinal Study, the authors collected information from eighth graders in 1988. Follow up studies were performed in 1990 and 1992. This study includes data from those who dropped out from eighth to tenth grades. Parents and students alike were given surveys to fill out concerning their relationships with each other. Their results find that living with a divorced father reduces social capital, as does a child living with a stepparent, all due to reduced parent-school interaction. Furthermore, their research also finds that one reason for dropping out of high school is high student mobility. Attending Catholic school, however, reduces the chance of dropping out. They suggest this is because of a more close-knit environment by attending this religious school. There is a strong, positive
relationship between Catholic schools and parent-school connectivity. One fact remains true in most recent economic literature – student mobility has a profound negative effect on graduation from high school.

Clearly, another contribution to dropping out is student mobility. South, Haynie, and Bose (2005) study some underlying correlates that affect mobility, including parent-child relationships, peer networks, academic performance, school attachment, and psychological well being. The authors study these factors to gain a better understanding of why students move from school to school, which adversely affects their graduation rates. They collect data from the National Longitudinal Study of Adolescent Health in an attempt to measure the effect student mobility has on dropping out. They try to, at first, determine what may cause mobility, and then secondly how this mobility affects the dropout rates.

South, Haynie, and Boses’ results suggest that mobile students are almost twice as likely to dropout when compared to their non-mobile counterparts. They find mobile students show lower quality relationships with their parents. Parents also reported that they were less involved in their child’s life. Furthermore, this study finds that mobile students have smaller but denser peer networks, participating in fewer extracurricular activities and maintaining a lower GPA than non-mobile pupils. In sum, this article finds that due to several underlying causes, student mobility has a drastic effect on dropout rates.

Lillard and DeCicca (2000) supply information regarding the effect of minimum mandatory coursework on high school dropout rates. Here, the authors examine a change in mandatory coursework to find an influence on dropout rates. The authors propose a cost-benefit analysis of increasing the minimum standards to graduate, and conceptually determine “…for the individual on the cusp of dropping out, an increase in graduation requirement of one unit must
impose greater costs than benefits.” The “unit” term is the Carnegie unit, which reflects the time spent on studying a subject for a year. Lillard and DeCicca (2000) use individual and aggregate data from surveys of students. With aggregate data, they include those who receive GEDs as dropouts, while those who receive GEDs in the individual data are not considered dropouts. The aggregate data shows dropout rates are positively correlated with coursework changes; if the minimum coursework increased, so did the number of dropouts.

The dropout rate also serves as a measure of a school’s success. As Rumberger and Thomas (2000) mention, “…school dropout rates - the proportion of students who drop out of a school - have been used as a measure of school performance in a number of studies on school effectiveness.” Three main conclusions were discovered. The first was that turnover rates were higher than dropout rates. This difference in distribution can be accounted for by the variation in student characteristics, which is the authors’ second conclusion. This conclusion is the quintessential aspect in this study, as it shows that the relationship between dropout and mobility rates can be influenced by the educational background of a student. Ethnicity and academically at risk students have a profound impact on the relationship between mobility and dropout rates as well. Their third and final conclusion maintains that schools actually have an influence on a student’s academic success. Rumberger and Thomas go on to declare that mobility and dropout rates adversely affect test scores, and to reduce this, schools release more students to increase their “legitimacy”.

One of the most often discussed sources of the high school dropout rate is poverty. Orthner and Randolph (1999) analyze how parental work and welfare receipt impact the decision for a pupil to drop out of high school. This study used longitudinal monitoring of parents and children who were a part of a welfare program. The authors use event history, which is an
umbrella term used to describe a set of procedures for time-series data analysis, to further their results. Orthner and Randolph (1999) note that event history should be reviewed to see if a child’s environment would be affected if a parent enters the labor force after a period of unemployment. More a propos to the subject at hand, the authors declare, “Event history analysis allows researchers to determine whether the pattern of parental employment in the semester immediately preceding students’ decisions to drop out was a predictive indicator.” Their results prove to be unexpected. The authors find that students growing up in poverty are more susceptible to dropping out of school, regardless of the work patterns or welfare programs their parents may be a part of. However, families mired in poverty, with parents who experience more consistent work or transition from welfare to the labor force and do not return, have children who will graduate more often than those whose parents do not work consistently or are on welfare. These results are countered with the finding that even for poor students whose parents work more consistently, graduation rates are still lower when compared to students coming from a middle-class family. This shows that poverty quite possibly may be considered the quintessential factor that determines the dropout rate, as poorer students may feel the monetary incentive of working may be more important than finishing school.

Vartanian and Gleason (1999) also study poverty effects on educational achievement, but do so in a slightly different manner from Orthner and Randolph (1999). Here, Vartanian and Gleason (1999) present information regarding the neighborhoods of black and white students, respectively. Their study focuses on how neighborhoods of these students will affect educational attainment. Vartanian and Gleason (1999) propose four theoretical models, which include the social isolation theory, the epidemic theory, the relative deprivation theory, and finally the personal utility maximization theory. According to the authors, the first two theories suggest
that living in poorer neighborhoods would significantly decrease the rate of graduation, as there would be a lack of positive adult role models in the community. The epidemic theory is a special case of the social isolation theory; it argues that neighborhood conditions do not matter much in affecting outcomes until conditions reach epidemic proportions. As a result, living in a neighborhood where few adults are out of work or living in poverty may not affect educational progress. On the other hand, living in a neighborhood where almost all of the inhabitants are unemployed, poorly educated, and living in poverty will make a young person less likely to graduate from high school or college. The relative deprivation theory maintains that living in a wealthy neighborhood would increase the rate of dropout, as disadvantaged youths would have to compete with wealthier children who have more resources at home, while the personal utility maximization theory insists that neighborhood conditions have no effect on dropout rates.

Vartanian and Gleason (1999) used sampled data of the Panel Study of Income Dynamics from 1968 to 1992, with the minimum age of 14 for their study. They found that whites dropped out at a rate of 8%, while blacks were at 13%. Upon further analysis, the authors come to the conclusion that those living in wealthier neighborhoods are more likely to graduate. The social isolation theory holds for black students as well, as poorer black students were more susceptible to dropping out. Wealthier blacks were not affected at all. For whites, there is a positive effect of the wealth of the neighborhood on college graduation. There is no such effect on high school graduation. The social isolation theory does not fit well with the results for white students. Clearly, poverty and neighborhood surroundings have an effect on high school dropout rates, as is evidenced by this paper as well as Orthner and Randolph (1999).

Haveman, Wolfe, and Spaulding (1991) propose information regarding the effects of family and economic circumstances on high school completion. The work hints at mobility as
being a large factor in determining the rates, as well as relationships between the parent and the child. The authors use data from the University of Michigan’s Panel Study of Income Dynamics for the year 1987, with subjects ranging from 19 to 23 years of age. The information relevant to each subject included personal characteristics, characteristics of the immediate family and of the grandparents, income, parents’ labor force participation, welfare receipt, as well as if they lived in an urban area or not. These factors were combined into six groups, which include gender, child’s family position, childcare time, family stress, family educational orientation, and family economic circumstances.

Parental education and the extensiveness of the mother’s work have positive effects on graduation, while the number of siblings, being on welfare during a child’s adolescent years, and the mobility of the student has negative effects on the completion of high school. They explain their results by stating if there is an increase of mothers in the labor force producing a source of income, children will not feel the need to leave school to find a job throughout his/her high school career. The results explicitly show that student mobility during the ages of 12 to 15 produces severe detrimental effects on graduation. Haveman, Wolfe, and Spaulding’s work in 1991 serve as corroborating evidence that mobility, education, and poverty level should be considered when trying to determine what affects high school drop out rates.

Sandefur, McLanahan, Wojtkiewicz (1992) put forth information regarding graduation rates based on solely the marital status of the parents during the child’s adolescent years. Perusing longitudinal data from the 1979-1985 NLSY, the writers performed research on the effect family types and structure has on graduation rates. Sandefur, McLanahan, and Wojtkiewicz observe:

Individuals who live apart from one or both parents when they are growing up are less likely to graduate from high school, more likely to work at low-wage jobs,
and more likely to form unstable families themselves than individual who grow up with both biological parents.

The group also examines relationships with a stepparent, and even no parent families. They measured 11 variables, including race, college attendance, the extent of high school education (or GED receipt), parental education, sex, family structure at age 14, the structure from ages fourteen to seventeen, parental education, and number of siblings. They also measured the adjusted family income, self-esteem (through a brief number of options), and finally perceived attitudes. Their results show that at age 14, if the child was growing up in a single parent or a parent with a stepparent home, the student was more likely to drop out when compared to a fellow 14 year old growing up with both biological parents. Children who encountered a disruption of stability from their parents, i.e. a divorce, the child was more likely to drop out, even when compared to children of that age range who come from a single parent or a biological parent and a stepparent home. Furthermore, the effect family structure and type has on the receipt of a GED is similar to the effect it has on the receipt of a high school diploma at the conclusion of the tenure. Clearly, this study proves that familial disruption has a negative effect on the graduation of high school.

Kenkel, Lillard and Mathios (2006) explain the relation between the receipt of a GED or a high school diploma and the effects each have on the two leading causes of preventable death – obesity and smoking. This study goes on to describe if a difference exists between these two factors in relation to those who enter a high school equivalency program as opposed to those who simply get a degree in high school after completion. The results in Kenkel, Lillard and Mathios (2006) show that there is no statistically significant difference between the GED recipients and
the high school diploma recipients when it comes to obesity. However, those who go through high school and procure a degree smoke less than those who obtain their GED.

This data suggests that education plays a primary role in health awareness. Though this information is vital in trying to raise a healthier generation, the information lacking in this paper serves as a starting point for my study. Kenkel, Lillard, and Mathios (2006) show how education can contribute to the healthy habits of individuals. This upcoming study will show how smoking, along with other factors, has an effect on the tendency to drop out. Since Kenkel, Lillard and Mathios (2006) perform analyses and provide results that show education has a profound effect on health returns, perhaps health related issues, among other variables, will have an effect on the dropout rate.

The combination of these papers provides the groundwork for the following study. By determining what factors can attribute to high school dropout rates, this study can try to show additional factors that have never been brought to the surface. Clearly, searching for more information as to what has been increasing the dropout rate will serve as an instrument to reduce this phenomenon.

**Data and Methods:**

Utilizing data from a multitude of sources, this study provides a preliminary model which includes six variables, hypothesized to explain the high school dropout rate. The variables include the youth unemployment rate, cigarette usage, the percentage of mothers in the labor force with children aged 6 – 18, the percentage of children in poverty status under 18, real median family income, and finally the amount of federal school revenue per student. The data
are from the United States and range from the year 1970 to 2001, as 2001 is the most current year in which information for all of the variables exists.

Data for the percentage of children under 18 in poverty and information for the dropout rate, the dependent variable, are from the United States Census site. Real household median income data was found perusing the United States Census site as well. Information regarding youth unemployment rates and the percentage of mothers in the labor force was procured from the Bureau of Labor Statistics website. Cigarette usage data was available through “Monitoring the Future”, a website created by the University of Michigan – Ann Arbor to put forth information regarding cigarette usage. In particular, I used daily usage of cigarettes for 12th graders, as the lifetime and thirty-day usage variables did not seem to be relevant to my study. By monitoring daily use, the data obtained from this survey would be more inclusive of daily smoking behavior of high school students. This means that no information for potential lifetime usage would be included in this data, illustrating the immediate effect of cigarette usage on the student. Furthermore, this survey included information from roughly 17,000 students each year, from 1970 to 2001. This statistic was given in percentage form. Finally, the data for federal school revenue per student was found on the National Center for Education Statistic’s website.

Though cigarette usage has declined over the past few years, more than one in four high school students still engage in this activity. As a result, one of the hypotheses of this study is that the high school dropout rate may be correlated with students engaging in this activity. The coefficient would be positive, as the higher the number of students who smoke cigarettes, the higher the number who will dropout. This is in part due to the notion that those students willing to commit the crime of under-aged smoking may also participate in other illegal activities, allowing students to stray from high school.
Another regressor is the percentage of mothers in the labor force. This explanatory variable is proposed because if there is an increase of mothers in the labor force producing a source of income, children will not feel the need to forego their education to maintain a job throughout his/her high school tenure, as Haveman, Wolfe, and Spaulding (1991) mention. As a result, a negative correlation between this variable and the dropout rate is expected.

An increase in the percentage of children under the age of 18 in poverty would lead to an increase in the high school dropout rate. Following the literature, the more children in poverty, the more the high school dropout rate would increase. Furthermore, an increased amount of federal school revenue per student would lead to a decreased dropout rate. As more resources are geared towards students, this could perhaps be an incentive to remain in school, as resources are available to help students who are not succeeding. As a result, a negative relationship between this variable and the dropout rate would exist; as there is more per student revenue, the dropout rate should decrease.

An increase in the real median family income would also see a decrease of the high school dropout rate. As the literature suggests, increased familial income would allow a family to locate to a better school district, increasing the probability of a child graduation. As a result, a negative relationship should be seen between this regressor and the high school dropout rate.

Finally, the youth unemployment rate may dictate whether or not a student would remain in school. If the rate is low, perhaps dropout rates would increase due to the higher probability of finding jobs. Using extra income as an incentive, the coefficient for this variable should be negative; as the youth unemployment rate increases, a student would find it harder to obtain a job, and therefore would stay in school. The dropout rate would, as a result, decrease. In fact, this study at first utilized the unemployment rate instead of the youth unemployment rate.
However, after scrutinizing over regression analysis, the youth unemployment rate functions in the same way as the unemployment rate, but is more statistically significant than the regular unemployment rate. As a result, this study only considers the youth unemployment.

Because the high school dropout rate is a percentage, for regression purposes, the logged version of the dropout rate is used. This was done to reduce the variability of the dependent variable, generating consistent regression estimates. Reduced variation in the extremes may also help reduce heteroskedasticity.

As a result of the variables used, a general model can be proposed as such:

$$ DR = \beta_1 \cdot SR + \beta_2 \cdot CU + \beta_3 \cdot ML + \beta_4 \cdot P + \beta_5 \cdot YUR + \beta_6 \cdot INC + \beta_7 \cdot YEARS + E $$

where DR is equal to the log of the dropout rates, SR is the federal school revenue per student, CU is cigarette usage, ML is the percentage of mothers in the labor force, P is the percentage of children in poverty, YUR is the youth unemployment rate, INC is real median family income, and YEARS is simply years. From this model, other models are necessary to create a functional, statistically significant model. In a later section of this study, we will see the effect regression analysis has on the general model.

Model 1 in Table 2 [see end] illustrates the original, general model used at first. This reflects the variables I posited would lend themselves to the dropout rate. However, we can see that most of the variables are statistically insignificant. Though the variables are not statistically significant, the F statistic of 18.54 suggests that the model, as a whole, is significant. This is clearly one of the tell tale signs of multi-collinearity. Though multi-collinearity will not make the predictors biased in any way, correcting this problem is still necessary. A way to correct this problem is to acquire several more data points and observations, but such information was unavailable. As a result, another approach was necessary to correct this problem. Explanatory
variables are added in numerous combinations so as to depict their relationship to the high school dropout rate.

Table 1 presents a correlation matrix regarding all of the variables, as well as the years the data ranges from. We can see that the variable of mothers in the labor force is very strongly correlated with years, as is the federal school revenue per student variable and real median family income. As a result, we can consider these variables to be trend variables. The subsequent models will use de-trended versions of the three variables, so that a model can be proposed that includes these variables. These regressors, as a result, could be utilized in finding a statistically significant model to explain the high school dropout rate. Perusing this table, we also see a strong correlation between the youth unemployment rate and the unemployment rate. This is also one of the reasons the unemployment rate was excluded, as it was very highly related to the youth unemployment rate. The youth unemployment rate was used, as it would be more indicative of the behavior of high school students in terms of obtaining a job. This seems more relevant to the study, as opposed to examining a general unemployment rate.

**Results:**

Clearly, model 1, table 2 illustrated a model that is statistically significant, but includes variables that are not statistically significant. In an attempt to disentangle from the model laden with multi-collinearity (model 1, table 2), a model that includes de-trended variables is used. These variables include the percentage of mothers in the labor force and real median household income. Previously, federal school revenue per student was defined as a trend variable, but in over 90 test runs, this variable was not statistically significant. As a result, model 2, table 2 reflects only two de-trended variables. From this model, the results find that the percentage of
mothers in the labor force as well as real median household income, when regressed in conjunction with years, are statistically significant. The adjusted R-squared value of .6796 is quite high; lucidly, these variables are good indicators of the high school dropout rate.

Model 3, table 2 reflects an equation with the non-trend variables. Of course, these variables are cigarette usage, the youth unemployment rate, and the percentage of children in poverty. From this model, one can discern only the unemployment rate and the children in poverty are significant variables. The model has an adjusted R-squared value of .6059, which is a number that does not reflect a model that can explain the dependent variable well.

Model 4 includes only the unemployment rate and the children in poverty. One can see that both of these variables are significant, and the R-squared value is at .6157. This model, as a result, contains two variables that can be declared as significant when it comes to predicting the dropout rate. A fifth model reflects the introduction of the de-trended variable of mothers in the labor force, along with the youth unemployment rate and the percentage of children in poverty. We can see that in model 5 all of the variables used are insignificant. The t statistic for the percentage of children in poverty sharply declines from -6.61 in our third model to -1.61. This illustrates that perhaps the variable of mothers in the labor force may have picked up part of the variability of the children in poverty variable.

Model 6 refers to yet another model that includes the de-trended variable of mothers in the labor force, but we also see the introduction of cigarette usage in conjunction with the percentage of children in poverty and the youth unemployment rate. We find that none of the variables are significant, but the t value of -.93 for the percentage of children in poverty and .64 for the youth unemployment rate has decreased yet again. This intimates that perhaps the introduction of cigarette usage may have picked up some of the variability of the youth
unemployment rate variable, similar to the effect that the mothers in the labor force variable had in the previous model.

In model 7 we see cigarette usage, the youth unemployment rate, and mothers in the labor force as the variables we test for significance. Cigarette usage, as well as years, is statistically significant. The R-squared value is at .6545. Model 8 measures the significance of cigarette usage, mothers in the labor force, years, and the percentage of children in poverty to see if there is any significance to this model. We see that cigarette usage is significant once again, as well as years. However, the percentage of children in poverty and mothers in the labor force are still not significant. The R-squared value is .6604. Clearly, we are slowly progressing to a model that may adequately represent an explanation for our dependent variable. Model 9 reflects three variables, cigarette usage, years, and the percentage of mothers in the labor force. We finally encounter a model where these variables are significant; the R-squared value is also fairly high, at a .665 level.

Interestingly, whenever cigarette usage is statistically significant, a positive relationship is found between this explanatory variable and the dependent variable. This result was predicted earlier in the study. Furthermore, when the percentage of mothers in the labor force is significant, we see a negative relationship to the dropout rate; this result is similar to what Haveman, Wolfe, and Spaulding (1991) found as well.

However, when the percentage of children in poverty is statistically significant, a negative correlation is noted to the high school dropout rate. This directly opposes the findings in Orthner and Randolph (1999). As previously mentioned, they found that those in poverty were more likely to dropout. The phenomenon seen in this study’s results can be explained by the notion that perhaps these students in poverty do not want to live in squalor any longer. As a
result, these students may want to graduate high school, procure a diploma, and try to maintain a steady job.

The results also show that when the youth unemployment rate variable is significant, a positive relation is seen, relative to the dependent variable. That is, as the youth unemployment rate goes up, so too does the dropout rate. These results cannot be explained.

Previously, we found multicollinearity in our original model. Other problems, like heteroskedasticity and autocorrelation may exist as well. Utilizing the White Test, the results show that heteroskedasticity does not exist for any model.

As far as autocorrelation is concerned, a utilization of the Breusch-Godfrey test shows that there is no positive nor is there any negative correlation in any of the models detailed in my study. As a result, the findings are significant and error free.

In conclusion, the inclusion of the cigarette usage variable tends to yield significant results. Model 9 includes both cigarette usage and mothers in the labor force in conjunction with years; these variables are statistically significant. It can be clearly determined that these explanatory variables certainly are good predictors of the high school dropout rate. The results show that cigarette usage and mothers in the labor force are strong correlates for the high school dropout rate.

**Conclusion:**

Recent economic literature has shown a multitude of variables and factors that may attribute to the high school dropout rate. These factors include student mobility, the relationship between parents and teachers, and the family structure of a student’s immediate family. However, the addition of cigarette usage, the youth unemployment rate, the percentage of
children in poverty, the percentage of mothers in the labor force, the youth unemployment rate, real median family income, and finally the federal school revenue per student provides new insight to what other factors may be predictors of the high school dropout rate. After realizing that a general model utilizing all 6 variables proved to be significant, even though some of the individual variables were not, a series of altered models have shown that cigarette usage is a prime predictor of the high school dropout rate. Perhaps engaging in this now illegal activity for high school students, as the legal age to use tobacco products has risen to 19, may encourage other activities that allow students to stray from the rigors of school. Other de-trended variables in conjunction with cigarette usage, like the percentage of mothers in the labor force, also can be deemed as a good predictor of the dependent variable. As the literature shows, education has a profound effect on the percentage of those who use cigarettes. Now, according to this study, one can see that cigarette usage can have an effect on educational attainment.

References


