A study of birth weight as a predictor of cognitive ability in childhood: applications of Loess regression and generalized propensity score methods

Xiaoyuan Tan

University at Albany, State University of New York, xt636168@hotmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Developmental Psychology Commons, Educational Psychology Commons, and the Quantitative Psychology Commons

Recommended Citation

https://scholarsarchive.library.albany.edu/legacy-etd/778

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive.
Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
A STUDY OF BIRTH WEIGHT AS A PREDICTOR OF COGNITIVE ABILITY IN CHILDHOOD: APPLICATIONS OF LOESS REGRESSION AND GENERALIZED PROPENSITY SCORE METHODS

by

Xiaoyuan Tan

A Dissertation

Submitted to the University at Albany, State University of New York in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

School of Education

Department of Educational and Counseling Psychology

2012
A Study of Birth Weight as a Predictor of Cognitive Ability

in Childhood: Applications of Loess Regression and

Generalized Propensity Score Methods

By

Xiaoyuan Tan

COPYRIGHT 2012
Acknowledgements

I feel deeply grateful to many people. First and for most, very special thanks are given to Dr. Robert Pruzek for his continuous supervision, advisement, and support for my dissertation research and for my entire graduate-level study. I have made the rightest decision, following him learning statistics and counting on him for any other important decision I had to make; to Dr. David Dai for giving me the freedom to explore what I am interested in but providing support whenever I need; to Dr. Joan Newman for having served as a mentor and a dear friend particularly in those dark moments; and to Dr. Bruce Dudek for his insightful comments on my dissertation research.

My gratitude is also extended to Yi Sun, Kuangnan Xiong, and many of my colleagues in both the educational psychology program and the biostatistics program. I thank them for the generous help they offered to me when I was struggling through my study and my research.

Finally, I would like to thank my parents, Wanyou Tan and Yongbi Gu, for keeping themselves happy and healthy when I was away pursuing my dream. And thank my brother, Hao Tan, my sister in law, Xufang Zhang, and my nephew, Tian Tan, for taking good care of my parents and themselves.
Table of Contents

Acknowledgements...........................................................................................................iii

Table of Contents.............................................................................................................iv

List of Tables..................................................................................................................vii

List of Figures.................................................................................................................xi

Abstract.............................................................................................................................xviii

Chapter 1 Introduction......................................................................................................1

  Statement of the Problem...............................................................................................1

  Major Objectives and Research Questions.....................................................................6

Chapter 2 Literature Review............................................................................................9

  Birth Weight and Cognitive Ability in Childhood.........................................................9

    Low birth weight and cognitive ability in childhood..................................................10

    Normal birth weight and cognitive ability in childhood............................................11

    Confounding and confounders....................................................................................13

    Potential avenues for future study suggested by the literature..................................19

  Loess Regression............................................................................................................20

    Mathematical rationale...............................................................................................21

    Graphic visualization of results..................................................................................23

    Advantages and disadvantages...................................................................................24

  Propensity Score Methods to Control for Confounding..............................................25

    General introduction.................................................................................................25

    Strengths and limitations...........................................................................................30

Chapter 3 Methods.........................................................................................................32
Samples…………………………………………………………………………………………32
  National Longitudinal Survey of Youth 1979-Child…………………………..32
  Panel Study of Income Dynamics-Child Development Supplement…………..33
  Comparisons of the two samples……………………………………………………..34
Data Preparation……………………………………………………………………………..36
  Adjustment for age differences.................................................................37
  Generation of cognitive composite score..................................................37
  Imputation of missing data........................................................................37
Data analysis.................................................................................................38
  Loess regression to depict the relationship of birth weight with cognitive ability38
  Generalized propensity score methods to control for confounding.............39

Chapter 4 Results ..........................................................................................43

Results for the NLSY79-C............................................................................43
  Loess regression to depict the relationship of birth weight with cognitive ability43
    Locally linear regression of cognitive composite score on
    birth weight and gestational age..............................................................50
    Simple locally linear regression of cognitive composite score on
    birth weight..........................................................................................55
  Generalized propensity score methods to control for confounding..............62
    Results for the non-black-non-Hispanics whose gestational ages are normal.65
    Results for the blacks whose gestational ages are normal........................77
Results for the PSID-CDS...........................................................................85
  Loess regression to depict the relationship of birth weight with cognitive ability85
Locally linear regression of cognitive composite score on birth weight and gestational age

Simple locally linear regression of cognitive composite score on birth weight

Generalized propensity score methods to control for confounding

Results for the whites whose gestational ages are normal

Results for the blacks whose gestational ages are normal

Chapter 5 Summary of the Results and Discussion

Summary of the Results

Discussion

Limitations and Recommendations for Future Research

References

Appendix
List of Tables

Table 4.1  Descriptive Statistics for Birth Weight, Gestational Age in Weeks, and Cognitive Composite Score for the Non-Black-Non-Hispanic (n=781) and the Black (n=653) Children in the NLSY79-C .................................................................48

Table 4.2  Descriptive Statistics of Covariates for the 654 Non-Black-Non-Hispanic Children in the NLSY79-C Whose Gestational Age Is in the Normal Range from 37 Weeks to 42 Weeks.................................................................67

Table 4.3  Correlations of Covariates with BIRTH and Cognitive Composite Score for the 654 Non-Black-Non-Hispanic Children in the NLSY79-C Whose Gestational Age Is in the Normal Range from 37 Weeks to 42 Weeks.................................69

Table 4.4  Means of Cognitive Composite Score, Counts of Children, Estimated Means after Adjusting for the Quintiles of the Balancing Score and Their 95% Confidence Intervals for the 557 Non-Black-Non-Hispanic Children in the NLSY79-C Whose Gestational Ages Are in the Normal Range from 37 Weeks to 42 Weeks and Whose Balancing Scores Are in the Overlapping Region from 0.6 to 0.95........................................76

Table 4.5  Descriptive Statistics of Covariates for the 551 Black Children in the NLSY79-C Whose Gestational Ages Are in the Normal Range from 37 Weeks to 42 Weeks.................................78
Table 4.6  Correlations of Covariates with BIRTH and Cognitive Composite Score for the 551 Black Children in the NLSY79-C Whose Gestational Ages Are in the Normal Range from 37 Weeks to 42 Weeks……………………………………..79

Table 4.7  Means of Cognitive Composite Score, Counts of Children, Estimated Means after Adjusting for the Quintiles of the Balancing Score and Their 95% Confidence Intervals for the 493 Black Children in the NLSY79-C Whose Gestational Ages Are in the Normal Range from 37 Weeks to 42 Weeks and Whose Balancing Scores Are in the Overlapping Region from 0.65 to 0.95………………………………………………………...84

Table 4.8  Descriptive Statistics for Birth Weight, Gestational Age in Days, and Cognitive Composite Score for the white (n=931) and the black (n=832) children in the PSID-CDS……………………….89

Table 4.9  Descriptive Statistics of Covariates for the 833 white Children in the PSID-CDS Whose Gestational Age Is in the Normal Range from 259 Days to 294 Days…………………………………….106

Table 4.10  Correlations of Covariates with BIRTH and Cognitive Composite Score for the 833 white Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days…………………………………….107
Table 4.11  Means of Cognitive Composite Score, Counts of Children, Estimated Means after Adjusting for the Quintiles of the Balancing Score and Their 95% Confidence Intervals for the 804 White Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days and Whose Balancing Scores Are in the Overlapping Region from 0.55 to 0.9…………………………………………………………...112

Table 4.12  Descriptive Statistics of Covariates for the 741 Black Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days…………………………………………………………113

Table 4.13  Correlations of Covariates with BIRTH and Cognitive Composite Score for the 741 Black Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days…………………………………………………………114

Table 4.14  Means of Cognitive Composite Score, Counts of Children, Estimated Means after Adjusting for the Quintiles of the Balancing Score and Their 95% Confidence Intervals for the 643 Black Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days and Whose Balancing Scores Are in the Overlapping Region from 0.55 to 0.9…………………………………………………………...119

Table 5.1  Descriptive Statistics for Birth Weight and Gestational Age for the Non-Black-Non-Hispanics and for the Blacks in the NLSY79-C and for the Whites and for the Blacks in the PSID-CDS…………………………………………………………………………………………122
<table>
<thead>
<tr>
<th>Table 5.2</th>
<th>Comparisons of Confounding Covariates between the NLSY79-C and the PSIDCDS</th>
<th>130/131</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 5.3</td>
<td>Comparisons of Observed Mean Scores for Each Birth Weight Category with Their Estimated Mean Scores after Adjusting for the Quintiles of the Balancing Score</td>
<td>137</td>
</tr>
</tbody>
</table>
List of Figures

Figure 3.1 Outline of data analysis ................................................................. 42

Figure 4.1 Scatterplot matrix for the three PIAT subtest scores of the
1434 non-black-non-Hispanic and black children in the
NLSY79-C ........................................................................................................ 45

Figure 4.2 Scatterplots of age against each cognitive score with loess,
linear, and quadratic fits superimposed on the data for the
1434 non-black-non-Hispanic and black children in the
NLSY79-C ........................................................................................................ 46

Figure 4.3 Scatterplot matrices for cognitive composite score, gestational
age in weeks, birth weight, and sex for the non-black-non-Hispanic
(n=781) and the black (n=653) children in the NLSY79-C ...................... 49

Figure 4.4 Fitted surface for locally linear bivariate regression (span=0.85)
of cognitive composite score on birth weight and gestational age for
the 781 non-black-non-Hispanic children in the NLSY79-C ................. 51

Figure 4.5 Fitted surfaces for locally linear bivariate regression (span=0.85)
of cognitive composite score on birth weight and gestational age
for the 401 non-black-non-Hispanic males (the two graphics above)
and for the 380 non-black-non-Hispanic females (the two graphics
below) in the NLSY79-C ........................................................................... 53

Figure 4.6 Fitted surface for locally linear bivariate regression (span=0.85) of
cognitive composite score on birth weight and gestational age for
the 653 blacks in the NLSY79-C ................................................................. 54
Figure 4.7  Fitted surfaces for locally linear bivariate regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 347 black males (the two graphics above) and for the 306 black females (the two graphics below) in NLSY79-C…………………………55

Figure 4.8  2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by race for the 1206 non-black-non-Hispanic and black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks…………………………………………………………………..58

Figure 4.9  2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by gender for the 654 non-black-non-Hispanic children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks……60

Figure 4.10  2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by gender for the 552 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks……………………………………………….62

Figure 4.11  Plot to demonstrate division of the birth weight distribution of the 654 non-black-non-Hispanic children and that of the 551 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks………………………………...64

Figure 4.12  Histograms showing overlap between the balancing score distributions for the 654 non-black-non-Hispanic children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks…………………………………………72
Figure 4.13 \textit{cbal.psa} plots checking covariate balance for the 654 non-black-non-Hispanic children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks........74

Figure 4.14 Fitted surface for locally linear bivariate regression (span=0.85) of cognitive composite score on birth weight and balancing score for the 557 non-black-non-Hispanic children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks and whose balancing scores are in the overlapping region from 0.6 to 0.95……………………………………………………………………..75

Figure 4.15 Histograms showing overlap between the balancing score distributions for the 551 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks……81

Figure 4.16 \textit{cbal.psa} plots checking covariate balance for the 551 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks……………………………………82

Figure 4.17 Fitted surface for locally linear bivariate regression (span=0.85) of cognitive composite score on birth weight and balancing score for the 493 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks and whose balancing scores are within the overlapping region from 0.65 to 0.95…………………………………………………………………83

Figure 4.18 Scatterplot matrix for the four WJ-R subtest scores of the 1763 white and black children in the PSID-CDS………………………………87
Figure 4.19  Scatterplots of age against each cognitive score with loess, linear, and quadratic fits superimposed on the data for the 1763 white and black children in the PSID-CDS…………………………………………88

Figure 4.20  Scatterplot matrices for cognitive composite score, gestational age in days, birth weight, and sex for the white (n=931) and the black (n=832) children in the PSID-CDS…………………………………91

Figure 4.21  Fitted surface for locally linear bivariate regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 931 white children in the PSID-CDS…………………………………92

Figure 4.22  Fitted surfaces for locally linear bivariate regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 459 white males (the two graphics above) and for the 472 white females (the two graphics below) in the PSID-CDS……………….93

Figure 4.23  Fitted surface for locally linear bivariate regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 832 blacks in the PSID-CDS…………………………………………94

Figure 4.24  Fitted surfaces for locally linear bivariate regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 443 black males (the two graphics above) and for the 389 black females (the two graphics below) in the PSID-CDS……………….95

Figure 4.25  2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by race for the 1575 white and black children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days…………………………………………………98
Figure 4.26 2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by gender for the 834 white children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days…………………………………………………………100

Figure 4.27 2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by gender for the 741 black children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days…………………………………………………………102

Figure 4.28 Plot to demonstrate division of the birth weight distribution of the 833 white children and that of the 741 black children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days…………………………………………………………104

Figure 4.29 Histograms showing overlap between the balancing score distributions for the 833 white children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days……109

Figure 4.30 *cbal.psa* plots checking covariate balance for the 833 white children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days……………………………………110

Figure 4.31 Fitted surface for local linear bivariate regression (span=0.85) of cognitive composite score on birth weight and balancing score for the 804 white children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days and whose balancing scores are in the overlapping region from 0.55 to 0.9…………………………………………………………111
Figure 4.32  Histograms showing overlap between the balancing score
distributions for the 741 black children in the PSID-CDS whose
gestational ages are in the normal range from 259 days to 294 days……..116

Figure 4.33  cbal.psa plots checking covariate balance for the 741 black
children in the PSID-CDS whose gestational ages are in the
normal range from 259 days to 294 days…………………………………117

Figure 4.34  Fitted surface for locally linear bivariate regression (span =0.85)
of cognitive composite score on birth weight and balancing score
for the 643 black children in the PSID-CDS whose gestational
ages are in the normal range from 259 days to 294 days and whose
balancing scores are in the overlapping region from 0.55 to 0.9…………..118

Figure 5.1  Scatterplot matrices for cognitive composite score, gestational
age, birth weight, and sex for the Non-Black-Non-Hispanics and
for the Blacks in the NLSY79-C and for the Whites and for the
Blacks in the PSID-CDS…………………………………………………………123

Figure 5.2  Comparisons of loess curves relating cognitive composite
score to birth weight when gestational age was restricted to
be normal between the NLSY79-C and the PSID-CDS first
by race then by gender for whites (or non-black-non-Hispanics)
and for blacks……………………………………………………………………..128

Figure 5.3  Plots to demonstrate the creation of three birth weight categories
for each subsample……………………………………………………………..133
Figure 5.4  Comparisons of fitted locally linear regression surfaces relating cognitive composite score to birth weight and balancing score between the NLSY79-C and the PSID-CDS when gestational age was restricted to be normal and balancing score was restricted to the region of overlap………………………………………………….135
Abstract

This study applied nonparametric loess regression to describe the predictive association between birth weight and cognitive ability in childhood and generalized propensity score methods to control the confounding of multiple covariates that summarize prenatal differences.

The data from wave one of the National Longitudinal Study of Youth 1979-Children (NLSY-C) and the data from wave one of the Panel Study of Income Dynamics-Child Development Supplement (PSID-CDS) were analyzed in parallel and the results were compared.

Analysis of the two datasets proceeded in two stages. At the first stage, the association between birth weight and cognitive ability was examined first using loess regression models in which both birth weight and gestational age were the predictors, and then by simple loess regression models in which birth weight was the sole predictor while gestational age was restricted to the normal range, i.e., between 37 complete weeks (259 days) and 42 complete weeks (294 days) to help control its confounding effects. The association was examined separately for subsamples defined by race and sex. At the second stage, the distribution of birth weight was broken into low, medium, and high, and the challenge of controlling multiple covariates was addressed by stratifying the three groups on the balancing score estimated from generalized propensity score models that can accommodate more than two groups.

The shape of the association between birth weight and cognitive ability as depicted by loess regression was not consistent across the two datasets. Among the children in the NLSY-C who were born near term and with normal birth weights, the
association had an upside-down U-shape. Among the children in the PSID-CDS who were born near term and with normal birth weights, the association was roughly a continuously increasing one. The shape of the association persisted, and was slightly strengthened when multiple covariates that account for prenatal differences were controlled by balancing score-based stratification.

Results from the two data sets demonstrated consistent racial and gender differences. The association was stronger for whites than for blacks. White males had slightly higher cognitive scores than white females, while black females had slightly higher scores than black males.
Chapter 1 Introduction

Advances in neonatal care have improved perinatal health to the point where almost all infants born in the west beyond a threshold gestational age survive (Hintz et al., 2005). Concern has shifted from survival toward understanding the relationship between perinatal characteristics, like birth weight, and health and developmental outcomes in childhood and later adulthood (Gorman, 2002b). This study applied modern nonparametric regression methods on the data collected in the United States by the National Longitudinal Study of Youth 1979-Children (NLSY79-C) in 1986 and the data also collected in the United States by the Panel Study of Income Dynamics-Child Development Supplement (PSID-CDS) in 1997 to reveal the relationship between birth weight and cognitive ability in childhood. The central goal of this study is to learn how birth weight relates to cognitive ability after more extensive controls are applied concerning potential confounders. This is done by studying how the relationship between birth weight and cognitive ability compares for the two sexes, and for different racial groups for the two named datasets when using nonparametric regression methods. This is also done by studying how adjusting for the effects of multiple confounders through the use of generalized propensity score methods may provide further information about this relationship.

Statement of the Problem

With regard to the relationship between birth weight and cognitive ability of the general population in their childhood, findings in the literature are surprisingly consistent. Virtually all suggest a significant positive association: the outcomes from cognitive tests improved significantly with increasing birth weight, although the associations are often
complex and the magnitude is fairly small (Astone & Guyer, 2004; Cheadle & Goosby, 2010; Silva et al., 2004; Metha, & O’Callaghan, 2006; Tong, Baghurst, & McMichael, 2006; Yang et al., 2008). When confounders were accounted for in the analyses, most of the studies showed an attenuation of the association (e.g., Jefferis et al., 2002; Yang et al., 2008), while a few others showed an increase (e.g., Matte et al., 2001; Richards et al., 2002).

Research on this topic has usually used either binned birth weight categories (e.g., Sorensen et al., 1997) or parametric regression models (e.g., Jefferis, Power, & Hertzman, 2002) to investigate the relationship. Binned birth weight categories have been criticized for arbitrarily defining the cutting points, while regression models often require unrealistic assumptions about the functional form of the relationship between the outcome and the predictor(s), and inappropriate assumptions can make substantial differences to the results (Winship & Morgan, 1999). Given what is known to be a complex set of relationships between birth weight and cognitive outcomes, simple parametric assumptions are generally unwarranted. Although there are a variety of post-estimation diagnostic methods (residual plots, leverage calculations, etc) to help identify better models, these have rarely been used to improve knowledge about the relationship between birth weight and cognitive ability.

By contrast, nonparametric regression models, such as loess regression (locally weighted polynomial regression), relax stringent assumptions about the nature of the relationship between the outcome and the predictor(s), typically substituting a weaker and more realistic assumption that the average value of the outcome is a smooth, continuous function of the predictor score values. When loess regression is used to generate
predictions, the predictions are usually better than their parametric counterparts. That is, by making weaker assumptions, loess regression yields smaller average squared errors of prediction, when compared to either linear or polynomial regression counterparts. A major premise of the current study is that the use of nonparametric loess regression can yield more information about such relationships than heretofore have been known to exist.

Another persistent challenge faced by the research on this topic is to disentangle at least the primary confounders that tend to mediate the relationships between birth weight and cognitive outcomes. Confounding is a major concern in causal studies whenever relationships between predictors and outcomes tend to differ with respect to one or more confounders. If an association is seen, even a weak one, does it mean that lower (higher) birth weight tends to cause retarded (or enhanced) cognitive development? Problems of interpretation tend to become prominent when substantial differences are found in the predictions of cognitive outcomes across the range of birth weight with respect to multiple dimensions of confounders.

Traditional methods, such as stratification (or matching) or parametric multiple regression, have limited capabilities in coping with the challenge of confounding. In those studies that used parametric multiple regression, confounders were usually put into the regression model together with birth weight as predictors; while in those studies that used the stratification (or matching) method, the sample was divided by one or two confounders each time, and then the means of cognitive outcomes were compared across different cells (or strata). However, stratification (or matching) method becomes impractical and difficult to implement when there are a large number of confounders to
control for (Winship & Morgan, 1999), which is true for research on this topic. The stratification (or matching) method is often used together with binned birth weight categories. If the distribution of birth weight is divided into three categories, i.e., less than 2,500g, more than 4,000g, and those in between, in case of \( n \) dichotomous confounders, there are \( 3 \times 2^n \) possible cells (or strata) over which the means of cognitive outcomes need to be compared. Obviously, the problem gets worse if some of the confounders are continuous. It is also known that one problem with a linear regression approach is that it imposes a linearity constraint on the functional form of the relationship between the outcome and the confounder(s). Nonlinear terms can be added, but it is generally difficult to know how nonlinearity can be approximated (Winship & Morgan, 1999).

Propensity score methods have recognized advantages over traditional methods in reducing confounding. For example, these methods permit many observed confounders to be considered simultaneously. The propensity score replaces multiple confounders such that just one score is applied as a predictor rather than multiple individual confounders (Rubin, 1997), thus greatly simplifying the model. The use of multiple predictors is not only accepted in propensity score models, it is invited. In the context of this study, any variables that might confound interpretations about the effects of birth weight are generally recommended for use in the model; further adjustments can also involve interactions of confounders and their polynomial terms (Rosenbaum & Rubin, 1984). Second, propensity score methods do not require any particular specification of the relationship between confounders and the outcome, which contrasts with uses of parametric regression models (Drake, 1993; Rubin, 1997). The propensity score is estimated without attention to the ultimate outcome.
Previous research has generally been informative showing the confounding effects of sex, race, gestational age, birth order, maternal age, maternal IQ (or its proxy, maternal educational attainment), maternal smoking and drinking habits during pregnancy, maternal physical and mental conditions, socioeconomic status of the family, postnatal family and school environment, and etc..

Among all the potentially important confounders other than sex and race, gestational age, a measure of fetal age at birth, needs particular attention. One common practice has been to use birth weight as a principal predictor, while gestational age has fallen into the shadows as a confounder whose role has not been studied as systematically and as thoroughly as it deserves. It is a simplistic argument to say that gestational age is a “cause” of birth weight, or “prematurity” causes “immaturity”. For instance, parturition could be accelerated or delayed because of large or small fetal size, respectively; or a small and seriously impaired infant might be delivered prematurely. On the other hand, gestational age can not be simply taken as a direct indicator of birth weight, since many infants born preterm are not seriously below the birth weights of those whose gestational age is ‘normal’; also, many full term infants are below normal birth weight standards at birth. It has argued that both birth weight and gestational age are needed to study etiology, to predict outcome, and to inform pre- and postnatal care. In fact, modern analyses of birth outcomes, such as perinatal mortality, include both birth weight and gestational age as predictors (e.g., Tentoni, et al., 2004). It has been found that in large unselected samples birth weight accounts for more than 90 percent of the variance in perinatal mortality, while gestational age accounts for 4 to 6 percent of the variance (Susser, et al., 1972). It is reasonable to believe that gestational age will also account for
additional variance in cognitive outcomes besides what has been accounted for by birth weight.

**Major objectives and research questions**

The major objective of the proposed study was to use modern nonparametric regression methods to improve general understanding of the predictive association between birth weight and cognitive ability in childhood. That is, the goal was to sharpen or clarify the relationship between these two variables, when compared to what other researchers have done. This objective was further refined to examine, and thereby reduce the confounding effects of sex, race, and gestational age, related to prediction of cognitive ability from birth weight. A secondary objective was to use generalized propensity score methods that consider birth weight as “multiple doses of a treatment” in order to accommodate the confounding of multiple variables that summarize prenatal differences other than sex, race, and gestational age.

Because the observed association between birth weight and cognitive ability in childhood could be spurious for one particular sample, data from two large prospective studies in the United States that track development from childhood into adulthood were examined. These were the National Longitudinal Study of Youth 1979-Children (NLSY79-C) and the Panel Study of Income Dynamics-Child Development Supplement (PSID-CDS). In addition to learning about the cross-cohort generalizability of results, the use of data from two cohorts was expected to have notable value in reducing the uncertainty of findings.

Data analysis proceeded by two stages with each stage aiming at one of the two objectives mentioned above. At the first stage, the relationship between birth weight and
cognitive ability in childhood was examined primarily with loess regression. At first, both birth weight and gestational age were used as predictors of cognitive ability. The objective was to study whether both birth weight and gestational age contribute to variance in cognitive ability. Then gestational age was restricted to the normal range, i.e. between 37 complete weeks (259 days) and 42 complete weeks (294 days) to help control its confounding effects; the relationship between birth weight and cognitive ability became the primary concern with birth weight being the sole predictor for cognitive ability. Because it was expected that the relationship would be found to vary by race and sex, the relationship was examined for males and for females separately within each racial group to the extent that subsample sizes allowed and data on race were available. At the second stage, confounding from multiple variables other than gestational age that summarize fundamental prenatal differences was adjusted for using propensity score stratification.

The specific research questions addressed were the following:

(A) What is the nature of the predictive association of birth weight and gestational age for cognitive ability in childhood?

(a.1) What is the nature of the predictive association of birth weight and gestational age for the cognitive ability of white males?

(a.2) What is the nature of the predictive association of birth weight and gestational age for the cognitive ability of white females?

(a.3) What is the nature of the predictive association of birth weight and gestational age for the cognitive ability of black males?
(a.4) What is the nature of the predictive association of birth weight and gestational age for the cognitive ability of black females?

(B) What is the nature of the predictive association of birth weight for cognitive ability in childhood when gestational age is restricted within the normal range, i.e., between 37 weeks (259 days) and 42 weeks (294 days)?

(b.1) What is the nature of the predictive association of birth weight for the cognitive ability of the white males whose gestational age is within the normal range?

(b.2) What is the nature of the predictive association of birth weight for the cognitive ability of the white females whose gestational age is within the normal range?

(b.3) What is the nature of the predictive association of birth weight for the cognitive ability of the black males whose gestational age is within the normal range?

(b.4) What is the nature of the predictive association of birth weight for the cognitive ability of the black females whose gestational age is within the normal range?

(C) Do observed associations increase or decrease after adjusting for numerous confounders that concern prenatal differences?

(D) Do the observed associations differ across the two national data sets? If so, how, and to what degree?
Chapter 2 Literature Review

*Birth Weight and Cognitive Ability in Childhood*

Birth weight, as a crude summary of genetic and prenatal environmental influences, has been linked to illnesses in adult life, such as coronary heart disease, hypertension, and type 2 diabetes mellitus. Barker (1998) argues that this is because biological and environmental factors “program” the structure and physiology of children in utero and during infancy. Any insults experienced during these critical periods of development will have direct, long-term consequences for health and disease susceptibility in later life. While the “programming” hypothesis is most often applied to physical health outcomes, this perspective has also been utilized to explain the observed association between birth weight and developmental delay in cognition. Disadvantaged prenatal conditions have been related to abnormalities in brain structure and thus are thought to affect cognitive development. Experimental evidence from animal models (Bhutta & Anand, 2001) suggests that disadvantaged prenatal conditions, such as malnutrition and poor maternal health and physical states, can promote neuronal cell death in the immature brain, increased rates of neuronal cell death could lead to volumetric losses in specific brain regions and may at least partially explain the cognitive abnormalities noted later in life. Several investigators (e.g. Nosarti et al., 2006) have observed atypical activations in certain brain regions when survivors of very low birth weight (<1500g) were challenged with cognitive tasks. Isaacs et al. (2000, 2001) found that a sample of adolescents born at very low birth weight had a selective impairment in mathematics computation that was accompanied by reduced bilateral hippocampal
volumes. Peterson et al. (2006) observed that smaller head circumference at age 8 was associated with lower scores in cognitive tests among children born preterm.

Babies born weighing less than 5 pounds, 8 ounces (2500 grams) are considered being born with low birth weights (LBW), and less than 1500 grams are considered very low birth weights (VLBW). Gender and racial differences have been noted. Research shows that the entire birth weight distribution of females is shifted to the left of that of males and thus females have a greater risk of being born at low birth weights than males (Kramer, 1987), but the differences are decreasing over time (Van Vliet et al., 2009). And blacks are about twice as likely as whites to be born at low birth weights, even after controlling for socioeconomic status (Collins, 1997; Cramer, 1995; Frisbie, Biegler, and de Turk, 1997; Hummer, 1993; Wilcox and Russell, 1990).

Low Birth Weight and Cognitive Ability in Childhood

Improvements in neonatal care have enabled an increasing percentage of children who were born preterm (<37 weeks) or with LBW to survive. Multiple studies of children who were born preterm or with LBW have followed up cohorts from birth to school age (>=5 years). A large number of these children have adverse outcomes such as cerebral palsy, hydrocephalus, blindness, deafness, or seizures (Hack et al., 1994; Roussounis, Hubley, & Dear, 1993; Wood et al., 2000). Even in children without obvious disabilities, selective deficits were observed in the domains of mental and motor skills, visual recognition memory, language, executive, attentional, and other neuropsychological skills (Espy et al., 2002; Goyen, Lui, & Woods, 1998; Landry et al., 1997; Sullivan & McGrath, 2003) that might directly contribute to cognitive weaknesses. Compared to their normal counterparts, children born with LBW tend to have lower cognitive and
educational test scores (see Taylor, Espy, & Anderson, 2009 for a review), more school problems (Klebanov et al., 1994a, 1994b; McCormick et al., 1990), grade failure and placement in special classes (Klebanov et al., 1994a, 1994b), and decreased likelihood of timely high school completion (Conley et al., 2003). These disadvantages are evident by the time of school entry, and persist throughout the school-age years. However, there are also studies that found no differences between preterm-born cases and term-born controls (e.g., Drillien, Thomson, & Burgoyne, 1980).

**Normal Birth weight and Cognitive Ability in Childhood**

Most studies of children born with LBW or VLBW employed hospital-based samples in their analyses (Gorman, 2002b). A natural question to ask at this point is whether the association between birth weight and cognitive abilities observed for LBW or VLBW children—low birth weight children having relatively low cognitive test scores, with scores improving as birth weight increases—would also exist in the general population, particularly among those who were born with normal birth weights (>2500 g) or on term (37-42 weeks). Confirmation of the association in the general population is important since this would imply that risk factors are similar and interventions targeted on birth weight would have a beneficial impact on the population level.

Shenkin, Starr, & Deary (2004) did a systematic search for studies on the relationship between normal birth weights and cognitive ability in childhood, including published journal articles and unpublished reports, papers, dissertations, theses, and conference proceedings. This search resulted in six studies that met their inclusion criterion of using data from general population surveys. Although the data from these six studies were collected from different cohorts who were born over more than half century
(from 1921 to 1987) on two different continents (North America and Europe), there were some surprising consistencies in their results. They all suggest a statistically significant positive association. The outcomes from cognitive tests improved significantly with increasing birth weight, but the magnitude was fairly small. Some studies (e.g. Jefferis, Power, & Hertzman, 2002; Richards et al., 2002) also observed a slight decline in test scores among children with high birth weight (>= 4000g), but Richards et al. (2002) found that the decline no longer existed after correcting for birth order. Binned birth weight categories and parametric multiple regression were the popular methods in these earlier studies,

This significant positive association was further confirmed by studies conducted later with data from cohorts born from 1970’s to present (e.g. Astone & Guyer, 2004; Cheadle & Goosby, 2010; Silva, Metha, & O’Callaghan, 2006; Tong, Baghurst, & McMichael, 2006; Yang et al., 2008). For example, Astone & Guyer (2004) explored the relationships between birth weight and a number of cognitive outcomes using data from the Panel Study of Income Dynamics Child Development Supplement (PSID-CDS). Before any potential confounders were controlled, there were significantly positive associations between birth weight and most of the cognitive outcomes for both boys and girls, but the effects of birth weight again were very small. Yang et al. (2008) used data from the National Longitudinal Study of Youth 1979-Children (NLSY1979-C). It was found that, before any potential confounders were taken into account, there were positive associations between birth weight and cognitive ability at ages 5 to 6, 7 to 9, and 11 to 12. In these later studies, new analytical techniques were developed or applied, which included fixed-effects models (Astone & Guyer, 2004), structural equation modeling
(Silva, Metha, & O’Callaghan, 2006), and latent class growth modeling (Cheadle & Goosby, 2010).

However, Shenkin, Starr, & Deary (2004) argue that the small, statistically significant association between birth weight and cognitive ability suggested by the literature may have been due to 1. publication bias, where studies that produced non-significant results may not have been published; 2. selection bias, where subjects included in these studies may not be representative of the general population; 3. inclusion of births in the low birth weight range, since almost all published studies included children born with LBW and there are suggestions that the association is mainly driven by those with the smallest birth weights.

Furthermore, Shenkin, Starr, & Deary (2004) point out residual confounding is another major threat to the validation of this small but statistically significant association, that is, the association may have been due to other variables not accounted for in the analyses.

Confounding and Confounders

If an association was observed between birth weight and cognitive ability, even a weak one, does it mean that lower birth weight has a causal effect on cognitive development? The causal connection between birth weight and cognitive ability is what any studies on this topic have endeavored to establish.

Randomized experiments are considered the gold standard for establishing causal effects. In randomized experiments, causal effects are established by comparing outcomes for subjects who were randomly assigned to treatment and control groups over a certain period of time. Naturally, however, with a ‘treatment’ such as birth weight
randomization is not possible. So to study the association between birth weight and cognitive ability requires use observational rather than experimental data. If there are systematic differences between two or more birth weight groups with respect to relevant covariates, then the problem of using observational data to make causal inferences is substantial (Winship & Morgan, 1999). In general, observational studies can lead to comparisons that entail “confounding,” which means that results can lead to a false picture of causal effects. In general, when treated and control subjects differ in their outcomes for reasons other than effects of ‘treatments’ confounding can be said to be present (Greenland, Pearl, & Robins, 1999; Robins & Morgenstern, 1987; Weinberg, 1993).

A confounder must meet three criteria (Greenland, Pearl, & Robins, 1999; Robins & Morgenstern, 1987; Weinberg, 1993): 1. a confounder must be an ancestor (cause) of the outcome; 2. it must be associated with the treatment; 3. and it cannot be affected by the treatment. Psychologists tend to use the term “covariate” when a confounder is implicated (Shenkin, Star, & Deary, 2004).

To date, studies on the relationship between birth weight and cognitive ability of the general population have used either stratification (e.g., Sorensen et al., 1997) or parametric regression models (e.g., Jefferis, Power, & Hertzman, 2002) to cope with the problem of confounding. Stratification is often used together with binned birth weight categories. For example, if we break the distribution of birth weight into categories by every 500 grams, to adjust for gender difference, the subsample falling into each birth weight category can be further divided into males and females, and then comparisons could be carried out across strata. In theory, stratification can be extended to multiple
confounders, but as the number of strata increases, the size of each decreases and the cross-strata comparisons become unstable, which can lead to biased results (Winship & Morgan, 1999). Furthermore, there is a question of where to place the boundaries or cuts when forming bins.

Parametric regression, particularly its extended model forms, is more flexible than stratification in dealing with multiple confounders. But one weakness of parametric regression is that it requires strong assumptions to be made about the form of the relationship between the outcome and each confounder in the model. The assumed form of the relationship between cognitive ability and any of the confounders, usually linear, has never been verified with real data.

The confounders that have in previous research been identified or assumed include sex, race, gestational age, birth order, maternal age (Shenkin, Star, & Deary, 2004), as well as breast feeding (Jefferis et al., 2002), maternal intelligence (Deary, Der, & Shenkin, 2005; Yang et al., 2008), parental education (Jefferis et al., 2002; Matte et al., 2001; Richards et al., 2002; Silva, Metha, & O’Callaghan, 2006), maternal smoking and drinking habits during pregnancy (Yang et al., 2008), social class or socioeconomic status of the family (Jefferis, Power, & Hertzman, 2002; Richards et al., 2002; Shenkin et al., 2001), and between-family and within-family effects (Astone & Guyer, 2004; Cheadle & Goosby, 2010; Yang et al., 2008). When confounders were accounted for in the analyses, most of the studies showed an attenuation of the association (e.g., Jefferis et al., 2002; Yang et al., 2008), but there have also been some studies that showed an increase (e.g., Matte et al., 2001; Richards et al., 2002). A few important “confounders” are discussed below.
Sex and Race. Almost all the studies have accounted for infant sex and most studies from the United States have also accounted for race in the context of using stratification or parametric regression. But accounting for sex and race in these ways can disguise important gender differences and racial differences.

Studies of the relationship between birth weight and infant mortality have shown gender differences and racial differences. For example, Susser et al. (1972) found that among the births in New York City in the period of 1958-1961, perinatal mortality was highest at the lowest birth weights and decreased gradually until an optimum weight around 3500 grams, after that the rates again began to rise. This pattern held within an array of sex and racial groups. In comparisons between groups defined by sex and race, however, perinatal mortality rates did not necessarily reach the same level given a birth weight at a particular gestation interval. In other words, sex and race were found to be indicators of effects on mortality over and above those of birth weight and gestational age.

In view of the possible differences between cognitive outcomes among groups defined by sex and race, it is advisable that studies analyze sexes and races separately. Literature search has revealed a number of studies that provided separate results for boys and for girls, but no consistent conclusions were reached regarding gender difference. Shenkin et al. (2001) and Jefferis et al. (2002) found that birth weights were significantly associated with cognitive test scores among both boys and girls and this association was a little stronger among girls than among boys. In contrast, Matte et al. (2001) found that the association between birth weight and IQ at age seven was stronger among boys than girls. Only a few studies (e.g., Rowe, 2002; Kiweon, 1992; Hardy & Mellits, 1977) that
analyzed data from the United States provided separate results for whites and for blacks and the results have been consistent. Correlations between birth weight and IQ or cognitive test scores were reported to be consistently higher for whites than for blacks.

*Gestational Age.* Gestational age is a measure of fetal age at birth. A common method of calculating gestational age starts counting from the first day of the woman's last menstrual period (LMP). Using the LMP method, gestational ages of 40 complete weeks (280 days) are considered to be a full-term delivery, and those between 37 (or 38) and 42 complete weeks are considered normal (or term).

Past studies, however, have often failed to take gestational age into account, when studying effects of birth weight. To serve most epidemiological and clinical needs, birth weight has been typically been treated as a primary indicator while gestational age has fallen into the shadows as a confounder. In fact, there is evidence to support assessment of gestational age as an important subject of study in its own right. First, low and normal birth weights do not exactly coincide with any dichotomy that separates preterm and term delivery (McKeown & Gibson, 1951). Analysis of the distribution and divergence of perinatal mortality in relation to the two factors also found birth weight, regardless of length of gestation, accounted for somewhat more than 90 percent of the variance in perinatal mortality. Gestational age, with birth weight controlled, accounted for 4 to 6 percent of the variance (Susser et al., 1972). Models have been developed using both birth weight and gestational age to study infant mortality (e.g., Tentoni et al., 2004). No efforts have seem to have been focused on study of developmental outcomes such as cognitive ability in childhood.
Within- and Between-Family Effects. Lower birth weight children are selective on many social and economic dimensions, and researchers (e.g., Gorman, 2002a; Gorman, 2002b) have even argued that social and economic factors which manipulate their post-natal grown-up environment have far more important effects on developmental outcomes than birth weight.

These social and economic factors can be categorized into family effects and outside-family school and neighborhood effects. Sufficient control for these effects has been a big challenge because some effects, such as intellectual climate at home, intimacy degree with caregiver, and maternal social support, are hard to observe and therefore hard to control.

Analyses using sibling data solved this problem to some degree. It allows separation of within-family effects from between-family effects and simultaneous estimation of both. Various models have been developed. For example, Gorman (2002b) used a model similar to what is known as fixed-effects models in econometrics. Instead of using the birth weight of individuals as the predictor and cognitive score as the outcome, the difference in birth weights between a sibling pair was used as the predictor and difference in their cognitive scores was used as the outcome. Regressing the difference in cognitive scores on the difference in birth weights was supposed to produce unbiased estimates of the within-family effects shared by the sibling pair. Yang et al. (2008) used a regression model that was originally developed to analyze twin data (Carlin et al., 2005). To simultaneously estimate the within- and between-family effects, a mean birth weight was calculated for each family as an indicator of between-family effects and
a deviation from the family mean was calculated for each sibling as an indicator of within-family effects.

Potential Avenues for Further Study Suggested by the Literature

Shenkin, Star, & Deary (2004) recommended a few potential avenues that future studies can follow. The studies reviewed in Shenkin, Star, & Deary (2004) analyzed data from cohorts who were born from 1921 to 1987, so the associations discovered for these old cohorts may not hold in today’s population. Therefore, they suggest it is important that data from children born later than 1970 reach the literature, and the data should be analyzed in ways that allow comparisons across studies and across cohorts. Now we have studies that analyzed data from children born later, such as Astone & Guyer (2004) and Yang et al. (2008), but differences in their research design and methodology prevent meaningful comparisons. For example, Astone & Guyer (2004) only included the children who were interviewed during PSID-CDS data collection wave one in 1997, regardless their age when they were interviewed. Yang et al. (2008), in contrast, included all the children who were interviewed across the entire NLSY79-C study period from 1986 to present, but the age of these children when they were interviewed must be the same.

They also suggest that studies analyze sexes separately. The majority of these studies they reviewed analyzed data from Europe, so what was considered to be important may have been sex difference only, but racial difference could be also very important to a country like the United States whose population are from a variety of racial backgrounds.
Furthermore, there are suggestions that the relationship is not linear, at least the slight decline in the upper end of the birth weight distribution was observed. So methods that don't require stringent assumptions regarding the functional form of the relationship between birth weight and cognitive outcomes are highly recommended.

Finally, Shenkin, Star, & Deary (2004) also advocate use of both birth weight and gestational age.

**Loess Regression**

Loess regression, locally weighted polynomial regression, originally developed by Cleveland (Cleveland, 1979; Cleveland & Devlin, 1988), is a nonparametric modeling technique that preserves the simplicity of parametric modeling but enjoys far more flexibility. Unlike its parametric counterpart that fits one global function to the overall data, loess regression fits weighted low-degree polynomial regression models to localized subsets of the data to build up a function that accounts for variation in the data to the full extent. Loess regression is based on the idea that any function, no matter how complex it is, can be well approximated in a small neighborhood by an easy-to-fit weighted regression model.

Parametric regression specifies both the functional form for the relationship between \(x\) and \(y\) as well as the error distribution, leaving only parameters to be estimated. In loess regression, however, the functional form for the relationship between \(x\) and \(y\) is left to be estimated. The error distribution is also left unspecified or is assumed to have very general properties.
Mathematical Rationale

Fox (2002) summarized the rationale that stands behind loess regression. In case of a simple model with only one predictor, \( y_i = f(x_i) + \varepsilon_i \), the process of fitting loess regression usually involves four steps.

1. Specify the “span”, \( s \), that determines how much of the data is used to fit each weighted polynomial regression. The span, \( s \), is generally chosen to make what is judged to be a good tradeoff between a good fit and an appropriately smoothed fit. The subset of data used in each weighted fit is comprised of the \( n*s \) (rounded to the next integer) points whose \( x \) values are closest to a particular \( x \)-value, \( x_0 \), at which \( y \) is being estimated. The smaller the span is, the closer the overall estimated function will conform to the data. In most applications, however, intermediate values of this ‘span’ argument are ordinarily used, as the loess approach entails a trade-off between getting results that change substantially across the \( x \) range, and becoming uninterpretable, and getting results that tend to be wholly insensitive to \( x \) vs. \( y \) relationship across the range of \( x \), which is what parametric regression usually does.

2. Use a kernel function to generate a weight for each observation within a subset of data. The use of weights is based on the idea that points closer to \( x_0 \) are more likely to influence estimation of \( y \) from \( x_0 \) than points that are further apart. Thus, the weights for observations that are close to \( x_0 \) should be the greatest and then falls symmetrically and smoothly as \( |x_i - x_0| \) grows. The
traditional weight function used for loess regression is the tri-cube weight function.

\[ W(z) = \begin{cases} 
(1 - |z|^3)^3 & \text{for } |z| < 1 \\
0 & \text{for } |z| \geq 1 
\end{cases} \]

where \( z_i = (x_i - x_0) \) divided by \( \max_{x_i \in N(x_0)} |x_i - x_0| \). The weight for a specific point \( x_i \) in any localized subset of data is obtained by evaluating the weight function at the distance between that point \( x_i \) and \( x_0 \), after scaling the distance so that the maximum absolute distance over all of the points in the subset of data is exactly one.

3. Find \( \hat{y}(x_0) \) at \( x_0 \) using a \( p \)-th-order weighted polynomial regression,

\[ \hat{y}(x_0) = \hat{a} + \hat{b}_1 x_0 + \hat{b}_2 x_0^2 + \ldots + \hat{b}_p x_0^p \]

and find \( \hat{a} \), \( \hat{b}_1 \), \( \hat{b}_2 \), \ldots \( \hat{b}_p \) that minimize the weighted residual sum of squares,

\[ \sum_{i=1}^{n} w_i (y_i - \hat{a} - \hat{b}_1 x_i - \hat{b}_2 x_i^2 - \ldots - \hat{b}_p x_i^p)^2 \]

The local polynomial fit to each subset of the data are almost always either locally linear or locally quadratic. Higher-degree polynomials would tend to over-fit the data in each subset and are numerically unstable.

4. This procedure is repeated for each \( n \) \( x \)-values, and the fitted values are connected, i.e., smoothed, to give an overall curve showing the predictions of \( y \) from \( x \).

The loess approach has been extended (in R) for up to four predictors, and the process is conceptually straightforward.
1. Specify a multivariate subset of data close to \( x_0 = (x_{01}, x_{02}, \ldots, x_{0k}) \) to be used to fit a weighted polynomial multiple regression.

2. Weights are defined by the scaled distances from \( x_i = (x_{i1}, x_{i2}, \ldots, x_{ik}) \) to
\[ x_0 = (x_{01}, x_{02}, \ldots, x_{0k}). \]

3. Find \( \hat{y}(x_0) \) at \( x_0 \) using, for example, a locally weighted linear regression:
\[ \hat{y}(x_0) = \hat{\alpha} + \hat{b}_1 x_{01} + \hat{b}_2 x_{02} + \ldots + \hat{b}_k x_{0k} \]
and find \( \hat{\alpha}, \hat{b}_1, \hat{b}_2, \ldots, \hat{b}_k \) that minimize the weighted residual sum of squares. Equations are readily generalized to second degree forms.

4. The procedure is repeated for each \( n \) \( x_0 \).

5. Results for successive values of the predictors are then smoothed by connecting the conditional means across the ranges of the predictors.

Similar to parametric multiple regression, the statistical significance of each predictor in loess regression can be assessed by dropping it from the model and performing an approximate incremental \( F \)-test for the change in the residual sum of squares (Fox, 2002).

**Graphic Visualization of Results**

As discussed above, in loess regression there are no parameter estimates. To see results, the fitted loess curve (in the simple regression case) or surface (in the case where there are two predictors) is examined graphically. In the simple regression case, a two-dimensional scatterplot that shows all the data with loess curves superimposed on the data allows researchers to discern patterns or features in the predictive function that tend to be masked or hidden when parametric regression methods and statistics are used. Put
in another way, the general procedure of fitting a loess curve supplies a means for “letting the data speak” in a way that effectively characterizes virtually any extant relationship that may exist between the response and the predictor without specifying strong prior assumptions other than that the relationship be smooth. The degree of smoothing can be controlled by the user.

Loess regression tends to reduce, often substantially, the danger of misrepresentation of the ‘true’ predictive association of the predictor to the response. Graphical tools have been developed to display the results for the simple regression case, such as, the R function `loess` in the `stats` library and the R function `loess.psa` in the `PSAgraphics` library (Helmreich & Pruzek, 2009). When two predictors are employed then loess regression yields a (smoothed response) surface; graphical methods, such as the R function `scatter3d` in the `car` library (Fox & Weisberg, 2011), can be used in such cases to represent projections onto two dimensional surface—from the three-dimensional space that might in principle be examined from many angles or perspectives. Many applications of these graphical tools have visualization of the target association as their goal. In these instances, statistical inference is of secondary interest. However, modern graphic techniques, such as the R function `xyplot` in the `lattice` library (Sarkar, 2008), have made it possible to construct a confidence band around the fitted loess curve to serve inference purposes. Theoretically, an upper surface and a lower surface can be also constructed around the fitted surface when there are two predictors.

**Advantages and Disadvantages**

The biggest advantage of loess regression, as discussed above, is that it does not require the specification of the functional form of the relationship between the response
and the predictor(s) to fit a model to all the data. This advantage makes loess regression ideal for modeling complex data for which no theoretical models exist.

Because of this special feature, however, the results of loess regression can not be easily represented in an easy-to-understand mathematical formula with relevant statistics reported, such as regression parameter estimates and corresponding effect sizes. Depending on the research question, this could be either a major or a minor drawback in using loess regression.

Another disadvantage is that loess regression requires a fairly large sample in order to produce good and reliable fits, and it can be sensitive to the effects of outliers, like parametric methods (however, various features of the method can help ameliorate this problem).

*Propensity Score Methods to Control for Confounding*

*General Introduction*

In observational studies, subjects are not randomly assigned to treatments. As a result, groups that received different treatments may be systematically different in terms of various characteristics. These differences can confounded with the treatment they receive and therefore lead to biased estimates of treatment effects. Propensity score methods, originated from Rosenbaum and Rubin (1983), have been widely used to control for differences between the treatment and control groups in observational studies and to provide support for causal inferences that can sometimes approach those that might have derived from a corresponding randomized, controlled experiment, if the latter are was feasible.
The basic idea of propensity score methods is to derive a scalar variable based on covariates that account for any observed group differences prior to the treatment. The scalar variable is called the propensity score and it is defined as the conditional probability of being assigned to the treatment group given pretreatment observed covariates; it is also a “balancing score” such that the conditional distribution of the pretreatment covariates, given the propensity score, is the same between the treatment and control groups (Rosenbaum & Rubin, 1983). Suppose each subject has observed covariates, X, and an indicator of treatment, Z=1 if treated and Z=0 if control. The propensity score, \( e(X) \), is the probability that a person will be in the treatment group, conditional on his or her observed covariates, X. That is, \( e(X) = \text{prob}(Z=1|X) \). If the investigator pairs subjects with the same propensity score \( e(X) \) or sorts them into strata according to their propensity scores, then treated and control subjects within these pairs or strata will tend to have similar distributions of X since \( e(X) \) is a balancing score. Comparisons of treatment and control groups within matched pairs or strata tend to be unconfounded if covariate selection has been effective (Rubin, 1997). To the extent that this is true, this means that causal inferences may be better supported.

There are two phases of an application of propensity score methods. In phase I, covariates are selected and used to estimate the propensity score. By definition, a covariate is identical to a confounder, although people who use propensity score methods usually prefer the term “covariate” to “confounder”. Thus, selected covariates should be associated with the outcome and they must be measured prior to the start of treatment so that they are not affected by the treatment (Joffe & Rosenbaum, 1999). In practice, however, adjustments for confounding are not always confined to pretreatment variables.
(Rosenbaum, 1984). They could be the variables whose quantity is a constant such as sex; the variables that change with time but the change is not affected by the treatment; the variables whose quantity may be affected by the treatment, but the effects are minor compared to the effects of the treatment on the outcome; or the variables that are affected by the treatment but are plausible surrogates for a clearly relevant unobserved pretreatment variables. Although logistic regression appears to be the most commonly used method for estimating the propensity score, the uses of bagging or boosting, recursive partitioning or tree-based methods, random forests, and neural networks for estimating the propensity score have also been examined (Austin, 2011). At the end of phase I it is important to assess the overlap of the propensity score distributions in the respective treatment groups. One of the key features of propensity score methods is that they help show when data are such as to permit comparisons of units or persons within narrow bands of propensity scores; when overlap is limited or non-existent there may be little basis for comparisons of groups using outcome variable scores.

In phase II, treated and control subjects with identical or very similar propensity score values are usually matched (Rosenbaum, 2002; Stuart, 2010), or the sample is classified into strata according to quantiles of the propensity score distribution (Rosenbaum & Rubin, 1984). Treatment effects are estimated by comparing the treated and control subjects who were matched into a pair or who were classified into the same stratum. Matching or stratifying on the propensity score makes it possible to control for confounding from any observed covariates. In an ideal situation, the distributions of the observed covariates are “balanced” between the treated and control subjects within a matched pair or within a stratum that is homogeneous in the propensity score, and
therefore, if any differences are observed between the treated and control subjects, the analyst can be more confident that the differences are the result of the treatment they received rather than pretreatment covariate differences.

Assessing balance after Phase I has been completed is therefore an important additional step. The degree of covariate balance achieved by propensity score stratification was initially examined by two-way analysis of variance as recommended by Rosenbaum & Rubin (1984) in which each covariate is used as the dependent variable, and a treatment indicator and a propensity score stratum index are the two factors. A non-significant main effect of treatment and a non-significant interaction of treatment and propensity score stratum index are often used to confirm that the covariate is balanced across treated and control groups within strata.

However, more and more researchers have argued against the idea of using inferential tests to assess covariate balance. Imai, King, & Stuart (2007) argue that any statistic that is used to evaluate balance should be “a characteristic of the sample and not of some hypothetic population and the sample size should not affect the value of the statistic (p. 498).” An example that meets this standard can be a standardized difference in means or a SMD. The function cbal.psa in the PSAgraphics library (Helmreich & Pruzek, 2009) produces a plot that checks covariate balance based on the comparisons of SMDs, defined as covariate differences in means divided by pooled standard deviation. Such comparisons are not based on inferential tests. In the cbal.psa plot, balance for each covariate is indicated if the absolute value of its SMD, after adjusting for the covariate differences using the propensity score, is “relatively close to 0.0.” It is unrealistic, in most situations, to expect the difference between the groups to be exactly 0.0. Cochran (1968)
suggested a rule of thumb that a mean difference should not differ by more than a quarter of a standard deviation. Other alternatives for assessing balance include use of quantile-quantile plots (Imai, King, & Stuart, 2007), higher order moments, non-parametric density plots and propensity score summary statistics (e.g. Austin and Mamdani (2006) and Rubin (2001)).

If the SMD between the two groups, for any covariate, remains ‘substantial’ after matching or stratification, the propensity score can be re-estimated by trying interactions or polynomial terms of the covariates that remain unbalanced. Also, other methods than logistic regression can be used to construct the propensity score. If such differences remain after repeated trials, unbalanced covariates can be further adjusted for by ordinary regression at the final stage (Zanutto, 2006).

Besides matching and stratification, other propensity score based adjustments have been made available, such as weighting by the inverse probability of treatment (Rosenbaum, 1987) or regressing the outcome variable on the propensity score and on an indicator variable denoting treatment status (Austin, 2011).

The situation is more complex when there are multiple treatment “doses” (e.g., low, medium, and high), particularly when an underlying continuum can be used to order several treatments. Lu et al. (2001) and Zanutto, Lu, & Hornik (2005) extended propensity score methods to accommodate multiple ordered treatments by matching or stratifying on a balancing score, say \( b(X) \). Such a balancing score can be estimated using McCullagh’s (1980) ordinal logit model. In McCullagh’s model, the distribution of treatment doses for person \( i \), \( Z_i \), given observed covariates, \( X_i \), is modeled as:

\[
\log\left( \frac{P(Z_i \geq d)}{P(Z_i < d)} \right) = \alpha_d + \beta^T X_i \quad \text{for } d = 2, 3, 4, 5, \ldots
\]
The key feature of this model is that the distribution of doses given covariates depends on the observed covariates only through $\beta^T X_i$; note that this term is the same for all $d$’s, but the intercept alpha ($\alpha_d$) generally varies across $d$’s. If $\beta^T X_i$, or the estimated logit is used as the balancing score $b(X)$, then persons with the same balancing score in different dose groups are expected to have the same distribution of the covariates $X$, that is, $\text{prob}(X|b(X), Z=z) = \text{prob}(X|b(X), Z=z')$ for each $z, z'$. The balancing score estimated from McCullagh’s (1980) ordinal logit model maintains the main advantage of the propensity score for binary treatment cases, which is that matching or stratifying on this score can balance the distribution of all the observed covariates simultaneously (Zanutto et al., 2005).

Strengths and Limitations

Zanutto and her colleagues (Zanutto, Lu, & Hornik, 2005; Zanutto, 2006;) compared propensity score methods with traditional methods such as parametric regression, and they pointed out some advantages propensity score methods enjoy in dealing with the challenge of confounding. First, propensity score methods do not require specification of the functional form of the relationship between covariates and the outcome because the two-phase design designates that the propensity score be estimated from confounding covariates without access to the outcome. In contrast, regression models usually rely on the assumption of linearity or log linearity. If the distributions are nonlinear, the results from regression models may be misleading. Second, propensity score methods allow many observed confounders to be accounted for simultaneously. Unlike parametric regression, which has to include both treatment and confounding covariates in the model to predict the outcome, propensity score methods replace multiple
covariates with just one score, and the score is applied together with treatment as predictors thus greatly simplifying the model. Any variables that may be confounded with treatment effects should be included in the estimation of the propensity score, and when necessary, even the interactions and polynomial terms that involve these variables should be considered. Furthermore, simple diagnostics are available in propensity score methods to inspect the degree the distributions of confounding covariates in the treatment and control groups overlap and therefore the range over which the data will support causal effects. This is not easy to be done with conventional regression methods. When the covariate distributions are very different, regression models depend on the specific form of the model to “extrapolate estimates of treatment effects” (Zanutto, Lu, & Hornik, 2005, p. 62).

On the other hand, propensity score methods are not perfect, they are subject to a number of weaknesses. In randomized experiments, any differences between groups, both observed and unobserved, are balanced through randomization. But propensity score adjustment can at best ensure no systematic differences in observed covariates only, because the propensity score can be estimated only from observed covariates. Consequently, observational studies, even if analyzed carefully with propensity score methods, generally provide somewhat weaker evidence than would a corresponding experimental study (Pruzek & Helmreich, 2004).
Chapter 3 Methods

*Samples*

In consideration that the observed associations could be the spurious results of one particular sample, data from two latest national longitudinal studies conducted in the United States were examined. The two studies were the National Longitudinal Study of Youth 1979-Children (NLSY79-C) and the Panel Study of Income Dynamics-Child Development Supplement (PSID-CDS). Using essentially the same methods to analyze the data from two nationally representative samples substantially reduced uncertainty in the results and therefore improved generalizability.

*National Longitudinal Survey of Youth 1979-Child (NLSY79-C)*

The National Longitudinal Survey of Youth 1979 is a longitudinal study of a representative sample of 12686 U.S. male and female youths who were 14 to 21 years of age by December 31, 1978. The sample includes substantial oversamples of African-Americans and Hispanics, but it does not include individuals who were not living in the US in 1979, but subsequently immigrated into the US. The NLSY79 contains extensive information about the employment, education, training, and family experiences of the respondents.

Starting in 1986, the NLSY79 included the children of NLSY79 female respondents as an independent sample. The objective was to examine the linkages between maternal-family behaviors and attitudes and subsequent child development and transition to adulthood. In 1986, a total of 4971 children under age 15 were interviewed. Since then, the children and their mothers have been interviewed and assessed every two years, and new children were added at subsequent waves of data collection. Data have
been collected on the children’s cognitive ability, temperament, motor and social development, behavior problems, and self-competence as well as the quality of their home environment. Since 1988, data have also been collected from the children age 10 and over on their schooling, family, peer-related and other attitudes and behaviors. A detailed description of the NLSY79-C can be found in the User’s Guide (Center for Human Resource Research, the Ohio State University, 2009).

**Panel Study of Income Dynamics Child Development Supplement (PSID-CDS)**

The Panel Study of Income Dynamics (PSID) is a longitudinal study of a representative sample of U.S. individuals and the families in which they reside since 1968. Besides the initial 5000 families, the sample also includes new families formed by the children and families immigrated to the U.S later. Information collected includes employment, income, wealth, expenditures, health, education, marriage, childbearing, philanthropy, and numerous other topics. However, the only information on children was limited to age, sex, and race.

In 1997, the PSID supplemented its main data collection with additional data on children and their development. The objective was to study how early family, school, and community experience influences later life as an adult. No more than two children aged twelve and under can be selected from each family for inclusion in the PSID-CDS, this process resulted in a total of 3586 children from over 2500 families to be recruited. Data were collected directly from the children if over three years old, and from their primary caregiver, secondary caregiver, absent parent, teacher, and school administrator on a broad array of developmental outcomes such as physical health, emotional well-being, academic achievement, cognitive ability, social relationships with family and peers, time
diaries, and much more. Data were collected again in 2002/2003 and 2007/2008 from children who remained under 18, including those who were born later than 1997. A detailed description of the PSID-CDS can be found in the User’s Guide (Institute for Social Research, the University of Michigan, 2010) and in Duncan, Hofferth, & Stafford (2002).

Comparisons of the Two Samples.

There are many similarities between the two samples besides their sample design as described above. First of all, the size of each of the two samples is comparable. In 1986, 4971 children born to the NLSY79 mothers were successfully interviewed, in comparison with the 3586 children from the over 2500 PSID families who were successfully interviewed in 1997. Both samples were born later than 1970 and the age of the two samples when they were first interviewed is close. The age of the children recruited by the NLSY79-C were under 15 in 1986 and the age of these children whose data were analyzed in the current study ranged from 5 to 13, while the age of the children recruited by the PSID-CDS were under 13 in 1997 and the age of these children whose data were analyzed in the current study ranged from 3 to 13. Both studies collected rich resources of data on prenatal care and birth circumstances, cognitive development, family context, community context, education, and school experiences. Data on prenatal care and birth circumstances include birth weight, gestational age, as well as birth order, age and marital status of the mother at birth of the child, maternal physical and mental conditions during pregnancy, maternal alcohol, cigarette, drug use during pregnancy, doctor visits during pregnancy, length of hospital stay, and at what age breastfeeding
stopped. Most of the data on prenatal care and birth circumstances were collected by means of mothers’ recall.

However, the cognitive abilities of the two samples were assessed using different cognitive tests. In the NLSY79-C, the Peabody Individual Achievement Tests for Math, Reading Recognition, and Reading Comprehension (Dunn & Markwardt, 1970) were employed to assess children’s math skills and reading capabilities. The Peabody Individual Achievement Tests were designed to be used with students in kindergarten-grade 12 or ages 5 through 18. This assessment is unique in that it provides a multiple-choice format, requiring a minimum amount of writing. While in the PSID-CDS, the tests employed were the Woodcock-Johnson Revised Tests of Achievement (WJ-R) for Letter-Word Identification, Passage Comprehension, Applied Problems, and Calculation Skills (Woodcock & Johnson, 1989). These tests may be used with individuals from age 2 through adult. The two nationally recognized achievement tests report results in raw score, standardized score, percentile rank, and grade or age equivalent, and the two tests were normed with children in both public and private schools throughout the United States. But to date, little empirical evidence exists to support comparability of the two tests in terms of their psychometric characteristics such as validity and reliability. Thus, any differences between the results from the two samples, if detected, could possibly be caused by the different tests they used. Besides the two achievement tests, the Digit Span subtest from the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) was used in both studies to assess short-term memory. The NLSY79-C also included Memory for Locations, Parts of the Body, the McCarthy Scales of Children’s Abilities, and the Peabody Picture Vocabulary Test-Revised as part of the cognitive test battery in 1986.
Another difference that needs to be noted is the 11-year time interval between the two samples. As described above, the ages of the children in the two samples when they were first interviewed were similar, but the children in the NLSY79-C were first interviewed in 1986, while the children in the PSID-CDS were first interviewed in 1997. This means that children in the PSID-CDS were born about 11 years later than the children in the NLSY-C, during a period when advances in neonatal care enabled almost all infants born in the United States to survive. This might make it reasonable to expect that birth weights of children in the PSID-CDS data set could have a wider range and a greater variety than the NLSY79-C sample. Little is known about how the general population has changed with respect to the relationship between perinatal characteristics and developmental outcomes. Using the same methods to analyze the data from the two samples could provide interesting information about changes in the general population in these ten years, but interpretations of the details seem likely to be difficult. Detailed information about the two samples, such as descriptive statistics, was provided in Chapter 4 Results.

Data Preparation

This study used only the data collected during the first wave of data collection in both studies, that is, the data collected in 1986 for the NLSY79-C and the data collected in 1997 for the PSID-CDS. And only these children who were singleton births rather than multiple births, who had their birth weight and gestational age recorded, and whose cognitive ability was assessed were included. The two datasets were prepared in multiple ways to facilitate data analyses.
Adjustment for age differences. The age of the children in the NLSY79-C whose data were analyzed ranged from 5 to 13 in 1986, while the age of the children in the PSID-CDS whose data were analyzed ranged from 3 to 13 in 1997. Therefore, variance in their performance in the cognitive tests could be the results of age differences at the time of testing. To adjust for age differences, the predicted score from a parametric regression model in which age is the sole predictor was subtracted from each observed cognitive test score.

Generation of cognitive composite score. The cognitive tests, including the Peabody Individual Achievement Tests used in the NLSY79-C and the Woodcock-Johnson Revised Tests of Achievement used in the PSID-CDS, aim to assess cognitive ability for multiple domains including calculation, problem solving, reading recognition, reading comprehension, and the like. That is, multiple test scores were collected from each child. To simplify the analyses, the residual of each cognitive test score after the value predicted by age was subtracted was standardized (converted to z-scores), and the standardized residuals (from the predictions from age) were added together to construct a composite score as the single outcome variable. The resultant composite score was then converted to z-scores. Graphics are used in the next chapter to further describe the process that was used.

Imputation of missing data. Only the data from the children who were singleton births rather than multiple births, who had their birth weight and gestational age recorded, and whose cognitive ability was assessed were analyzed. Therefore, there were no missing data in birth weight, gestational age, and cognitive test scores. But there was a considerable amount of missingness in the covariates that describe the prenatal
differences of those children. The missing values in these covariates were imputed using a multiple imputation procedure (Schafer, 1997). The multiple imputation procedure assumes the data generating mechanism to be multivariate normality and the missing data mechanism to be ignorable missing at random (MAR). A formal definition of MAR states that probabilities of missingness are related to observed values but not to missing values. There is no statistical test to prove this assumption. Ten imputed data sets were produced and the imputed values were averaged across the ten data sets to reduce imputation uncertainty.

Data Analysis

Loess Regression to Depict the Relationship of Birth Weight with Cognitive Ability

Since the relationship between birth weight and cognitive ability in childhood is expected to vary across subgroups such as defined by sex and race, each full sample was split into subsamples based on their sex and race, then the relationship was studied using nonparametric loess regression for each subsample, to be specific, first for whites and blacks separately, and then for white males, white females, black males, and black females separately. To answer the first research question, “What is the nature of the predictive association of birth weight and gestational age for cognitive ability in childhood?”, for each subsample defined by race and sex, a locally linear regression was conducted with both birth weight and gestational age being the predictors and with cognitive composite score being the response. It was supplemented by a fitted locally linear regression surface relating the cognitive outcome to both birth weight and gestational age. The objective was to study if gestational age accounted for additional variance in cognitive ability besides what was accounted for by birth weight. Interpretation of the fitted surfaces was also aided by the comparable graphs produced by
the function $scatter3d$, although they are not presented. To answer the second research question, “What is the nature of the predictive association of birth weight for cognitive ability in childhood when gestational age is restricted within the normal range?”, birth weight was the sole predictor in the locally linear regression while gestational age was restricted within the normal range, i.e. between 37 weeks (259 days) and 42 weeks (294 days) to control for its confounding effects. For each subsample analysis, a two-dimensional scatterplot of the cognitive outcome against birth weight with a locally linear regression curve overlaid was used to display both the data and the association.

The loess regression models that include both birth weight and gestational age as the predictors appear earlier in the current study than simple loess regression models that include only birth weight as the sole predictor. Readers may find it useful to examine the simple loess regression results in the second section below before studying results based on two predictors.

**Generalized propensity score methods to control for confounding**

To answer the third research question, “Do observed associations increase or decrease after adjusting for numerous confounders that concern prenatal differences?”, generalized propensity score methods were used to reduce effects of multiple covariates that may confound the relationship between birth weight and cognitive ability in childhood. Most propensity score applications have been developed to control for confounding across two groups: treatment and control. Generalized propensity score methods, in which McCullagh’s (1980) ordinal logit model is used to estimate a balancing score for each individual, are able to accommodate treatment with multiple doses (e.g., high, medium, low, control). Since birth weight is continuously distributed,
no simple dichotomization is sufficient for any study of the relationship between birth weight and cognitive ability. The generalized methods provide a means for propensity score adjustment for confounding to be feasible in the situations where the distribution of birth weight is broken into multiple categories on a continuum, and also to make adjustments for multiple covariates that can confound interpretations of effects.

The stage that employed generalized propensity score methods was conducted separately for the whites and the blacks whose gestational age was in the normal range, i.e. between 37 weeks (259 days) and 42 weeks (294 days). A search of the relevant literature had found that the associations of the cognitive test scores with birth weight tended to positive, so that cognitive test scores improved with increases in birth weight. But a slight decline in the test scores among the children with high birth weight (>= 4000g) was also observed in some studies. Therefore, the birth weight distribution of each racial group was divided into three continuously positioned groups with the children whose birth weights were positioned in the middle as the group in the middle and with the children whose birth weights positioned on either end as the other two groups. If the association of the cognitive score with birth weight is a continuously increasing one, we would expect, after the effects of multiple confounders are controlled by generalized propensity score methods, that the children with high birth weights would have the highest cognitive score. If there is a decline among the children with high birth weight, we would expect the group in the middle would have the highest cognitive score. The determination of cutting points was informed not only by literature but also by the results from simple locally linear regression of cognitive composite score on birth weight when
gestational age was restricted to be normal. The determination of cutting points also took account of racial differences in birth weight and increases in birth weight over time.

Covariates to be controlled for included those that summarize prenatal differences and hence are unaffected by birth weight, such as maternal age in the birth year of the child, maternal highest educational attainment by the birth year of the child, maternal physical and mental conditions during pregnancy, maternal smoking and drinking habits during pregnancy, and socioeconomic status of the family in the birth year of the child. Covariates also included those whose quantity is a constant such as birth order and maternal IQ. McCullagh’s (1980) ordinal logit model was used to estimate the balancing score with the main effects of multiple confounding covariates and their polynomial terms and interactions being the predictors. Several ordinal logit models were compared and the optimal model was the most parsimonious one that best balanced the covariates.

Then a fitted locally linear regression surface was generated to relate cognitive composite score to both birth weight and the balancing score. The purpose was to show how the association of cognitive composite score with birth weight compared across the distribution of the balancing score and if adding the balancing score as another predictor modified the association.

The results were also presented numerically. The three groups of children were divided into 5 strata according to the quintiles of their balancing score, and cognitive composite score was compared across the three birth weight categories within each stratum. The number of children within birth weight category \( j \) and within stratum \( i \) and their mean cognitive score were summarized in a 5 by 3 table. In addition, the overall
mean score for the children within each birth weight category after stratifying on the balancing score to adjust for covariate differences was estimated as:

$$\bar{y}_{Ej} = \sum_{i=1}^{5} \frac{1}{5} \bar{y}_{ij}$$  (Conniffe, Gash & O’Connell, 2000)

Where $\bar{y}_{Ej}$ is the estimated mean score for children within birth weight category $j$ and $\bar{y}_{ij}$ is the observed mean among children within birth weight category $j$ and within stratum $i$. And its corresponding standard error was estimated as:

$$\text{s.e.}(\bar{y}_{Ej}) = \frac{1}{5} \sqrt{\sum_{i=1}^{5} \frac{s_{ij}^2}{n_{ij}}}$$  (Conniffe, Gash, & O’Connell, 2000)

Where $s_{ij}^2$ is the sample variance among children within birth weight category $j$ and within stratum $i$, and $n_{ij}$ is the number of children within birth weight category $j$ and within stratum $i$.

![Figure 3.1 Outline of data analysis.](image)

**Figure 3.1** Outline of data analysis.
Chapter 4 Results

Results for the NLSY79-C

Loess Regression to Depict the Relationship of Birth Weight with Cognitive Ability

In 1986, for the NLSY79-C, a total of 4971 children under age 15 and their parents and teachers were successfully interviewed. Birth weight, gestational age, and most of the potential confounding covariates were collected by means of mothers’ recall. These children included those who were born before 1979 when the mothers were first interviewed and those whose usual residence was outside the mother’s household.

Birth weight was recorded in ounces and converted into grams. Gestational age was estimated in complete weeks from the first day of the mother’s last menstrual period to the birthday of the child.

Multiple measures were employed to assess the cognitive capabilities of these children, which included Memory for Location, Parts of the Body, the McCarthy Scales of Children’s Abilities, the Peabody Picture Vocabulary Test-Revised, the Digit Span subtest from the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974), and the Peabody Individual Achievement Tests (PIAT) Math, Reading Recognition, and Reading Comprehension (Dunn & Markwardt, 1970).

The PIAT measures academic achievement of children age five and over. The math subtest contains multiple-choice items that test knowledge and application of math concepts and facts. The reading recognition subtest includes items that measure recognition of printed letters and the ability to read words aloud. In the reading comprehension subtest, the student chooses one of four pictures that best illustrates a sentence.
Race of these children was coded as Hispanics, blacks, and non-black-non-Hispanics. Since the sample size for Hispanics is small, only the 1434 non-black-non-Hispanic and black children who are singleton births rather than multiple births, who had their birth weight and gestational age recorded, and who completed the two PIAT subtests, math and reading recognition in 1986 were included in the analyses. The non-Black-non-Hispanic sample was treated as a counterpart to the white sample from the PSID-CDS. According to US census data (Censusscope, 2011), in the 1980s, less than 2.5% of the US population were American Indians, Asians, Hawaiian and Pacific Islanders, and others. Therefore, it was assumed that the majority of the non-black-non-Hispanic children are whites, with a small proportion of 2.5% possibly being minorities.

Figure 4.1 displays the scatterplot matrix for the three PIAT subtest scores of the 1434 non-black-non-Hispanic and black children in the NLSY79-C. The score from the reading comprehension subtest was excluded from the analyses because there is a perfect correlation between the reading recognition score and the reading comprehension score among those children who scored low in both subtests as shown in Figure 4.1. The impossible perfect correlation implies there were some unknown manipulations, if not mistakes, made to the data. In this case, reading recognition was chosen over reading comprehension because it has fewer missing values.
Figure 4.1 Scatterplot matrix for the three PIAT subtest scores of the 1434 non-black-non-Hispanic and black children in the NLSY79-C.

Adjustment for age differences. The variable “Age of Child at 1986 Interview Date of Mother” was used as the age variable in this study. Age of child at 1986 interview date of mother was recorded in months and was converted into years. The ages of these 1434 children varied from 5 to 13. The scatterplots of age against each of the two cognitive scores (Figure 4.2) show how the cognitive score increases as age increases, which confirms that the variance in these students’ performance in the two cognitive tests is strongly related to their age differences at the time of testing. In the two scatterplots, the black straight line shows the predicted score from the linear regression with age being the predictor and the cognitive score being the response, the dotted black line shows the
predicted score from the quadratic polynomial regression for age, and the dashed grey line shows a loess fit. The quadratic regression fits the data better than the linear regression as shown in Figure 4.2, and is closer to that of loess. Therefore, because it entails use of an explicit formula to adjust for age differences at the time of testing, the predicted score from the quadratic regression was subtracted from each observed score to obtain residuals which were taken to be the adjusted cognitive scores for subsequent analyses.

*Figure 4.2* Scatterplots of age against each cognitive score with loess, linear, and quadratic fits superimposed on the data for the 1434 non-black-non-Hispanic and black children in the NLSY79-C.

*Generation of cognitive composite score.* As can be seen in Figure 4.1 the math score is positively correlated with the reading recognition score and the magnitude is reasonably strong. The positive correlation is evident even after adjusting for age differences at the time of testing. The residuals for the two cognitive scores, after subtracting variance explained by age (the quadratic fit), were standardized and added
together to generate a cognitive composite score. The resultant composite score was then converted to a z-score.

Table 4.1 presents the descriptive statistics for birth weight, gestational age in weeks, and cognitive composite score for the non-black-non-Hispanic (n=781) and black (n=653) children. The mean birth weight of these non-black-non-Hispanic children is 3326 grams, and the minimum birth weight is 936 grams while the maximum is 4933 grams. The mean birth weight of these black children is 3060 grams, about 250 grams lower than that of the non-black-non-Hispanics. The minimum birth weight of these black children is 907 grams while the maximum is 5613 grams. The gestational ages of these non-black-non-Hispanic children have a range between 26 weeks and 46 weeks and a mean of 39 weeks, while the gestational ages of these black children have a range between 26 weeks and 45 weeks and a mean of 38 weeks, slightly lower. The cognitive composite scores of these non-black-non-Hispanic children have a range between -3.04 and 3.78 and a mean of 0.19. The mean cognitive composite score of these black children is much lower, being -0.23.
Table 4.1

*Descriptive Statistics for Birth Weight, Gestational Age in Weeks, and Cognitive Composite Score for Non-Black-Non-Hispanic (n=781) and Black (n=653) Children in the NLSY79-C*

<table>
<thead>
<tr>
<th></th>
<th>Non-black-non-Hispanics</th>
<th>blacks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>birth weight</td>
<td>gestational age</td>
</tr>
<tr>
<td>mean</td>
<td>3326</td>
<td>39</td>
</tr>
<tr>
<td>sd</td>
<td>561</td>
<td>2.37</td>
</tr>
<tr>
<td>median</td>
<td>3345</td>
<td>39</td>
</tr>
<tr>
<td>minimum</td>
<td>936</td>
<td>26</td>
</tr>
<tr>
<td>maximum</td>
<td>4933</td>
<td>46</td>
</tr>
<tr>
<td>skewness</td>
<td>-0.44</td>
<td>-1.37</td>
</tr>
<tr>
<td>kurtosis</td>
<td>1.42</td>
<td>5</td>
</tr>
</tbody>
</table>

The scatterplot matrices in Figure 4.3 show relationships between the cognitive composite score, gestational age in weeks, birth weight, and sex for two subsamples defined by race in the NLSY79-C; these provide details about the joint distributions of the variables. For the non-black-non-Hispanics, their cognitive composite scores are mainly distributed between -2 and 2.5 (z-score units) and their gestational ages are mainly distributed between 34 weeks and 44 weeks, and their birth weights are mainly distributed between 2000 grams and 4500 grams. Birth weight is moderately correlated with gestational age as expected ($r = 0.47$). And it is slightly positively correlated with cognitive composite score ($r = 0.06$). Cognitive composite score tends to increase as birth weight increases. There is also a small positive association between gestational age and cognitive composite score ($r = 0.03$). We also see non-black-non-Hispanic males (coded
as 1) are heavier in birth weight in comparison with non-black-non-Hispanic females (coded as 2), but there are hardly any evident differences between males and females in their gestational ages and cognitive composite scores. For the blacks, their cognitive composite scores are mainly distributed between -2.5 and 2, their gestational ages are mainly distributed between 34 weeks and 42 weeks, and their birth weights are mainly distributed between 1500 grams and 4500 grams. Their birth weights are moderately correlated with their gestational ages ($r = 0.49$), but only modestly correlated with their cognitive composite scores ($r = 0.04$). The gestational age variable is essentially uncorrelated with and the cognitive composite score ($r = 0.01$). Black males (coded as 1) have higher birth weight than black females (coded as 2), but there are hardly any evident differences between males and females in their gestational ages and cognitive composite scores.

Figure 4.3 Scatterplot matrices for cognitive composite score, gestational age in weeks, birth weight, and sex for the non-black-non-Hispanic (n=781) and the black (n=653) children in the NLSY79-C.
Locally linear regression of cognitive composite score on birth weight and gestational age. To answer the first research question, “What is the nature of the predictive association of birth weight and gestational age for cognitive ability in childhood?” for each subsample defined by race and sex, a locally linear regression was conducted with both birth weight and gestational age being the predictors and with cognitive composite score being the response. The objective was to study whether gestational age accounted for additional variance in cognitive composite score besides what was accounted for by birth weight. (In the loess function, the degree was set to 1 to specify locally linear regression, and the span was set as 0.85.)

A locally linear regression was first fitted to the data of the 781 non-black-non-Hispanic children. Figure 4.4 shows the fitted surface relating cognitive composite score to the independent variables birth weight and gestational age for these children. The graphic on the left gives a clear view of the surface using birth weight as the key predictor, and the graphic on the right gives a clear view using gestational age as the key predictor. The data are mainly distributed within the interval from 2000 grams to 4500 grams along the distribution of birth weight with a number of outliers on both ends and within the interval from 34 weeks to 44 weeks along the distribution of gestational age with a number of outliers also on both ends. The surface shows no consistent association between cognitive composite score and birth weight in the area where most data are distributed. For those whose gestational ages are about 40 weeks or below, their cognitive composite scores increase as their birth weights increase. For those whose gestational ages are beyond 40 weeks, their cognitive composite scores decrease as their birth weights increase. The surface also shows no consistent association between cognitive
composite score and gestational age in the area where most data are distributed. For those children with extremely low birth weights, their cognitive composite scores increase as their gestational ages increase, while for those children with extremely high birth weights, the association is upside-down U-shaped.

Figure 4.4 Fitted surface for locally linear regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 781 non-black-non-Hispanic children in the NLSY79-C.

The non-black-non-Hispanic sample was further divided into males and females. A locally linear regression was fitted to males and then to females. The two graphics above in Figure 4.5 demonstrate the fitted surface for the loess regression of cognitive composite score on birth weight and gestational age for the non-black-non-Hispanic males with the one on the left giving a clear view of the surface using birth weight as the key predictor and the one on the right giving a clear view using gestational age as the key predictor. The surface shows no consistent association between cognitive composite score and birth weight in the area where most data are distributed. For those whose gestational ages are about 40 weeks or below, their cognitive composite scores increase as their birth weights increase. For those whose gestational ages are beyond 40 weeks,
their cognitive composite scores decrease as their birth weights increase. The surface also shows no consistent association between cognitive composite score and gestational age in the area where most data are distributed.

The two graphics below in Figure 4.5 demonstrate the fitted locally linear regression surface for the non-black-non-Hispanic females. The surface shows no consistent association either between cognitive composite score and birth weight at any fixed level of gestational age or between cognitive composite score and gestational age at any fixed level of birth weight.
Figure 4.5 Fitted surfaces for locally linear regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 401 non-black-non-Hispanic males (the two graphics above) and for the 380 non-black-non-Hispanic females (the two graphics below) in the NLSY79-C.

Figure 4.6 shows the fitted surface for the locally linear regression of cognitive composite score on birth weight and gestational age for the 653 blacks. The data are mainly distributed within the interval from 1500 grams to 4500 grams along the distribution of birth weight with a number of outliers on both ends and within the interval from 34 weeks to 42 weeks along the distribution of gestational age with a number of outliers also on both ends. The surface shows no consistent association between cognitive
composite score and birth weight. For those with very short gestational ages, their cognitive composite scores increase as their birth weights increase, but for those whose gestational ages are within the normal range, the association is upside-down U-shaped, and the association changes gradually to a negative one for those whose gestational ages are beyond the normal range. The surface also shows that the association between cognitive composite score and gestational age in the area where most data are distributed is upside-down U-shaped, suggesting there is an optimum gestational age around 35 weeks.

Figure 4.6 Fitted surface for locally linear regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 653 blacks in the NLSY79-C.

The black sample was further divided into males and females. The two graphics above in Figure 4.7 demonstrate the fitted surface for the locally linear regression of cognitive composite score on birth weight and gestational age for the black males. The surface shows a positive association between cognitive composite score and birth weight and an upside-down U-shaped association between cognitive composite score and gestational age with an optimum gestational age being around 35 weeks.
The two graphics below in Figure 4.7 demonstrate the three-dimensional scatterplot with a fitted surface for the black females. The surface shows no consistent association either between cognitive composite score and birth weight at any fixed level of gestational age or between cognitive composite score and gestational age at any fixed level of birth weight.

![Figure 4.7 Fitted surfaces for locally linear regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 347 black males (the two graphics above) and for the 306 black females (the two graphics below) in NLSY79-C.](image)

*Simple locally linear regression of cognitive composite score on birth weight.* To answer the second research question, “What is the nature of the predictive association of
birth weight for cognitive ability in childhood when gestational age is restricted within the normal range, i.e., between 37 weeks and 42 weeks?” those children whose gestational ages are below or beyond the normal range were excluded from the sample. The birth weights of these non-black-non-Hispanic children whose gestational ages are normal range from 1500 grams to 5000 grams, while the birth weights of these black children whose gestational ages are normal range from 1000 grams to 5000 grams, with one outlier being 5613 grams.

The top graph in Figure 4.8 is a two-dimensional scatterplot of birth weight against cognitive composite score in which non-black-non-Hispanics and blacks define the two separate groups. The data for the non-black-non-Hispanics are distributed slightly to the upper right of those of the blacks.

Locally linear simple regression curves were overlaid separately for the two racial groups with the degree set to 1 and the span set as 0.85. The two curves are displayed again in the second graph in Figure 4.8 by a scale that can capitalize the association pattern. The loess curve of the non-black-non-Hispanics (the black line) shows that for those non-black-non-Hispanic children whose gestational ages are normal, their cognitive composite scores increase gradually with birth weight until the reach of an optimum weight of about 3600 grams; after that point composite scores begin to fall with increasing birth weights. The mean cognitive score of those birth weights below 3000 grams is 0.15 (n = 130, 95% CI = [-0.006, 0.3]), and the mean cognitive score increases to 0.25 for birth weights between 3000 grams and 4000 grams (n = 463, 95CI = [0.17, 0.33]), and then decreases to 0.02 for birth weights beyond 4000 grams (n= 61, 95CI = [-0.23, 0.27]).
The loess curve for those black children whose gestational age is normal (the grey line) is almost flat in the area where most data are distributed with a slight decline on either end. The mean cognitive score of those birth weights below 2800 grams is -0.25 (n = 123, 95%CI = [-0.43, -0.07]), and the mean cognitive score increases to -0.2 for birth weights between 2800 grams and 3800 grams (n = 378, 95CI=[-0.3, -0.1]), and then decreases to -0.42 for birth weights beyond 3800 grams (n= 51, 95CI=[-0.65, -0.19]).

The upper horizontal dashed line shows the mean cognitive composite score for these non-black-non-Hispanic children, while the lower horizontal dashed line shows the mean cognitive composite score for the black children. The lines are separated from each other, indicating notable racial differences on average. The SMD between the mean of the non-black-non-Hispanics and that of the blacks is 0.45, i.e., just under one standard deviation unit for these data.
Figure 4.8 2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by race for the 1206 non-black-non-Hispanic and black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks.

The top graph in Figure 4.9 is a two-dimensional scatterplot that shows how average cognitive composite scores change across the distribution of birth weight for the non-black-non-Hispanics whose gestational ages are within the range from 37 weeks to
42 weeks in the case of the NLSY79-C data. The mean difference, standardized “effect size”, is 0.03. Locally linear regression curves were overlaid separately for the males (the black line) and for the females (the grey line). The two curves are magnified in the second graph in Figure 4.9. The loess curve for the males shows that for those non-black-non-Hispanic males whose gestational ages are normal, cognitive composite scores increase slightly with increases in their birth weight until an optimum weight around 3600 grams, after that composite scores begin to fall. The loess curve for the females shows that for those non-black-non-Hispanic females whose gestational ages are normal, their cognitive composite scores improve slightly with increases in their birth weights until an optimum weight around 3600 grams, after that composite scores begin to fall. The horizontal dashed line shows the mean cognitive composite score for these non-black-non-Hispanic children. Both the loess curve for males and that for females are sufficiently close to the horizontal dashed line to indicate that gender differences are minor.
Figure 4.9 2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by gender for the 654 non-black-non-Hispanic children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks.

The top graph Figure 4.10 shows a two-dimensional scatterplot generated for the blacks whose gestational ages are within the normal range from 37 weeks to 42 weeks with locally linear regression curves overlaid separately for the black males (the black
line) and for the black females (the grey line). The two curves are magnified in the graph below. The loess curve for the males shows that for those black males whose gestational ages are normal, cognitive scores improve slightly with increases in their birth weights. But this growth pattern is not evident for those black females whose gestational ages are also normal, and the slight decline at the upper end was mainly caused by the outlier being 5613 grams. The horizontal dashed line shows the mean cognitive composite score for these black children. The two loess curves are sufficiently close to the horizontal dashed line to indicate gender differences are minor. The mean difference, standardized “effect size” (SMD), is -0.13.
Figure 4.10 2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by gender for the 552 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks.

Generalized propensity score methods to control for confounding.

To answer the third research question, “If any of the observed associations are going to increase or decrease after confounding of numerous covariates that account for
prenatal differences is controlled?” generalized propensity score methods were used to accomplish the goal of reducing the effects of multiple covariates that may confound the relationship between birth weight and cognitive ability in childhood.

The results from locally linear regression of cognitive composite score on birth weight while gestational age was restricted to be within the normal range from 37 weeks to 42 weeks show that there are notable differences between the non-black-non-Hispanics and the blacks, but gender differences within each racial group are trivial. Furthermore, if each racial group is further divided by gender, the sample size will be too small to make loess regression resistant to outliers. Therefore, propensity score analysis was conducted only for the non-black-non-Hispanic sample and for the black sample which include both the males and the females whose gestational ages are normal.

The results from locally linear regression of cognitive composite score on birth weight show that the association between birth weight and cognitive composite score has an upside-down U shape. For the 654 non-black-non-Hispanics whose gestational ages are normal, their cognitive composite scores increase gradually with birth weights until an optimum weight around 3600 grams, after that their composite scores begin to fall. The loess curve for the 552 black children whose gestational ages are normal is almost flat in the area where most data are distributed, but there is also a slight decline on either end. To learn whether there is an optimum range in the middle, the distribution of birth weight of the non-black-non-Hispanics was divided into three categories by 3000 grams and by 4000 grams as shown in Figure 4.11. Similarly, the distribution of birth weight of the blacks was divided into three categories by 2800 grams and by 3800 grams because they have averagely lower birth weights. The determination of the cutting points was
based on observation of the loess curves but was also informed by the literature. The literature found the cognitive test scores improve with increasing birth weights, but some researchers, such as Jefferis, Power, & Hertzman (2002) and Richards et al. (2002), also observed a slight decline among the children whose birth weights are beyond 4000 grams.

Figure 4.11 Plot to demonstrate division of the birth weight distribution of the 654 non-black-non-Hispanic children and that of the 551 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks.

In this case, generalized propensity score methods in which a balancing score estimated from ordinal logit regression serves as the basis to stratify the three groups of children are suitable. Covariates to be adjusted for included sex, age of the child by 1986, maternal age in the birth year of the child, birth order to mother, marital status of mother in the birth year of the child, mother’s IQ, mother’s highest educational attainment by the birth year of the child, family net income in the birth year of the child, maternal weight
before pregnancy, if the mother drank alcohol or smoked during pregnancy, and in which month during pregnancy the mother first visited a medical person for prenatal care.

The covariate of birth order has possible values from 1 to 5. Since in these samples there were very few children not born as the first child in the family, this covariate was recoded as a binary variable with these children born as the first child in the family coded as 1, and those not born first coded as 2. Mother’s IQ was assessed in 1980 by the Armed Forces Qualification Test (AFQT). AFQT scores are not raw scores, but rather percentile scores indicating how each examinee performed compared with all other examinees. Thus, someone who receives an AFQT of 55 scored better than 55 percent of all other examinees. The maximum possible score is 99 as a person can do better than 99 percent of those who took the test, but he or she cannot do better than himself, so the high percentile is 99. Family net income in the birth year of the child was converted to year 1979 US dollars by adjusting for inflation according to the Consumer Price Index. The mothers started visiting a medical person for prenatal care as early as in the first month of their pregnancy or as late as in the seventh month. There were also mothers who didn’t visit a medical person during their pregnancy. Their value for this covariate was coded as 11.

Results for the non-black-non-Hispanics in the NLSY79-C whose gestational age is normal. Propensity score analysis was first conducted for the 654 non-black-non-Hispanics whose gestational ages are within the normal range from 37 weeks to 42 weeks.

The covariates “marital status of mother in the birth year of the child”, “mother’s highest educational attainment by the birth year of the child”, and “family net income in
the birth year of the child”, the three variables extracted from the NLSY79 main data file have the highest number of missing values as shown in Table 4.2. A number of the children interviewed in 1986 were born before 1979 when the NLSY79 study started collecting data, therefore, only the parents of those children who were born in 1979 and after had their information collected on the three covariates. The missing values in these covariates were imputed using a multiple imputation procedure (Schafer, 1997). The multiple imputation procedure is based on the assumption of multivariate normality and the missing data mechanism is taken to be ignorable missing at random (MAR). The extreme values for skewness and kurtosis of some covariates, such as “month in pregnancy of 1st doctor visit”, could be the evidence for violation of the multivariate normality assumption. And the assumption of MAR can hardly hold in this case either for the reason that has been explained above. However, simulation studies demonstrate that the multiple imputation procedure provides less biased estimates as compared to analyses using complete data even when the MAR assumption can not be totally defended (Vargas-Chanes, 2000). Considering the large number of missing values in some covariates, ten imputed data sets were produced and the imputed values were averaged across the ten data sets to reduce imputation uncertainty.
### Table 4.2

Descriptive Statistics of Covariates for the 654 Non-Black-Non-Hispanic Children in the NLSY79-C Whose Gestational Age Is in the Normal Range from 37 Weeks to 42 Weeks

<table>
<thead>
<tr>
<th>Covariate</th>
<th>mean/percentage</th>
<th>sd</th>
<th>median</th>
<th>min</th>
<th>max</th>
<th>skew</th>
<th>kurtosis</th>
<th>n of missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>50% (males)</td>
<td>0.5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.01</td>
<td>-2.01</td>
<td>0</td>
</tr>
<tr>
<td>ageby1986</td>
<td>7</td>
<td>1.86</td>
<td>7</td>
<td>5</td>
<td>13</td>
<td>0.83</td>
<td>-0.05</td>
<td>0</td>
</tr>
<tr>
<td>mothers.age</td>
<td>19</td>
<td>2.04</td>
<td>19</td>
<td>14</td>
<td>23</td>
<td>-0.02</td>
<td>-0.69</td>
<td>0</td>
</tr>
<tr>
<td>birth.order.to.mother</td>
<td>70% (first born)</td>
<td>0.46</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.86</td>
<td>-1.26</td>
<td>0</td>
</tr>
<tr>
<td>marital.status.of.mother</td>
<td>70% (married)</td>
<td>0.47</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-0.79</td>
<td>-1.38</td>
<td>309</td>
</tr>
<tr>
<td>marital.status.of.mother*</td>
<td>71% (married)</td>
<td>0.46</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-0.91</td>
<td>-1.16</td>
<td>NA</td>
</tr>
<tr>
<td>mothers.IQ</td>
<td>39</td>
<td>25.39</td>
<td>35</td>
<td>1</td>
<td>99</td>
<td>0.45</td>
<td>-0.75</td>
<td>22</td>
</tr>
<tr>
<td>mothers.IQ*</td>
<td>39</td>
<td>25.04</td>
<td>35</td>
<td>1</td>
<td>99</td>
<td>0.47</td>
<td>-0.68</td>
<td>NA</td>
</tr>
<tr>
<td>mothers.education</td>
<td>11</td>
<td>1.83</td>
<td>12</td>
<td>6</td>
<td>16</td>
<td>-0.32</td>
<td>0.31</td>
<td>470</td>
</tr>
<tr>
<td>mothers.education*</td>
<td>10</td>
<td>1.89</td>
<td>11</td>
<td>5</td>
<td>16</td>
<td>-0.1</td>
<td>-0.22</td>
<td>NA</td>
</tr>
<tr>
<td>Income</td>
<td>11766</td>
<td>8079.29</td>
<td>10573</td>
<td>0</td>
<td>52863</td>
<td>1.58</td>
<td>4.92</td>
<td>382</td>
</tr>
<tr>
<td>income*</td>
<td>12067</td>
<td>5627.31</td>
<td>11856</td>
<td>0</td>
<td>52863</td>
<td>1.83</td>
<td>10.53</td>
<td>NA</td>
</tr>
<tr>
<td>maternal.weight.before.pregnancy</td>
<td>124</td>
<td>19.54</td>
<td>120</td>
<td>83</td>
<td>200</td>
<td>0.76</td>
<td>0.71</td>
<td>6</td>
</tr>
<tr>
<td>maternal.weight.before.pregnancy*</td>
<td>124</td>
<td>19.46</td>
<td>120</td>
<td>83</td>
<td>200</td>
<td>0.76</td>
<td>0.73</td>
<td>NA</td>
</tr>
<tr>
<td>drink.alcohol</td>
<td>46% (drank)</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.14</td>
<td>-1.99</td>
<td>0</td>
</tr>
<tr>
<td>if.smoke</td>
<td>51% (smoked)</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-0.05</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>month.in.pregnancy.of.1st.doctor.visit</td>
<td>3</td>
<td>1.7</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>2.42</td>
<td>8.46</td>
<td>18</td>
</tr>
<tr>
<td>month.in.pregnancy.of.1st.doctor.visit*</td>
<td>3</td>
<td>1.7</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>2.39</td>
<td>8.44</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: * represents the covariates after their missing values were imputed.

Table 4.2 presents the mean, standard deviation, median, minimum, maximum of each covariate, as well as the skewness and kurtosis of its distribution. As shown in this table, for example, the average age of these non-black-non-Hispanic children when they were interviewed is 7 years old; and the average age of the mothers when their child was born is 19 years old. Most the mothers were married when their child was born because the distribution of “marital status of mother” is negatively skewed. The covariate “drank alcohol” is positively skewed, while the covariate “if smoke” is negatively skewed, indicating there are fewer mothers who drank alcohol during pregnancy than those who didn’t, but there are more mothers who smoked during pregnancy than those who didn’t.
Another major objective of Table 4.2 is to show changes in these statistics that come from imputation of missing values. For those covariates with missing values, descriptive statistics were presented twice: the ones without “*” are descriptive statistics before their missing values were imputed, the ones with “*” are descriptive statistics after missing values were imputed. Mother’s highest educational attainment and family net income have the highest numbers of missing values, there are several notable changes in their descriptive statistics after missing values were imputed.

According to the definition, a confounder must be an ancestor (cause) of the outcome; and it must be associated with treatment, but not be affected by the treatment (Greenland, Pearl, & Robins, 1999, Robins & Morgenstern, 1987; Weinberg, 1993). Therefore, BIRTH, a three-level ordinal categorical variable was created, with those whose birth weights are <= 3000 grams coded as 1, and those whose birth weights are between 3000 grams and 4000 grams coded as 2, and those whose birth weights are > 4000 grams coded as 3. The correlations of each covariate with BIRTH and with cognitive composite score were examined and are reported in Table 4.3.
Table 4.3

*Correlations of Covariates with BIRTH and Cognitive Composite Score for the 654 Non-Black-Non-Hispanic Children in the NLSY79-C Whose Gestational Age Is in the Normal Range from 37 Weeks to 42 Weeks*

<table>
<thead>
<tr>
<th></th>
<th>BIRTH</th>
<th>cognitive composite score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>-0.16</td>
<td>-0.02</td>
</tr>
<tr>
<td>ageby1986</td>
<td>-0.04</td>
<td>0.1</td>
</tr>
<tr>
<td>mothers.age</td>
<td>0.11</td>
<td>-0.09</td>
</tr>
<tr>
<td>birth.order.to.mother</td>
<td>0.02</td>
<td>-0.2</td>
</tr>
<tr>
<td>marital.status.of.mother</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>mothers.IQ</td>
<td>0.14</td>
<td>0.26</td>
</tr>
<tr>
<td>mothers.education</td>
<td>0.1</td>
<td>0.08</td>
</tr>
<tr>
<td>maternal.weight.before.pregnancy</td>
<td>0.23</td>
<td>-0.03</td>
</tr>
<tr>
<td>drink.alcohol</td>
<td>-0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>if.smoke</td>
<td>-0.18</td>
<td>-0.03</td>
</tr>
<tr>
<td>month.in.pregnancy.of.1st.doctor.visit</td>
<td>-0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Income</td>
<td>0.12</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Sex, maternal age in the birth year of the child, mother’s IQ, mother’s highest educational attainment, maternal weight before pregnancy, if the mother smoked during pregnancy, and family net income in the birth year of the child are related to BIRTH, either positively or negatively, with an absolute magnitude greater than 0.1. Non-black-non-Hispanic males tend to be heavier than females. The mothers who had children when they were older tended to have higher IQs than their counterparts who had children at younger ages. Also, older mothers tended to be better educated than their younger counterparts. Also the heavier the mother was before pregnancy, the heavier the birth weight of the child tended to be. If the mother smoked during pregnancy, the child had a higher risk of being low birth weight; the same was true of children born to poor families. But age of the child by 1986, birth order to mother, marital status of mother, if the mother drank alcohol during pregnancy, and in which month during pregnancy the mother first
visited a medical person for prenatal care shows little relationship with the three category variable BIRTH.

Age of the child by 1986, maternal age in the birth year of the child, birth order to mother, marital status of the mother when the child was born, mother’s IQ, mother’s highest educational attainment, if the mother drank alcohol during pregnancy, and family net income in the birth year of the child are notably related, either positively or negatively, with cognitive composite score. But sex, maternal weight before pregnancy, whether the mother smoked during pregnancy, and in which month during pregnancy the mother first visited a medical person for prenatal care show little relationship with cognitive composite score.

The main effects of all the 12 covariates, no matter whether they can differentiate those with high birth weights from those with low birth weights, were first included in the ordinal logit model as predictors to estimate the balancing score. Polynomial terms and interaction terms involving the covariates were tried in the ordinal logit model as an additional term one after another. The additional term would be kept in the model if covariate balance was improved after the sample was stratified based on the quintiles of the balancing score estimated from that model. This iterative process led into a final model which includes the main effects of the 12 covariates and a quadratic term for mother’s age. \( \hat{b}(x_i) = p(y \geq 2) = \frac{1}{1 + e^{(\alpha x_{i2} + \beta^T x_i)}} \) was used as the balancing score to stratify the three groups of children into five strata based on the quintiles of \( \hat{b}(x_i) \).

The next step was to assess overlap between the distributions of the balancing score across the three birth weight categories. Specifically, the goal concerned whether the balancing scores overlapped (or not) across all the three birth weight categories. The
comparison was then limited to the children whose balancing score was positioned in the overlapping region, with the children with extreme balancing scores outside the overlapping region (i.e., the children without comparable children in the other two groups) being excluded. Figure 4.12 shows how the distributions of the balancing score overlap each other across the three birth weight categories. It is shown that there is a balancing score region from 0.6 to 0.95 that contains scores for children from all the three birth weight categories, and this region can therefore support comparisons across the three groups.
Figure 4.12 Histograms showing overlap between the balancing score distributions for the 654 non-black-non-Hispanic children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks.

If the balancing score is appropriately estimated, covariates should look balanced within each stratum as if the children had been randomly assigned to one of the birth weight categories. Imbalanced covariates are those that can differentiate lower birth weight

weights from higher birth weights, but are not sufficiently well reflected in the balancing score (Zanutto, Lu, & Hornik, 2005).

Because generalized propensity score methods are relatively new, researchers haven’t reached full agreement about what the best methods could be to assess covariate balance. R function \textit{cbal.psa} (Pruzek & Helmreich, 2009), based on comparisons of SMDs defined as covariate differences in means divided by pooled standard deviation, was used to evaluate covariate balance. The \textit{cbal.psa} plot was originally developed for two-group (treatment and control) comparison, therefore, in the first comparison, the children whose birth weights are in the first category (≤ 3000 grams) were compared with the children whose birth weights are in the aggregated second and third categories (> 3000 grams); in the second comparison, the children whose birth weights are in the aggregated first and second categories (≤4000 grams) were compared with the children whose birth weights are in the third category (> 4000 grams). The plot on the left in Figure 4.13 compares children whose birth weights are ≤ 3000 grams (n=130) with those whose birth weights are >3000 grams (n=524); the plot on the right compares the children whose birth weights are ≤4000 grams (n=593) with those whose birth weights are > 4000 grams (n=61). In each plot, the grey unfilled circles correspond to the absolute SMDs without adjustment for balancing score stratification. The absolute SMD is defined as the absolute value of the mean difference between the two groups of children divided by the pooled standardized deviation. The covariates are listed on the Y axis, and they are ordered based on their corresponding absolute SMD without adjustment for balancing score stratification. The grey letters correspond to the 5 strata. The black filled circles correspond to the absolute SMDs with adjustment for balancing score stratification. A
SMD was first calculated for each covariate within each stratum, the absolute SMD with adjustment for balancing score stratification was achieved by averaging SMDs across strata and then taking the absolute value. The dashed vertical line on the left corresponds to zero. The absolute values of the SMDs for these covariates are nearer to zero following adjustment for balancing score stratification than the absolute SMDs without adjustment for balancing score stratification, particularly when comparing those whose birth weights are \( \leq 3000 \) grams with those whose birth weights are \( > 3000 \) grams. The two groups are of comparable size. The graphic shows that what is termed “the balancing score” actually does help balance all observed covariate distributions as theory indicates it should. This approach to assessing balance graphically in the case of generalized propensity score methods appears not to have been used before.

![Figure 4.13](cbal.psa) plots checking covariate balance for the 654 non-black-non-Hispanic children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks.
Figure 4.14 shows the fitted locally linear regression surface relating cognitive composite score to birth weight and balancing score for the non-black-non-Hispanic children after limiting their balancing scores to the region of overlap from 0.6 to 0.95. The span was set as 0.85. The surface shows that for those non-black-non-Hispanic children whose gestational ages are normal and whose balancing scores are between 0.6 and 0.95, the association of cognitive composite score with birth weight is upside-down U-shaped at any fixed level of the balancing score, their cognitive composite scores increase gradually with birth weights until an optimum weight around 3600 grams, after that their cognitive composite scores begin to fall.

Figure 4.14 Fitted surface for locally linear regression (span=0.85) of cognitive composite score on birth weight and balancing score for the 557 non-black-non-Hispanic children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks and whose balancing scores are in the overlapping region from 0.6 to 0.95.

Cognitive composite scores were then compared across the three groups of children within each stratum. Table 4.4 is a 5 by 3 table. The 557 non-black-non-Hispanic children whose gestational ages are normal and whose balancing scores are in the overlapping region from 0.6 to 0.95 was divided into 15 cells according to the
quintiles of their balancing scores and the birth weight category they belong to. Besides the number of children in each cell and their mean cognitive score, table 4.4 also displays the estimated overall mean score for children within each birth weight category after stratifying on the balancing score to adjust for covariate differences and its corresponding standard error.

Table 4.4

Means of Cognitive Composite Score, Counts of Children, Estimated Means after Adjusting for the Quintiles of the Balancing Score and Their 95% Confidence Intervals for the 557 Non-Black-Non-Hispanic Children in the NLSY79-C Whose Gestational Ages Are in the Normal Range from 37 Weeks to 42 Weeks and Whose Balancing Scores Are in the Overlapping Region from 0.6 to 0.95

<table>
<thead>
<tr>
<th>strata</th>
<th>balancing score range</th>
<th>birth weight&lt;=3000</th>
<th>3000&lt;birth weight&lt;=4000</th>
<th>birth weight&gt;4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>adjusted means</td>
<td>0.15</td>
<td>0.26</td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>[-0.1, 0.4]</td>
<td>[0.16, 0.36]</td>
<td>[-0.47, 0.25]</td>
<td></td>
</tr>
<tr>
<td>(0.6, 0.692)</td>
<td>0.19(29)</td>
<td>0.32(44)</td>
<td>-0.09(3)</td>
<td></td>
</tr>
<tr>
<td>(0.692, 0.788)</td>
<td>0.04(32)</td>
<td>0.24(92)</td>
<td>-0.63(7)</td>
<td></td>
</tr>
<tr>
<td>(0.788, 0.86)</td>
<td>0.38(24)</td>
<td>0.18(95)</td>
<td>0.26(11)</td>
<td></td>
</tr>
<tr>
<td>(0.86, 0.917)</td>
<td>0.36(14)</td>
<td>0.24(103)</td>
<td>-0.22(14)</td>
<td></td>
</tr>
<tr>
<td>(0.917, 0.95)</td>
<td>-0.2(5)</td>
<td>0.33(72)</td>
<td>0.12(12)</td>
<td></td>
</tr>
</tbody>
</table>

As displayed in Table 4.4, the children whose birth weights are > 4000 grams demonstrate an obvious disadvantage comparing with the children whose birth weights are <= 4000 grams, their mean score is notably lower than those of the other two groups in the first four strata. In stratum 1, 2, and 5, the children whose birth weights are <= 3000 grams have lower mean score than the children whose birth weights are between 3000 grams and 4000 grams, but in stratum 3 and 4, the children whose birth weights are <= 3000 grams have higher mean score. There are only 3 children whose birth weights
are > 4000 grams in stratum 1, while in stratum 5, there are only 4 children whose birth weights are <=3000 grams, any differences we see within these two strata can therefore be biased. The estimated mean score after adjusting for the quintiles of the balancing score is 0.15 for those children whose birth weights are <=3000 grams, the estimated mean score for those whose birth weights are between 3000 grams and 4000 grams is 0.26, and the estimated mean score for those whose birth weights are >4000 grams is -0.11.

Results for the blacks in the NLSY79-C whose gestational ages are normal.

Propensity score analysis was then conducted for the blacks whose gestational ages are within the normal range from 37 weeks to 42 weeks. The birth weights of these black children range from 1000 grams to 5000 grams, with one outlier being 5613 grams. The one outlier was excluded from further analysis, resulting in a sample size of 551.

Marital status of mother in the birth year of the child, mother’s highest educational attainment by the birth year of the child, and family net income in the birth year of the child, the three covariates extracted from the NLSY79 main data file have the highest number of missing values as shown in Table 4.5. The missing values in these covariates were imputed using the multiple imputation procedure (Schafer, 1997). Ten imputed data sets were produced and the imputed values were averaged across the ten data sets to reduce imputation uncertainty. Table 4.5 presents the descriptive statistics for these 12 covariates.
Table 4.5

Descriptive Statistics of Covariates for the 551 Black Children in the NLSY79-C Whose Gestational Ages Are in the Normal Range from 37 Weeks to 42 Weeks

<table>
<thead>
<tr>
<th>Description</th>
<th>mean/percentage</th>
<th>sd</th>
<th>median</th>
<th>min</th>
<th>max</th>
<th>skew</th>
<th>kurtosis</th>
<th>n of missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>52% (males)</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.08</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>age by 1986</td>
<td>7</td>
<td>1.99</td>
<td>7</td>
<td>5</td>
<td>13</td>
<td>0.69</td>
<td>-0.41</td>
<td>0</td>
</tr>
<tr>
<td>mothers' age</td>
<td>18</td>
<td>2.06</td>
<td>18</td>
<td>13</td>
<td>23</td>
<td>0.27</td>
<td>-0.5</td>
<td>0</td>
</tr>
<tr>
<td>birth order to mother</td>
<td>71% (first born)</td>
<td>0.45</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.92</td>
<td>-1.14</td>
<td>0</td>
</tr>
<tr>
<td>marital status of mother</td>
<td>22% (married)</td>
<td>0.44</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.13</td>
<td>-0.7</td>
<td>299</td>
</tr>
<tr>
<td>marital status of mother*</td>
<td>49% (married)</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.02</td>
<td>-2.01</td>
<td>NA</td>
</tr>
<tr>
<td>mothers' IQ</td>
<td>16</td>
<td>15.79</td>
<td>12</td>
<td>1</td>
<td>94</td>
<td>1.75</td>
<td>3.67</td>
<td>11</td>
</tr>
<tr>
<td>mothers' IQ*</td>
<td>17</td>
<td>15.89</td>
<td>12</td>
<td>1</td>
<td>94</td>
<td>1.67</td>
<td>3.33</td>
<td>NA</td>
</tr>
<tr>
<td>mothers' education</td>
<td>11</td>
<td>1.55</td>
<td>11</td>
<td>7</td>
<td>15</td>
<td>-0.14</td>
<td>-0.29</td>
<td>368</td>
</tr>
<tr>
<td>mothers' education*</td>
<td>10</td>
<td>1.87</td>
<td>10</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>-0.55</td>
<td>NA</td>
</tr>
<tr>
<td>Income</td>
<td>8506</td>
<td>6369.73</td>
<td>7000</td>
<td>0</td>
<td>35242</td>
<td>1.23</td>
<td>1.75</td>
<td>355</td>
</tr>
<tr>
<td>Income*</td>
<td>9731</td>
<td>4342.65</td>
<td>9716</td>
<td>0</td>
<td>35242</td>
<td>0.84</td>
<td>3.72</td>
<td>NA</td>
</tr>
<tr>
<td>maternal weight before pregnancy</td>
<td>126</td>
<td>20.78</td>
<td>125</td>
<td>85</td>
<td>274</td>
<td>1.53</td>
<td>5.7</td>
<td>19</td>
</tr>
<tr>
<td>maternal weight before pregnancy*</td>
<td>126</td>
<td>20.47</td>
<td>125</td>
<td>85</td>
<td>274</td>
<td>1.55</td>
<td>5.93</td>
<td>NA</td>
</tr>
<tr>
<td>drank alcohol</td>
<td>31% (drank)</td>
<td>0.46</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.81</td>
<td>-1.34</td>
<td>0</td>
</tr>
<tr>
<td>if smoked</td>
<td>30% (smoked)</td>
<td>0.46</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.87</td>
<td>-1.25</td>
<td>0</td>
</tr>
<tr>
<td>month in pregnancy of 1st doctor's visit</td>
<td>3</td>
<td>1.94</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>2.01</td>
<td>4.97</td>
<td>30</td>
</tr>
<tr>
<td>month in pregnancy of 1st doctor's visit*</td>
<td>3</td>
<td>1.89</td>
<td>3</td>
<td>1</td>
<td>11</td>
<td>1.99</td>
<td>5.13</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: * represents the covariates after their missing values were imputed.

As shown in Table 4.5, the average age of these black children when they were interviewed is 7 years old; and the average age of the mothers when their child was born is 18 years old. Most the mothers were not married when their child was born because the distribution of “marital status of mother” is positively skewed. The covariates “drank alcohol” and “if smoked” are positively skewed, indicating there were more mothers who drank alcohol and smoked during pregnancy than those who didn’t. Another major objective of Table 4.5 is to show changes in these statistics from imputation of missing values. Marital status of mother, mother’s education, and income have the highest
number of missing values, there are notable changes in their descriptive statistics after their missing values were imputed.

**BIRTH**, a three-level ordinal categorical variable was created, with those whose birth weights are <= 2800 grams coded as 1, and those whose birth weights are between 2800 grams and 3800 grams coded as 2, and those whose birth weights are > 3800 grams coded as 3. And the correlations of each covariate with **BIRTH** and with cognitive composite score were examined and reported in Table 4.6.

Table 4.6

*Correlations of Covariates with BIRTH and Cognitive Composite Score for the 551 Black Children in the NLSY79-C Whose Gestational Ages Are in the Normal Range from 37 Weeks to 42 Weeks*

<table>
<thead>
<tr>
<th></th>
<th>BIRTH</th>
<th>cognitive composite score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>-0.12</td>
<td>0.07</td>
</tr>
<tr>
<td>age by 1986</td>
<td>-0.10</td>
<td>-0.11</td>
</tr>
<tr>
<td>mothers.age</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>birth.order.to.mother</td>
<td>-0.06</td>
<td>-0.11</td>
</tr>
<tr>
<td>marital.status.of.mother</td>
<td>-0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>mothers.IQ</td>
<td>0.04</td>
<td>0.29</td>
</tr>
<tr>
<td>mothers.education</td>
<td>0.10</td>
<td>0.2</td>
</tr>
<tr>
<td>maternal.weight.before.pregnancy</td>
<td>0.12</td>
<td>-0.02</td>
</tr>
<tr>
<td>drink.alcohol</td>
<td>-0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>if.smoke</td>
<td>-0.03</td>
<td>-0.07</td>
</tr>
<tr>
<td>month.in.pregnancy.of.1st.doctor.visit</td>
<td>-0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td>Income</td>
<td>0.07</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Sex, age of the child by 1986, maternal age in the birth year of the child, mother’s highest educational attainment, and maternal weight before pregnancy are related with **BIRTH**, either positively or negatively, with an absolute magnitude close to 0.1. Black males tend to have heavier birth weight than females. The older the mother was when the child was born, the better educated the mother was. Also the heavier the mother was
before pregnancy, the heavier the birth weight of the child tended to be. But birth order to mother, marital status of mother, mother’s IQ, if the mother drank alcohol or smoked during pregnancy, in which month during pregnancy the mother first visited a medical person for prenatal care, and family net income in the birth year of the child show little relationship with BIRTH.

Age of the child by 1986, birth order to mother, mother’s IQ, mother’s highest educational attainment, if the mother drank alcohol during pregnancy, and family net income in the birth year of the child are notably related, either positively or negatively, with cognitive composite score. But sex, maternal age in the birth year of the child, marital status of mother, maternal weight before pregnancy, if the mother smoked during pregnancy, and in which month during pregnancy the mother first visited a medical person for prenatal care show little relationship with cognitive composite score.

The main effects of all the 12 covariates were first included in the ordinal logistic regression model as predictors to estimate the balancing score. Polynomial terms and interaction terms involving the covariates were tried in the ordinal logit model as an additional term one after another. The additional term would be kept in the model if covariate balance was improved after the sample was stratified based on the quintiles of the balancing score estimated from that model. This iterative process led into a final model which includes the main effects of the 12 covariates and an interaction between birth order to mother and mother’s IQ. \( \hat{b}(x_i) = p(y \geq 2) \) was used as the balancing score to stratify the three groups of children into five strata based on the quintiles of \( \hat{b}(x_i) \).

Figure 4.15 displays the histograms that show how the distributions of the balancing score overlap each other across the three birth weight categories. It is shown
that there is a balancing score region from 0.65 to 0.95 that contains scores for children from all the three birth weight categories, and this region can therefore support comparisons across the three groups.

Figure 4.15 Histograms showing overlap between the balancing score distributions for the 551 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks.
cbal.psa plots were used to assess whether covariate balance was achieved by the so-called balancing score. The plot on the left of Figure 4.16 compares the children whose birth weights are \( \leq 2800 \text{ grams} \) (n=123) with those whose birth weights are \( >2800 \text{ grams} \) (n=428), while the plot on the right compares the children whose birth weights are \( \leq 3800 \text{ grams} \) (n=501) with those whose birth weights are \( >3800 \text{ grams} \) (n=50). The absolute values of the SMDs for most of these covariates with adjustment for balancing score stratification (the black filled circles) are nearer to zero than the absolute SMDs without adjustment for balancing score stratification (the grey unfilled circles), particularly when comparing those whose birth weights are \( \leq 2800 \text{ grams} \) with those whose birth weights are \( >2800 \text{ grams} \), the two samples with a comparable size. The graphic shows that “the balancing score” actually does help balance most observed covariate distributions as theory indicates it should. This approach to assessing balance graphically in the case of generalized propensity score methods appears to be novel.

![Figure 4.16](cbal.psa plots checking covariate balance for the 551 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks.)
Figure 4.17 shows the fitted locally linear regression surface relating cognitive composite score to birth weight and balancing score for the blacks after limiting their balancing score to the region of overlap from 0.65 to 0.95. The span was set as 0.85. The surface shows that for those black children whose gestational ages are normal and whose balancing scores are between 0.65 and 0.95, the association of cognitive composite score with birth weight is upside-down U-shaped at any fixed level of balancing score, their cognitive composite scores increase gradually with birth weights until an optimum weight around 3300 grams, after that their cognitive composite scores begin to fall.

![Figure 4.17](image)

**Figure 4.17** Fitted surface for locally linear regression (span=0.85) of cognitive composite score on birth weight and balancing score for the 493 black children in the NLSY79-C whose gestational ages are in the normal range from 37 weeks to 42 weeks and whose balancing scores are within the overlapping region from 0.65 to 0.95.

Cognitive composite scores were then compared across the three groups of children within each stratum. As displayed in Table 4.7, the children whose birth weights are > 3800 grams demonstrate an obvious disadvantage comparing with the children whose birth weights are <= 3800 grams, their mean scores are notably lower than those of the other two groups in stratum 1, 3, 4, and 5. And the mean scores of the children
whose birth weights are $\leq 2800$ grams are lower than those of the children whose birth weights are between 2800 grams and 3800 grams across the five strata. There are only 2 children whose birth weights are $> 3800$ grams in stratum 1, and only 4 children whose birth weights are $>3800$ grams in stratum 2, any differences we see within these two strata could therefore be biased. The estimated mean score after adjusting for the quintiles of the balancing score is -0.25 for those children whose birth weights are $\leq 2800$ grams, is -0.16 for those whose birth weights are between 2800 grams and 3800 grams, and is -0.55 for those whose birth weights are $>3800$ grams.

Table 4.7

Means of Cognitive Composite Score, Counts of Children, Estimated Means after Adjusting for the Quintiles of the Balancing Score and Their 95% Confidence Intervals for the 493 Black Children in the NLSY79-C Whose Gestational Ages Are in the Normal Range from 37 Weeks to 42 Weeks and Whose Balancing Scores Are in the Overlapping Region from 0.65 to 0.95

<table>
<thead>
<tr>
<th>strata</th>
<th>means(counts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>balancing score</td>
<td>birth weight $\leq 2800$</td>
</tr>
<tr>
<td>range</td>
<td></td>
</tr>
<tr>
<td>(0.65, 0.714)</td>
<td>-0.19(18)</td>
</tr>
<tr>
<td>(0.714, 0.767)</td>
<td>-0.32(28)</td>
</tr>
<tr>
<td>(0.767, 0.806)</td>
<td>-0.33(20)</td>
</tr>
<tr>
<td>(0.806, 0.856)</td>
<td>-0.29(22)</td>
</tr>
<tr>
<td>(0.856, 0.95)</td>
<td>-0.12(12)</td>
</tr>
<tr>
<td>adjusted means</td>
<td>-0.25</td>
</tr>
<tr>
<td>95% CI</td>
<td>[-0.45, -0.05]</td>
</tr>
</tbody>
</table>
Results for the PSID-CDS

Loess Regression to Depict the Relationship of Birth Weight with Cognitive Ability

In 1997, 3586 children from over 2500 families and their parents and teachers were successfully interviewed. Birth weight, gestational age, and most of the potential confounding covariates were collected by means of mothers’ recall.

Birth weight was recorded in pounds and ounces and converted into grams. Gestational age was estimated from how many days before or after the due date when the child was expected to be born, with the due date defined as 40 complete weeks from the first day of the mother’s last menstrual period.

Four subtests, letter-word identification, passage comprehension, calculation, and applied problems, from the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R) Tests of Achievement (Woodcock & Johnson, 1989) and the digit span subtest from the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) were administered to measure cognitive capabilities. Letter-word identification assesses reading decoding by requiring the examinee to orally read a list of single words of increasing difficulty. This subtest is comparable to the PIAT reading recognition subtest. Passage comprehension requires the examinee to orally supply the missing word removed from a sentence or a very brief paragraph, it is comparable to the PIAT reading comprehension subtest. Calculation requires the examinee to perform mathematic operations that vary in difficulty. In applied problems, the examinee analyzes and solves practical mathematic problems. These two subtests are parallel to the PIAT math subtest. Digit span requires the examinee to repeat dictated series of digits (e.g., 4 1 7 9) forwards
and other series backwards. The digit span score was excluded from any analyses because of its relatively low reliability.

These children are from a variety of racial backgrounds. Since sample sizes for minorities, such as Hispanics, American Indians, and Asians, are small, analyses and results were given only for the 1763 white and black children who are singleton births rather than multiple births, whose mother stayed in the United States during pregnancy, who were born in the United States, who had their birth weight and gestational age recorded, and who completed two of the four WJ-R subtests, applied problem and letter-word, in 1997. Figure 4.18 displays the scatterplot matrix for the four WJ-R subtest scores. Only the score from the letter-word identification subtest and the score from the applied problems subtest were included in the analyses.
Adjustment for age differences. The age of these 1763 white and black children varies from 3 to 13. The scatterplots of age against each of the two cognitive scores (Figure 4.19) show the cognitive score increases as age increases, which confirms that the variance in these students’ performance in the two cognitive tests are strongly related to their age differences at the time of testing. In the two scatterplots, the dashed grey line is a loess fit, the black straight line is the linear regression fit, and the dotted black line is the quadratic fit which is closer to the loess fit than linear regression. Therefore, to adjust
for age differences at the time of testing, the predicted score from the quadratic regression was subtracted from each observed score to obtain residuals which were taken to be the adjusted cognitive scores for subsequent analyses.

*Figure 4.19* Scatterplots of age against each cognitive score with loess, linear, and quadratic fits superimposed on the data for the 1763 white and black children in the PSID-CDS.

*Generation of cognitive composite score.* As can be seen in *Figure 4.18*, the letter word score is positively correlated with applied problem score and the magnitude is reasonably strong. The positive correlation was evident even after adjusting for age differences at the time of testing. The residuals of the two cognitive scores, after subtracting the variance explained by age (the quadratic fit), were standardized and added together to generate a cognitive composite score. The resultant composite score was then converted to z-scores.

Table 4.8 presents the descriptive statistics for birth weight, gestational age in days, and cognitive composite score for the white (n=931) and black (n=832) children.
The mean birth weight of these white children is 3452 grams, and the minimum birth weight is 680 grams while the maximum is 6038 grams. The mean birth weight of these black children is 3160 grams, about 300 grams lower than that of the whites. The minimum birth weight of these black children is 496 grams while the maximum is 6917 grams. The gestational ages of these white children have a range between 182 days and 318 days and a mean of 278 days, while the gestational ages of these black children have a range between 168 days and 336 days and a mean of 276 days, slightly lower. The cognitive composite scores of these white children have a range between -3.67 and 3.46 and a mean of 0.33. The mean cognitive composite score of these black children is much lower, being -0.37.

Table 4.8

Descriptive Statistics for Birth Weight, Gestational Age in Days, and Cognitive Composite Score for the white (n=931) and the black (n=832) children in the PSID-CDS

<table>
<thead>
<tr>
<th></th>
<th>whites</th>
<th></th>
<th>blacks</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>birth weight</td>
<td>gestational age</td>
<td>cognitive composite score</td>
<td>birth weight</td>
</tr>
<tr>
<td>mean</td>
<td>3452</td>
<td>278</td>
<td>0.33</td>
<td>3160</td>
</tr>
<tr>
<td>sd</td>
<td>594</td>
<td>14.45</td>
<td>0.96</td>
<td>663</td>
</tr>
<tr>
<td>median</td>
<td>3430</td>
<td>280</td>
<td>0.38</td>
<td>3203</td>
</tr>
<tr>
<td>minimum</td>
<td>680</td>
<td>182</td>
<td>-3.67</td>
<td>496</td>
</tr>
<tr>
<td>maximum</td>
<td>6038</td>
<td>318</td>
<td>3.46</td>
<td>6917</td>
</tr>
<tr>
<td>skewness</td>
<td>-0.34</td>
<td>-1.56</td>
<td>-0.26</td>
<td>-0.5</td>
</tr>
<tr>
<td>kurtosis</td>
<td>1.41</td>
<td>5.56</td>
<td>0.39</td>
<td>2.85</td>
</tr>
</tbody>
</table>

Figure 4.20 are the scatterplot matrices for cognitive composite score, gestational age in days, birth weight, and sex for the two subsamples defined by race in the PSID-CDS.
CDS. For the whites, their cognitive composite scores are mainly distributed between -2 and 2.5, their gestational ages are intensively distributed between 250 days and 300 days, and their birth weights are mainly distributed between 1500 grams and 5000 grams. Birth weight is positively correlated with gestational age as expected ($r = 0.58$). And it is slightly positively correlated with cognitive composite score ($r = 0.09$). Cognitive composite scores increase as birth weight increases. There is also slightly a positive association between gestational age and cognitive composite score ($r = 0.03$). White males (coded as 1) are heavier in birth weight in comparison with white females (coded as 2). But there are few differences between males and females in their gestational ages and cognitive composite scores. For the blacks, cognitive composite scores are mainly distributed between -3 and 2, their gestational ages are intensively distributed between 260 days and 300 days; their birth weights are mainly distributed between 1000 grams and 5000 grams. Their birth weights are positively correlated with their gestational ages ($r = 0.6$), and slightly correlated with their cognitive composite scores ($r = 0.04$). There is also a small positive association between gestational age and cognitive composite score ($r = 0.04$). Black males (coded as 1) have higher birth weights than black females (coded as 2), but there are few differences between males and females in their gestational ages and cognitive composite scores.
Figure 4.20 Scatterplot matrices for cognitive composite score, gestational age in days, birth weight, and sex for the white (n=931) and the black (n=832) children in the PSID-CDS.

Locally linear regression of cognitive composite score on birth weight and gestational age. To answer the first research question, “What is the nature of the predictive association of birth weight and gestational age for cognitive ability in childhood?” for each subsample defined by race and sex, a locally linear regression was conducted with both birth weight and gestational age being the predictors and with cognitive composite score being the response. (In the loess function, the degree was set to 1 to specify locally linear regression, and the span was set as 0.85.)

Figure 4.21 shows the fitted surface for loess regression of cognitive composite score on birth weight and gestational age for the 931 whites. The graphic on the left gives a clear view of the surface using birth weight as the key predictor, and the one on the right gives uses gestational age as the key predictor. The data are mainly distributed within the interval from 1500 grams to 5000 grams along the distribution of birth weight with a number of outliers on both ends and within the short interval from 250 days to 300
days along the distribution of gestational age with a number of outliers particularly on the left side. The surface shows a positive association between cognitive composite score and birth weight at any fixed level of gestational age and no evident association between cognitive composite score and gestational age at any fixed level of birth weight.

Figure 4.21 Fitted surface for locally linear regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 931 white children in the PSID-CDS.

The white sample was further divided into males and females. The two graphics above in Figure 4.22 demonstrate the fitted surface for the loess regression of cognitive composite score on birth weight and gestational age for the white males. The surface shows a positive association between cognitive composite score and birth weight at any fixed level of gestational age and no evident association between cognitive composite score and gestational age in the area where most data are distributed. The rise of the fitted surface on the left side is caused by the outliers on the left end.

The two graphics below in Figure 4.22 demonstrate the fitted surface for the white females. The surface shows a positive association between cognitive composite score and birth weight at any fixed level of gestational age and a positive association between
cognitive composite score and gestational age at any fixed level of birth weight. These two views show the same relationship from two different perspectives.

Figure 4.22 Fitted surfaces for locally linear regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 459 white males (the two graphics above) and for the 472 white females (the two graphics below) in the PSID-CDS.

Figure 4.23 shows the fitted surface for loess regression of cognitive composite score on birth weight and gestational age for the blacks. The graphic on the left gives a clear view of the surface using birth weight as the key predictor, and the one on the right
uses gestational age as the key predictor. The data are mainly distributed within the interval from 1000 grams to 5000 grams along the distribution of birth weight with a number of outliers on both ends and within the short interval from 260 days to 300 days along the distribution of gestational age with a number of outliers particularly on the left end. The surface shows that there is almost no association between cognitive composite score and birth weight or between cognitive composite score and gestational age in the area where most data are distributed. The decline of the surface at the right corner is the result one outlier in this region.

![Surface Plot](image)

**Figure 4.23** Fitted surface for locally linear regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 832 blacks in the PSID-CDS.

The black sample was further divided into males and females. The two graphs above in Figure 4.24 demonstrate the fitted surface for the loess regression of cognitive composite score on birth weight and gestational age for the black males. The surface shows that the association of cognitive composite score with birth weight is upside-down U-shaped. And it also shows that there is hardly any association between cognitive composite score and gestational age at any fixed level of birth weight.
The two graphs below in Figure 4.24 demonstrate the fitted surface for black females. The surface shows a positive association between cognitive composite score and birth weight in the area where most data are distributed and no consistent association between cognitive composite score and gestational age at any fixed level of birth weight.

*Figure 4.24* Fitted surfaces for locally linear regression (span=0.85) of cognitive composite score on birth weight and gestational age for the 443 black males (the two graphics above) and for the 389 black females (the two graphics below) in the PSID-CDS.
Simple locally linear regression of cognitive composite score on birth weight. To answer the second research question, “What is the nature of the predictive association of birth weight for cognitive ability in childhood when gestational age is restricted within the normal range, i.e., between 259 days and 294 days?” those children whose gestational ages are below or beyond the normal range were excluded from the sample. The birth weights of these white children whose gestational ages are normal range from 2000 grams to 5000 grams, with one outlier being 6038 grams, while the birth weights of these black children whose gestational ages are normal range from 1000 grams to 5000 grams.

The top graph in Figure 4.25 shows a two-dimensional scatterplot of birth weight against cognitive composite score in which whites and blacks are the two separate groups. The data for whites are distributed slightly to the upper right of those of the blacks.

Locally linear simple regression curves were overlaid separately for the two racial groups with the degree set to 1 and the span set as 0.85. The two curves are magnified in the second graph in Figure 4.25. The loess curve for the whites (the black line) shows that for those white children whose gestational ages are normal, their cognitive composite scores increase gradually with birth weights until an optimum weight around 4337 grams, after that their composite scores begin to fall. The mean cognitive score of those birth weights below 3200 grams is 0.32 (n = 225, 95%CI = [0.2, 0.44]), and the mean cognitive score increases to 0.37 for birth weights between 3200 grams and 4200 grams (n = 539, 95CI = [0.29, 0.45]), and then decreases to 0.29 for birth weights beyond 4200 grams (n= 70, 95CI = [0.08, 0.5]).
The loess curve for the blacks (the grey line) shows that for those black children whose gestational ages are normal, their cognitive composite scores also slightly increase with birth weights until an optimum weight around 3400 grams, after that there is no evident rise or decline. The mean cognitive score of those birth weights below 3000 grams is -0.4 (n = 240, 95%CI = [-0.5, -0.3]), and the mean cognitive score increases to -0.34 for birth weights between 3000 grams and 4000 grams (n = 446, 95CI = [-0.42, -0.26]), and then increases again slightly to -0.3 for birth weights beyond 4000 grams (n= 55, 95CI = [ -0.57, -0.03]).

The upper horizontal dashed line shows the mean cognitive composite score for these white children, while the lower horizontal dashed line shows the mean cognitive composite score for these black children. The lines are separated from each other, indicating notable racial differences on average. The SMD between the mean of the whites and that of the blacks is 0.7.
Figure 4.25 2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by race for the 1575 white and black children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days.

The top graph in Figure 4.26 depicts a two-dimensional scatterplot that shows how the average cognitive composite scores change across the distribution of birth weight.
for the whites whose gestational ages are within the normal range from 259 days to 294 days. The mean difference, standardized “effect size”, is 0.06. Locally linear regression curves were overlaid separately for the males (the black line) and for the females (the grey line). The two curves are magnified in the second graph in Figure 4.26. The loess curve for the males shows that for those white males whose gestational ages are normal, their cognitive composite scores increase slightly with increases in their birth weights until an optimum weight around 3900 grams, after that their composite scores begin to fall. The loess curve for the females shows that for the white females whose gestational ages are normal, their cognitive composite scores generally improve slightly with increases in their birth weights. The horizontal dashed line shows the mean cognitive composite score for these white children. The two loess curves are sufficiently close to the horizontal dashed line to indicate that gender differences are minor.
Figure 4.26 2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by gender for the 834 white children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days.

The first graph in Figure 4.27 is a two-dimensional scatterplot generated for the blacks whose gestational ages are within the normal range from 259 days to 294 days.
with locally linear regression curves overlaid separately for the black males (the black line) and for the black females (the grey line). The two curves are magnified in the second graph in Figure 4.27. The loess curve for the black males whose gestational ages are normal is almost flat in the area where most data are distributed. The loess curve for the females shows that for those black females whose gestational ages are normal, their cognitive composite scores generally increase as their birth weights increase particularly in the area where most data are distributed. The horizontal dashed line shows the mean cognitive composite score for these black children. The two loess curves are sufficiently close to the horizontal dashed line to indicate gender differences are minor. The mean difference, the SMD, is -0.04.
Figure 4.27 2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by gender for the 741 black children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days.
Generalized propensity score methods to control for confounding

To answer the third research question, whether observed associations tend to increase or decrease after confounding of numerous covariates that account for prenatal differences is controlled, generalized propensity score methods were used to accomplish the goal of reducing the effects of multiple confounders.

Because the results from the locally linear regression of cognitive composite score on birth weight when gestational age was restricted to be within the normal range from 259 days to 294 days show that there are notable racial differences, but gender differences within each racial group are trivial, propensity score analysis was conducted only for the white sample and for the black sample which include both the males and the females whose gestational age is normal.

The results from the locally linear regression of cognitive composite score on birth weight show that the association between birth weight and cognitive composite score among the 834 white children whose gestational ages are normal has an upside-down U shape. Their cognitive composite scores increase gradually with birth weights until an optimum weight around 4337 grams, after that their composite scores begin to fall. The results for the 741 black children whose gestational ages are normal show that their cognitive composite scores also slightly increase with birth weights until an optimum weight around 3400 grams, after that there is no evident rise or decline. To prove if there is an optimum range in the middle, the distribution of birth weight of the whites was divided into three categories by 3200 grams and by 4200 grams as shown in Figure 4.28. Similarly, the distribution of birth weight of the blacks was divided into three categories by 3000 grams and by 4000 grams. The cutting points are 200 grams
higher than those used in the NLSY79-C. This is because birth weight increased with time, the birth weights of the children in the PSID-CDS are averagely 150 grams to 200 grams heavier than those of the children in the NLSY79-C which was conducted about 10 years earlier.

**Figure 4.28** Plot to demonstrate division of the birth weight distribution of the 833 white children and that of the 741 black children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days.

Covariates adjusted for included age of the child by 1997, maternal age in the birth year of the child, birth order to mother, marital status of mother in the birth year of the child, mother’s highest educational attainment by the birth year of the child, father’s age in the birth year of the child, birth order to father, and some covariates that describe social economic status of the mother during pregnancy such as whether some of the medical expenses related to the pregnancy or the delivery of the child were covered by private health insurance, whether Medicaid paid for any of these medical bills, whether
the mother was in the WIC program when she was pregnant with the child, whether the mother received any government food stamps while she was pregnant, and whether the mother received any payments from ADC or AFDC while she was pregnant.

The birth order of the white sample, either to mother or to father, has possible values from 1 to 5, while the birth order of the black sample has possible values from 1 to 10. Since there were very few children not born as the first child in the family, this covariate was recoded as a binary variable with these children born as the first child in the family coded as 1, and those not coded as 2. The covariate “whether some of the medical expenses related to the pregnancy or the delivery of the child were covered by private health insurance” was originally a three-level ordinal categorical variable and was recoded as a binary variable with those having private insurance coded as “1” and those having no private insurance coded as “0”. The PSID-CDS also collected information on whether the mother received assistance from any other public agencies and whether the mother got free food from any other government programs while she was pregnant. These two covariates were excluded because very few mothers did so during their pregnancy.

*Results for the whites in the PSID-CDS whose gestational ages are normal.*

Propensity score analysis was first conducted for the whites whose gestational ages are within the normal range from 259 days to 294 days. The birth weights of these white children range from 2000 grams to 5000 grams, with one outlier being 6038 grams. The one outlier was excluded from further analysis, resulting in a sample size of 833.

The missing values in these covariates were imputed using the multiple imputation procedure (Schafer, 1997). Ten imputed data sets were produced and the imputed values were averaged across the ten data sets to reduce imputation uncertainty. Table 4.9 presents the descriptive statistics for these 13 covariates.
Table 4.9

Descriptive Statistics of Covariates for the 833 white Children in the PSID-CDS Whose Gestational Age Is in the Normal Range from 259 Days to 294 Days

<table>
<thead>
<tr>
<th>Covariate</th>
<th>mean/percentage</th>
<th>sd</th>
<th>median</th>
<th>min</th>
<th>max</th>
<th>skew</th>
<th>kurtosis</th>
<th>n of missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>49%(males)</td>
<td>0.5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.02</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>ageby1997</td>
<td></td>
<td></td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>13</td>
<td>0.07</td>
<td>-1.21</td>
</tr>
<tr>
<td>mothers age</td>
<td>27</td>
<td>5.24</td>
<td>28</td>
<td>15</td>
<td>42</td>
<td>0.09</td>
<td>-0.24</td>
<td>3</td>
</tr>
<tr>
<td>mothers age*</td>
<td>27</td>
<td>5.23</td>
<td>28</td>
<td>15</td>
<td>42</td>
<td>0.1</td>
<td>-0.24</td>
<td>NA</td>
</tr>
<tr>
<td>birth order to mother</td>
<td>41%(first born)</td>
<td>0.49</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.38</td>
<td>-1.86</td>
<td>6</td>
</tr>
<tr>
<td>birth order to mother*</td>
<td></td>
<td></td>
<td>41%(first born)</td>
<td>0.49</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.38</td>
</tr>
<tr>
<td>marital status of mother</td>
<td>91%(married)</td>
<td>0.29</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-2.87</td>
<td>6.34</td>
<td>17</td>
</tr>
<tr>
<td>marital status of mother*</td>
<td></td>
<td></td>
<td>91%(married)</td>
<td>0.29</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-2.86</td>
</tr>
<tr>
<td>mothers education</td>
<td>14</td>
<td>2.01</td>
<td>13</td>
<td>8</td>
<td>17</td>
<td>0.05</td>
<td>-0.92</td>
<td>119</td>
</tr>
<tr>
<td>mothers education*</td>
<td></td>
<td></td>
<td>14</td>
<td>1.95</td>
<td>13</td>
<td>8</td>
<td>17</td>
<td>0.15</td>
</tr>
<tr>
<td>fathers age</td>
<td>30</td>
<td>5.75</td>
<td>29</td>
<td>15</td>
<td>57</td>
<td>0.47</td>
<td>0.81</td>
<td>41</td>
</tr>
<tr>
<td>fathers age*</td>
<td></td>
<td></td>
<td>29</td>
<td>5.78</td>
<td>29</td>
<td>15</td>
<td>57</td>
<td>0.48</td>
</tr>
<tr>
<td>birth order to father</td>
<td>60%(first born)</td>
<td>0.49</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.39</td>
<td>-1.85</td>
<td>45</td>
</tr>
<tr>
<td>birth order to father*</td>
<td></td>
<td></td>
<td>41%(first born)</td>
<td>0.49</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.38</td>
</tr>
<tr>
<td>private insurance</td>
<td>50%(yes)</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>private insurance*</td>
<td></td>
<td></td>
<td>50%(yes)</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>MEDICAID</td>
<td>13%(yes)</td>
<td>0.33</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2.26</td>
<td>3.17</td>
<td>2</td>
</tr>
<tr>
<td>MEDICAID*</td>
<td></td>
<td></td>
<td>12%(yes)</td>
<td>0.33</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2.27</td>
</tr>
<tr>
<td>WIC</td>
<td>16%(yes)</td>
<td>0.37</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.86</td>
<td>1.48</td>
<td>0</td>
</tr>
<tr>
<td>food stamp</td>
<td>7%(yes)</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3.41</td>
<td>9.75</td>
<td>1</td>
</tr>
<tr>
<td>food stamp*</td>
<td></td>
<td></td>
<td>7%(yes)</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3.38</td>
</tr>
<tr>
<td>ADC.AFDC</td>
<td>4%(yes)</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4.63</td>
<td>19.67</td>
<td>1</td>
</tr>
<tr>
<td>ADC.AFDC*</td>
<td></td>
<td></td>
<td>4%(yes)</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4.56</td>
</tr>
</tbody>
</table>

Note: * represents the covariates after their missing values were imputed.

As shown in Table 4.9, the average age of these white children when they were interviewed is 7 years old; the average age of the mothers when their child was born is 27 years old, and the average age of the fathers when their child was born is 29 years old.

Most the mothers were married when their child was born because the distribution of “marital status of mother” is negatively skewed. Half of the mothers had private insurance to cover expenses related to the pregnancy or the delivery of the child, and half of them didn’t. Most of the mothers didn’t rely on financial aid provided by government...
or public agencies, such as Medicaid, WIC, food stamp, ADC or AFDC, because those covariates are positively skewed. Mother’s highest educational attainment by the birth year of the child, the covariate extracted from the PSID main data file, had the highest number of missing data values; and there are notable changes in its descriptive statistics after its missing data were imputed.

BIRTH, a three-level ordinal categorical variable was created, with those whose birth weights are <= 3200 grams coded as 1, and those whose birth weights are between 3200 grams and 4200 grams coded as 2, and those whose birth weights are > 4200 grams coded as 3. And the correlations of each covariate with BIRTH and with cognitive composite score were examined and reported in Table 4.10.

Table 4.10

Correlations of Covariates with BIRTH and Cognitive Composite Score for the 833 white Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days.

<table>
<thead>
<tr>
<th></th>
<th>BIRTH</th>
<th>cognitive composite score</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>-0.08</td>
<td>-0.03</td>
</tr>
<tr>
<td>ageby1997</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>mothers age</td>
<td>-0.03</td>
<td>0.23</td>
</tr>
<tr>
<td>birth order to mother</td>
<td>0.08</td>
<td>-0.11</td>
</tr>
<tr>
<td>marital status of mother</td>
<td>0.04</td>
<td>0.16</td>
</tr>
<tr>
<td>mothers education</td>
<td>0.02</td>
<td>0.32</td>
</tr>
<tr>
<td>fathers age</td>
<td>-0.03</td>
<td>0.17</td>
</tr>
<tr>
<td>birth order to father</td>
<td>0.04</td>
<td>-0.09</td>
</tr>
<tr>
<td>private insurance</td>
<td>-0.03</td>
<td>-0.07</td>
</tr>
<tr>
<td>MEDICAID</td>
<td>0</td>
<td>-0.24</td>
</tr>
<tr>
<td>WIC</td>
<td>-0.08</td>
<td>-0.23</td>
</tr>
<tr>
<td>food stamp</td>
<td>-0.03</td>
<td>-0.2</td>
</tr>
<tr>
<td>ADC.AFDC</td>
<td>-0.05</td>
<td>-0.15</td>
</tr>
</tbody>
</table>
Sex, birth order to mother, and one covariate that describes socioeconomic status of the mother during pregnancy, WIC, are related with BIRTH, either positively or negatively, with an absolute magnitude being 0.08. But the rest 10 covariates show little or no relationship with BIRTH.

Maternal age in the birth year of the child, birth order to mother, marital status of mother in the birth year of the child, mother’s highest educational attainment by the birth year of the child, father’s age in the birth year of the child, and four covariates that describe socioeconomic status of the mother during pregnancy, MEDICAID, WIC, food stamp, and ADC.AFDC, are related with cognitive composite score, either positively or negatively, with an absolute magnitude larger than 0.1. But sex, age of the child by 1997, birth order to father, and whether some of the medical expenses related to the pregnancy or the delivery of the child were covered by private health insurance show little relationship with cognitive composite score.

The main effects of all the 13 covariates were first included in the ordinal logistic regression model as predictors to estimate the balancing score. Polynomial terms and interaction terms involving the covariates were tried in the ordinal logit model one after another, and this iterative process led into a final model which includes the main effects of the 13 covariates, a quadratic term of age by 1997, and an interaction between sex and age by 1997. 

\[ \hat{b}(x_i) = p(y \geq 2) = \frac{1}{1 + e^{-(x_{i0} + \beta^T x_i)}} \]

was used as the balancing score to stratify the three groups of children into five strata based on the quintiles of \( \hat{b}(x_i) \).

Figure 4.29 displays the histograms that show how the distributions of the balancing score overlap each other across the three birth weight categories. It is shown that there is a balancing score region from 0.55 to 0.9 that contains children from all the
three birth weight categories, and this region can therefore support comparisons across the three groups.

*Figure 4.29* Histograms showing overlap between the balancing score distributions for the 833 white children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days.

*cbal.psa* plots to check covariate balance are presented in Figure 4.30. The plot on the left in Figure 4.30 compares the children whose birth weights are <= 3200 grams.
(n=225) with those whose birth weights are >3200 grams (n=608), while the plot on the right compares the children whose birth weights are <=4200 grams (n=764) with those whose birth weights are > 4200 grams (n=69). When comparing children whose birth weights are <= 3200 grams with those whose birth weights are > 3200 grams, the absolute values of the SMDs for most covariates following adjustments using balancing score stratification (the black filled circles) are much closer to 0 than their counterparts for the unadjusted case. But when comparing those whose birth weights are <= 4200 grams with those above 4200 grams, the absolute SMDs for most covariates after adjustment for balancing score stratification are larger in magnitude than for the unadjusted case. However, none of these SMDs exceed 0.4, and all but one are below 0.3.

Figure 4.30 cbal.psa plots checking covariate balance for the 833 white children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days.

Figure 4.31 shows the fitted locally linear regression surface relating cognitive composite score to birth weight and balancing score for the whites after limiting their balancing scores to the region of overlap from 0.55 to 0.9. The span was set as 0.85. The
surface shows that for those white children whose gestational ages are normal and whose balancing scores are between 0.55 and 0.7, their cognitive composite scores increase with increases in birth weights. But for those whose balancing scores are beyond 0.7, the association of cognitive composite score with birth weight is upside-down U-shaped, their cognitive composite scores increase gradually with birth weights until an optimum weight around 4000 grams, after that their cognitive composite scores begin to fall.

*Figure 4.31* Fitted surface for local linear regression (span=0.85) of cognitive composite score on birth weight and balancing score for the 804 white children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days and whose balancing scores are in the overlapping region from 0.55 to 0.9.

Cognitive composite scores were then compared across the three groups of children within each stratum. As displayed in Table 4.11, the children whose birth weights are > 4200 grams have higher mean scores than the children whose birth weights are <= 4200 grams in stratum 1, 2, and 4. And the children whose birth weights are between 3200 grams and 4200 grams have higher mean scores than the children whose birth weights are <=3200 in stratum 1, 3, and 5. The estimated mean score after adjusting for the quintiles of the balancing score is 0.32 for those children whose birth weights are
<=3200 grams, is 0.37 for those whose birth weights are between 3200 grams and 4200 grams, and is 0.39 for those whose birth weights are >4200 grams.

Table 4.11

Means of Cognitive Composite Score, Counts of Children, Estimated Means after Adjusting for the Quintiles of the Balancing Score and Their 95% Confidence Intervals for the 804 White Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days and Whose Balancing Scores Are in the Overlapping Region from 0.55 to 0.9

<table>
<thead>
<tr>
<th>strata balancing score range</th>
<th>means(counts)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>birth weight&lt;=3200</td>
<td>3200&lt;birth weight&lt;=4200</td>
</tr>
<tr>
<td>(0.55, 0.664)</td>
<td>0.19(55)</td>
<td>0.26(76)</td>
</tr>
<tr>
<td>(0.664, 0.718)</td>
<td>0.56(59)</td>
<td>0.39(99)</td>
</tr>
<tr>
<td>(0.718, 0.761)</td>
<td>0.18(39)</td>
<td>0.39(116)</td>
</tr>
<tr>
<td>(0.761, 0.798)</td>
<td>0.33(38)</td>
<td>0.35(111)</td>
</tr>
<tr>
<td>(0.798, 0.9)</td>
<td>0.36(25)</td>
<td>0.44(119)</td>
</tr>
<tr>
<td>adjusted means</td>
<td>0.32</td>
<td>0.37</td>
</tr>
<tr>
<td>95% CI</td>
<td>[0.19, 0.45]</td>
<td>[0.28, 0.46]</td>
</tr>
</tbody>
</table>

Results for the blacks in the PSID-CDS whose gestational ages are normal.

Propensity score analysis was then conducted for the 741 blacks whose gestational ages are within the normal range from 259 days to 294 days.

The missing values in these covariates were imputed, again using the multiple imputation procedure (Schafer, 1997). Ten imputed data sets were produced and the imputed values were averaged across the ten data sets to reduce imputation uncertainty. Table 4.12 presents the descriptive statistics for these 13 covariates.
Table 4.12

*Descriptive Statistics of Covariates for the 741 Black Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days*

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Mean/Percentage</th>
<th>SD</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>N of Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>53% (males)</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.11</td>
<td>-1.99</td>
<td>0</td>
</tr>
<tr>
<td>Age by 1997</td>
<td>7</td>
<td>2.88</td>
<td>7</td>
<td>3</td>
<td>13</td>
<td>0.07</td>
<td>-1.22</td>
<td>0</td>
</tr>
<tr>
<td>Mothers age</td>
<td>25</td>
<td>5.46</td>
<td>24.5</td>
<td>12</td>
<td>41</td>
<td>0.19</td>
<td>-0.56</td>
<td>21</td>
</tr>
<tr>
<td>Mothers age*</td>
<td>25</td>
<td>5.44</td>
<td>25</td>
<td>12</td>
<td>41</td>
<td>0.19</td>
<td>-0.56</td>
<td>NA</td>
</tr>
<tr>
<td>Birth order to mother</td>
<td>39% (first born)</td>
<td>0.49</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.46</td>
<td>-1.79</td>
<td>28</td>
</tr>
<tr>
<td>Birth order to mother*</td>
<td>39% (first born)</td>
<td>0.49</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.46</td>
<td>-1.79</td>
<td>NA</td>
</tr>
<tr>
<td>Marital status of mother</td>
<td>39% (married)</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.44</td>
<td>-1.81</td>
<td>34</td>
</tr>
<tr>
<td>Marital status of mother*</td>
<td>40% (married)</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.4</td>
<td>-1.84</td>
<td>NA</td>
</tr>
<tr>
<td>Mothers education</td>
<td>12</td>
<td>1.81</td>
<td>12</td>
<td>3</td>
<td>17</td>
<td>0.06</td>
<td>1.45</td>
<td>185</td>
</tr>
<tr>
<td>Mothers education*</td>
<td>12</td>
<td>1.69</td>
<td>12</td>
<td>3</td>
<td>17</td>
<td>0.17</td>
<td>1.62</td>
<td>NA</td>
</tr>
<tr>
<td>Fathers age</td>
<td>28</td>
<td>6.72</td>
<td>27</td>
<td>15</td>
<td>57</td>
<td>0.79</td>
<td>1.19</td>
<td>310</td>
</tr>
<tr>
<td>Fathers age*</td>
<td>27</td>
<td>6.25</td>
<td>26</td>
<td>15</td>
<td>57</td>
<td>0.8</td>
<td>1.2</td>
<td>NA</td>
</tr>
<tr>
<td>Birth order to father</td>
<td>34% (first born)</td>
<td>0.47</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.67</td>
<td>-1.56</td>
<td>325</td>
</tr>
<tr>
<td>Birth order to father*</td>
<td>35% (first born)</td>
<td>0.48</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-0.61</td>
<td>-1.64</td>
<td>NA</td>
</tr>
<tr>
<td>Private insurance</td>
<td>53% (yes)</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-0.12</td>
<td>-1.99</td>
<td>5</td>
</tr>
<tr>
<td>Private insurance*</td>
<td>53% (yes)</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-0.13</td>
<td>-1.99</td>
<td>NA</td>
</tr>
<tr>
<td>MEDICAID</td>
<td>53% (yes)</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-0.13</td>
<td>-1.99</td>
<td>4</td>
</tr>
<tr>
<td>MEDICAID*</td>
<td>53% (yes)</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-0.13</td>
<td>-1.99</td>
<td>NA</td>
</tr>
<tr>
<td>WIC</td>
<td>66% (yes)</td>
<td>0.47</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-0.7</td>
<td>-1.51</td>
<td>1</td>
</tr>
<tr>
<td>WIC*</td>
<td>67% (yes)</td>
<td>0.47</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-0.7</td>
<td>-1.51</td>
<td>NA</td>
</tr>
<tr>
<td>Food stamp</td>
<td>40% (yes)</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.42</td>
<td>-1.83</td>
<td>4</td>
</tr>
<tr>
<td>Food stamp*</td>
<td>40% (yes)</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.42</td>
<td>-1.83</td>
<td>NA</td>
</tr>
<tr>
<td>ADC.AFDC</td>
<td>31% (yes)</td>
<td>0.46</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.84</td>
<td>-1.3</td>
<td>4</td>
</tr>
<tr>
<td>ADC.AFDC*</td>
<td>31% (yes)</td>
<td>0.46</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.83</td>
<td>-1.31</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: * represents the covariates after their missing values were imputed.

As shown in Table 4.12, the average age of these black children when they were interviewed was 7 years old; the average age of the mothers when their child was born was 25 years old, and the average age of the fathers when their child was born was 27 years old. Most of these PSID mothers were not married when their child was born as can be seen from the positive skew of the distribution of “marital status of mother.” More mothers had private insurance to cover expenses related to the pregnancy or the delivery
of the child than those who didn’t. More mothers received payments from Medicaid or joined WIC when they were pregnant than those who didn’t, but less mothers received government food stamps or received payments from ADC or AFDC. Mother’s highest educational attainment by the birth year of the child, father’s age in the birth year of the child, and birth order to father had the highest number of missing values, and there are notable changes in their descriptive statistics after their missing data were imputed.

BIRTH, a three-level ordinal categorical variable was created, with those whose birth weights are <= 3000 grams coded as 1, and those whose birth weights are between 3000 grams and 4000 grams coded as 2, and those whose birth weights are > 4000 grams coded as 3. And the correlations of each covariate with BIRTH and with cognitive composite score were examined and reported in Table 4.13.

Table 4.13

**Correlations of Covariates with BIRTH and Cognitive Composite Score for the 741 Black Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days**

<table>
<thead>
<tr>
<th></th>
<th>BIRTH</th>
<th>cognitive composite score</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>-0.1</td>
<td>0.02</td>
</tr>
<tr>
<td>ageby1997</td>
<td>0.02</td>
<td>-0.07</td>
</tr>
<tr>
<td>mothers age</td>
<td>0.07</td>
<td>0.13</td>
</tr>
<tr>
<td>birth order to mother</td>
<td>0.12</td>
<td>-0.08</td>
</tr>
<tr>
<td>marital status of mother</td>
<td>0.16</td>
<td>0.19</td>
</tr>
<tr>
<td>mothers education</td>
<td>0.15</td>
<td>0.27</td>
</tr>
<tr>
<td>fathers age</td>
<td>0.03</td>
<td>0.07</td>
</tr>
<tr>
<td>birth order to father</td>
<td>0.06</td>
<td>-0.09</td>
</tr>
<tr>
<td>private insurance</td>
<td>-0.05</td>
<td>-0.16</td>
</tr>
<tr>
<td>MEDICAID</td>
<td>-0.1</td>
<td>-0.22</td>
</tr>
<tr>
<td>WIC</td>
<td>-0.09</td>
<td>-0.16</td>
</tr>
<tr>
<td>food stamp</td>
<td>-0.11</td>
<td>-0.12</td>
</tr>
<tr>
<td>ADC.AFDC</td>
<td>-0.07</td>
<td>-0.16</td>
</tr>
</tbody>
</table>
Sex, birth order to mother, marital status of mother, mothers education, and three covariates that describe socioeconomic status of the mother during pregnancy, MEDICAID, WIC, and food stamp, are related with BIRTH, either positively or negatively, with an absolute magnitude larger than 0.08. But ageby1997, mother’s age, father’s age, birth order to father, private insurance, and ADC.AFDC show little or no relationship with BIRTH.

Maternal age in the birth year of the child, marital status of mother in the birth year of the child, mother’s highest educational attainment by the birth year of the child, and the five covariates that describe socioeconomic status of the mother during pregnancy, private insurance, MEDICAID, WIC, food stamp, and ADC.AFDC, are related with cognitive composite score, either positively or negatively, with an absolute magnitude larger than 0.1. But sex, age of the child by 1997, birth order to mother, father’s age, and birth order to father show little relationship with cognitive composite score.

The main effects of all the 13 covariates were first included in the ordinal logistic regression model as predictors to estimate the balancing score. Polynomial terms and interaction terms involving the covariates were tried in the ordinal logit model one after another, and this iterative process led to the finalized model that includes the main effects of the 13 covariates, the interaction between birth order to mother and marital status of mother, and the interaction between birth order to mother and birth order to father. The three groups of children were stratified on the balancing score into five strata.

Figure 4.32 displays the histograms that show how the distributions of the balancing score overlap each other across the three birth weight categories. It is shown
that there is a balancing score region from 0.55 to 0.9 that contains children from all the three birth weight categories, and this region shows support comparisons across the three groups.

Figure 4.32 Histograms showing overlap between the balancing score distributions for the 741 black children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days.
cbal.psa plots to check covariate balance are presented in Figure 4.33. The plot on the left in Figure 4.33 compares the children whose birth weights are \(\leq 3000\) grams (n=240) with those whose birth weights are \(>3000\) grams (n=501); while the plot on the right compares the children whose birth weights are \(\leq 4000\) grams (n=686) with those whose birth weights are \(> 4000\) grams (n=55). The absolute values of the SMDs for most of the covariates with adjustment for balancing score stratification (the black filled circles) are nearer to zero than the absolute SMDs without adjustment for balancing score stratification (the grey unfilled circles), particularly when comparing those whose birth weights are \(\leq 3000\) grams with those whose birth weights are \(> 3000\) grams, the two samples with a comparable size. The graphic shows that “the balancing score” actually does help balance all observed covariate distributions for both left and right plots.

Figure 4.33 cbal.psa plots checking covariate balance for the 741 black children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days.

Figure 4.34 shows the fitted locally linear regression surface relating cognitive composite score to birth weight and balancing score for the blacks after limiting their
balancing scores to the region of overlap from 0.55 to 0.9. The span was set as 0.85. The surface shows that for most of those black children whose gestational ages are normal, their cognitive composite scores increase gradually with increases in birth weights. But this positive association is not consistent across the entire distribution of the balancing score. For those black children with the lowest balancing scores, the association between birth weight and cognitive composite score is upside-down U-shaped; their cognitive composite scores increase gradually with birth weights until an optimum weight around 3400 grams, after which composite scores begin to fall.

Figure 4.34 Fitted surface for locally linear regression (span =0.85) of cognitive composite score on birth weight and balancing score for the 643 black children in the PSID-CDS whose gestational ages are in the normal range from 259 days to 294 days and whose balancing scores are in the overlapping region from 0.55 to 0.9.

Cognitive composite scores were then compared across the three groups of children within each stratum. As displayed in Table 4.14, the children whose birth weights are > 4000 grams have higher mean scores than the children whose birth weights are <= 4000 grams in stratum 2, 3, 4, and 5. The children whose birth weights are between 3000 grams and 4000 grams have higher mean scores than the children whose
birth weights are <=3000 in strata 1, 2, and 5. The estimated mean score after adjusting for the quintiles of the balancing score is -0.39 for those children whose birth weights are <=3000 grams, is -0.35 for those whose birth weights are between 3000 grams and 4000 grams, and is -0.52 for those whose birth weights are >4000 grams. There are only 2 children whose birth weights are > 4000 grams in stratum 1, any differences we see within this stratum could therefore be biased. If we ignore stratum 1, the estimated mean score across the rest four strata is -0.33 for those children whose birth weights are <=3000 grams and -0.34 for those whose birth weights are between 3000 grams and 4000 grams; it is -0.24 for those whose birth weights are >4000 grams.

Table 4.14

Means of Cognitive Composite Score, Counts of Children, Estimated Means after Adjusting for the Quintiles of the Balancing Score and Their 95% Confidence Intervals for the 643 Black Children in the PSID-CDS Whose Gestational Ages Are in the Normal Range from 259 Days to 294 Days and Whose Balancing Scores Are in the Overlapping Region from 0.55 to 0.9

<table>
<thead>
<tr>
<th>strata</th>
<th>balancing score range</th>
<th>means(counts)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>birth weight&lt;=3000</td>
<td>3000&lt;birth weight&lt;=4000</td>
<td>birth weight&gt;4000</td>
</tr>
<tr>
<td></td>
<td>-(0.55, 0.580)</td>
<td>-0.62(23)</td>
<td>-0.41(27)</td>
</tr>
<tr>
<td></td>
<td>(0.580, 0.643)</td>
<td>-0.61(66)</td>
<td>-0.4(73)</td>
</tr>
<tr>
<td></td>
<td>(0.643, 0.708)</td>
<td>-0.18(45)</td>
<td>-0.41(93)</td>
</tr>
<tr>
<td></td>
<td>(0.708, 0.783)</td>
<td>-0.22(38)</td>
<td>-0.35(100)</td>
</tr>
<tr>
<td></td>
<td>(0.783, 0.9)</td>
<td>-0.31(23)</td>
<td>-0.18(104)</td>
</tr>
<tr>
<td>adjusted means</td>
<td>-0.39</td>
<td>-0.35</td>
<td>-0.52</td>
</tr>
<tr>
<td>95% CI</td>
<td>[-0.55, -0.23]</td>
<td>[-0.44, -0.26]</td>
<td>[-1.13, 0.08]</td>
</tr>
</tbody>
</table>
Chapter 5 Summary of the Results and Discussion

Summary of the Results

The primary objective of the current study was to examine the predictive relationship between birth weight and cognitive ability in childhood principally through the effective use of nonparametric loess regression. To reduce the likelihood that observed associations for one particular sample could be spurious, data from two national longitudinal studies conducted in the United States were examined to improve the generalizability of the results. One study was the National Longitudinal study of Youth 1979-Children (NLSY79-C), the other was the Panel Study of Income Dynamics-Child Development Supplement (PSID-CDS). The NLSY79-C started their first wave of data collection with their subjects and their families in 1986, while the PSID-CDS began in 1997. Only the data collected from the first wave were analyzed and the results were compared across the two nationally representative samples, although there was a ten-year time difference in starting dates. Besides birth weight and cognitive capabilities, the two studies also collected data on gestational age and on the covariates that could confound the association between birth weight and cognitive ability in childhood.

Both studies oversampled minorities, particularly blacks and Hispanics. However, the racial background of the children in the NLSY79-C was coded in a simpler way, with Hispanics as one group, blacks as a second group, and non-black-non-Hispanics as a third group which included whites as well as minorities other than blacks and Hispanics. The sample of Hispanics was excluded from analyses because of its small size. The final sample for this study was composed of 1434 non-black-non-Hispanic and black singleton births. Similarly, the minorities other than blacks in the PSID-CDS were also excluded,
resulting in a final sample of 1763 white and black singleton births. The 781 non-black-non-Hispanic children in the NLSY79-C were treated as counterparts to the 931 whites in the PSID-CDS, although they might also include non-black-non-Hispanic minorities such as American Indians and Asians. The 653 black children in the NLSY79-C were compared with the 832 blacks in the PSID-CDS. It is known that the association between birth weight and cognitive ability varies by sex and race, the association was therefore examined first for whites (or non-black-non-Hispanics in the NLSY79-C), and for white (or non-black-non-Hispanic) males and for white (or non-black-non-Hispanic) females separately; then for blacks, and for black males and females separately.

The ages of the 1434 non-black-non-Hispanic and black children in the NLSY79-C sample varied from 5 to 13 in 1986 when they were first interviewed, while the ages of the 1763 white and black children in the PSID-CDS sample varied from 3 to 13 in 1997. As shown in Table 5.1, whites (or non-black-non-Hispanics) have heavier birth weights and longer gestational ages than blacks. In these ten years from 1986 to 1997, there was a slight increase in either birth weights or gestational ages for both whites (or non-black-non-Hispanics) and blacks.
Table 5.1

Descriptive Statistics for Birth Weight and Gestational Age for the Non-Black-Non-Hispanics and for the Blacks in the NLSY79-C and for the Whites and for the Blacks in the PSID-CDS

<table>
<thead>
<tr>
<th></th>
<th>Birth weight</th>
<th>Gestational Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>non-black-non-Hispanics in NLSY79-C</td>
<td>3326</td>
<td>3345</td>
</tr>
<tr>
<td>whites in PSID-CDS</td>
<td>3452</td>
<td>3430</td>
</tr>
<tr>
<td>blacks in NLSY79-C</td>
<td>3060</td>
<td>3090</td>
</tr>
<tr>
<td>blacks in PSID-CDS</td>
<td>3160</td>
<td>3203</td>
</tr>
<tr>
<td>non-black-non-Hispanics in NLSY79-C (weeks)</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>whites in PSID-CDS</td>
<td>278</td>
<td>280</td>
</tr>
<tr>
<td>blacks in NLSY79-C</td>
<td>38</td>
<td>45</td>
</tr>
<tr>
<td>blacks in PSID-CDS</td>
<td>276</td>
<td>280</td>
</tr>
</tbody>
</table>

The two data sets used different but comparable cognitive tests to evaluate the cognitive ability of their subjects. The NLSY79-C used the Peabody Individual Achievement Tests (PIAT). The score from the math subtest and the score from the reading recognition subtest were combined to generate a composite score as the sole indicator of cognitive ability. The PSID-CDS used the Woodcock-Johnson Revised Tests of Achievement (WJ-R). The scores from the two WJ-R subtests (applied problem and letter-word identification) that assess very close constructs as those assessed by the PIAT math and reading recognition subtests were combined to generate a composite score.

The scatterplot matrices (Figure 5.1) display similar joint distributions among cognitive composite score, gestational age, and birth weight across the two samples. Birth weight is moderately correlated with gestational age, but only modestly correlated with cognitive composite scores. The correlation of birth weight with cognitive composite
score is stronger for whites (or non-black-non-Hispanics) than for blacks ($r = 0.06$ for the non-black-non-Hispanics vs. $r = 0.04$ for the blacks in the NLSY79-C; $r = 0.09$ for the whites vs. $r = 0.04$ for the blacks in the PSID-CDS). There is also a slightly positive association between gestational age and cognitive composite score. Males are heavier at birth than females, but there are no evident differences between males and females in their gestational ages and average cognitive composite scores.

Figure 5.1 Scatterplot matrices for cognitive composite score, gestational age, birth weight, and sex for the Non-Black-Non-Hispanics and for the Blacks in the NLSY79-C and for the Whites and for the Blacks in the PSID-CDS.

To answer the first research question, “What is the nature of the predictive association of birth weight and gestational age for cognitive ability in childhood?” a
locally linear regression surface was generated for each subsample defined by sex and race relating cognitive composite score to both birth weight and gestational age.

The birth weights of the 781 non-black-non-Hispanic children in the NLSY79-C have a range from 936 grams to 4933 grams, while the birth weights of the 653 black children have a range from 907 grams to 5613 grams. In the area where most of the data are distributed, no consistent and interpretable association was observed between birth weight and cognitive composite score at any fixed level of gestational age either for the non-black-non-Hispanics or for the blacks. When the non-black-non-Hispanic sample and the black sample were broke into subgroups by sex, a positive association was observed between birth weight and cognitive composite score only for the black males. The birth weights of the 931 white children in the PSID-CDS have a range from 680 grams to 6038 grams, while the birth weights of the 832 black children have a range from 496 grams to 6917 grams. In the area where most of the data are distributed, a positive association was observed between birth weight and cognitive composite score at any fixed level of gestational age for the whites, but not for the blacks. When the white sample and the black sample were broken into subgroups by sex, a positive association between birth weight and cognitive composite score was observed for both the white males and the white females as well as for the black females. But the association among the black males has an upside-down U shape. The results from the two samples do not conform well with each other.

The gestational ages of the 781 non-black-non-Hispanic children in the NLSY79-C have a range from 26 weeks (182 days) to 46 weeks (322 days), while the gestational ages of the 653 black children have a range from 26 weeks (182 days) to 45 weeks (315
days). In the area where most of the data are distributed, no consistent and interpretable association was observed between gestational age and cognitive composite score at any fixed level of birth weight for any of the subsamples except for the black males among whom the association has an upside down U-shape with an optimum gestational age being around 35 weeks. The gestational ages of the 931 white children in the PSID-CDS have a range from 182 days (26 weeks) to 318 days (45 weeks), while the gestational ages of the 832 black children have a range from 168 days (24 weeks) to 336 days (48 weeks). In the area where most of the data are distributed, a positive association was observed between gestational age and cognitive composite score at any fixed level of birth weight only for the white females. The results from the two samples do not conform well with each other.

To answer the second research question, “What is the nature of the predictive association of birth weight for cognitive ability in childhood”, gestational age was restricted within the normal range, i.e., between 37 weeks (259 days) and 42 weeks (294 days) to control its confounding effects. Because birth weight is moderately correlated with gestational age, those children with extremely low and high birth weights were consequently excluded from further analyses. The birth weights of the remaining 654 non-black-non-Hispanic children in the NLSY79-C whose gestational ages are normal have a range from 1500 grams to 5000 grams, while the birth weights of the remaining 552 black children whose gestational ages are normal have a range from 1000 grams to 5000 grams, with one outlier being 5613 grams. Similarly, the birth weights of the remaining 834 white children in the PSID-CDS whose gestational ages are normal have a range from 2000 grams to 5000 grams, with one outlier being 6038 grams, while the birth
weights of the remaining 741 black children whose gestational ages are normal have a range from 1000 grams to 5000 grams.

Simple locally linear regression of cognitive composite score on birth weight was conducted separately for each subsample defined by sex and race whose gestational ages are normal. The results were summarized and presented in Figure 5.2. After the confounding effects of gestational age was controlled, the association of birth weight with cognitive composite score has an upside-down U shape among the non-black-non-Hispanic children in the NLSY79-C whose gestational ages are normal and also among the white children in the PSID-CDS whose gestational ages are normal. Their cognitive composite scores increase gradually with birth weights until an optimum weight, after that their composite scores begin to fall. The optimum weight for the non-black-non-Hispanic children in the NLSY79-C is around 3600 grams, while the optimum weight for the white children in the PSID-CDS is much higher, being around 4300 grams, and there are only a few data beyond this optimum weight.

When the non-black-non-Hispanic sample in the NLSY79-C and the white sample in the PSID-CDS were broken into subgroups by sex, the association of birth weight with cognitive composite score has an upside-down U shape for the non-black-non-Hispanic males and for the non-black-non-Hispanic females in the NLSY79-C whose gestational ages are normal as well as for the white males in the PSID-CDS whose gestational ages are normal. Their cognitive composite scores increase gradually with birth weights until an optimum weight, after that their composite scores begin to fall. But for the white females in the PSID-CDS whose gestational ages are normal, the association is generally a continuously increasing one.
The loess curve relating cognitive composite score to birth weight for the black children in the NLSY79-C whose gestational ages are normal and the loess curve for the black children in the PSID-CDS whose gestational ages are normal are almost flat in the area where most of the data are distributed. But for the black children in the NLSY79-C, a slight decline was observed in their cognitive scores both for the children with extremely low birth weights and for the children with extremely high birth weights, while for the black children in the PSID-CDS, a slight decline was observed only for the children with extremely low birth weights. The association is stronger for the whites (or the non-black-non-Hispanics) than for the blacks.

When the two black samples were broken into subgroups by sex, a slight growth pattern was observed for the black males in the NLSY79-C whose gestational ages are normal, but not for the black females. On the contrary, a slight growth pattern was observed for the black females in the PSID-CDS whose gestational ages are normal, but not for the black males.
Figure 5.2 Comparisons of loess curves relating cognitive composite score to birth weight when gestational age was restricted to be normal between the NLSY79-C and the PSID-CDS first by race then by gender for whites (or non-black-non-Hispanics) and for blacks.

The association of birth weight with cognitive composite score has a similar shape between whites (or non-black-non-Hispanics) and blacks. An upside-down U shape was observed both for the non-black-non-Hispanics and for the blacks in the NLSY79-C whose gestational ages are normal. And the association is generally a continuously increasing one both for the whites and for the blacks in the PSID-CDS whose gestational
ages are normal. The slight decrease on the upper end of the association for the whites is caused by a few outliers whose birth weights are beyond 4300 grams, the optimum weight. When comparing the cognitive composite scores of the non-black-non-Hispanics in the NLSY79-C whose gestational ages are normal with those of the blacks whose gestational ages are also normal, the SMD between the two groups is 0.45. And the SMD between the whites in the PSID-CDS whose gestational ages are normal and the blacks whose gestational ages are also normal is 0.7. These two SMDs show that there are notable differences between whites (or non-black-non-Hispanics) and blacks in their cognitive scores. At every fixed level of birth weight, whites (or non-black-non-Hispanics) have higher cognitive scores, on average, than blacks.

The SMD between the non-black-non-Hispanic males in the NLSY79-C whose gestational ages are normal and the non-black-non-Hispanic females whose gestational ages are normal is 0.03, while the SMD between the white males in the PSID-CDS whose gestational ages are normal and the white females whose gestational ages are also normal is 0.06, indicating gender differences are trivial among whites (or non-black-non-Hispanics), but white (or non-black-non-Hispanic) males have slightly higher cognitive scores than white (or non-black-non-Hispanic) females.

The SMD between the black males in the NLSY79-C whose gestational ages are normal and the black females whose gestational ages are normal is -0.13, while the SMD between the black males in the PSID-CDS whose gestational ages are normal and that of the black females whose gestational ages are normal is -0.04, indicating gender differences are trivial also among blacks, but black females have slightly higher cognitive scores than black males.
A second objective of the current study was to investigate whether observed associations between birth weight and cognitive ability increases or decreases after reducing (at least some) effects of confounding of multiple covariates. That is, besides gestational age, prenatal differences were adjusted using generalized propensity score methods using several available covariates. It had been found in the literature that the association between birth weight and cognitive ability in childhood is confounded with birth order, maternal age, maternal intelligence, parental education, maternal physical and mental conditions, maternal smoking and drinking habits during pregnancy, socioeconomic status of the family, and etc. Table 5.2 lists the covariates in each study whose confounding was controlled. Although the two datasets were carefully examined so that all of these covariates that had ever been studied in previous existing literature could be included, the covariates in some important areas such as maternal IQ, maternal physical and mental conditions, and maternal smoking and drinking habits during pregnancy are not available from the PSID-CDS. And some covariates selected from the NLSY79-C such as mother’s highest educational attainment, family net income, and marital status of mother suffer from a severe missing data problem.

Table 5.2

Comparisons of Confounding Covariates between the NLSY79-C and the PSID-CDS

<table>
<thead>
<tr>
<th></th>
<th>NLSY79-C</th>
<th>PSID-CDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>sex</td>
<td>sex</td>
</tr>
<tr>
<td>birth order</td>
<td>birth order to mother</td>
<td>1. birth order to mother</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. birth order to father</td>
</tr>
<tr>
<td>maternal age</td>
<td>maternal age in the birth year of the child</td>
<td>maternal age in the birth year of the child</td>
</tr>
<tr>
<td>maternal IQ</td>
<td>mother’s IQ that was assessed in 1980 by the armed forces</td>
<td>none</td>
</tr>
<tr>
<td>Category</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Parental Education</td>
<td>mother’s highest educational attainment by the birth year of the child</td>
<td></td>
</tr>
<tr>
<td>Maternal Physical and Mental Conditions</td>
<td>maternal weight before pregnancy</td>
<td></td>
</tr>
<tr>
<td>Maternal Smoking and Drinking Habits</td>
<td>1. if the mother drank alcohol during pregnancy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. if the mother smoked during pregnancy</td>
<td></td>
</tr>
<tr>
<td>SES of the Family</td>
<td>family net income in the birth year of the child</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Others</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. whether some of the medical expenses related to the pregnancy or the delivery of the child were covered by private health insurance</td>
</tr>
<tr>
<td></td>
<td>2. whether Medicaid paid for any of these medical bills</td>
</tr>
<tr>
<td></td>
<td>3. whether the mother was in the WIC program when she was pregnant with the child</td>
</tr>
<tr>
<td></td>
<td>4. whether the mother received any government food stamps while she was pregnant</td>
</tr>
<tr>
<td></td>
<td>5. whether the mother received any payments from ADC or AFDC while she was pregnant</td>
</tr>
</tbody>
</table>

Comparisons were made between the whites (or the non-black-non-Hispanics) and the blacks in those covariates. Comparing with the blacks, the whites (or the non-black-non-Hispanics) are more likely to be later in birth order. And their mothers are
more likely to be married when they were born, to have higher IQ score, to be better educated, to drink alcohol and to smoke during pregnancy, and to have higher family net income, but are less likely to have private insurance coverage, and to receive financial support from government and public agencies during pregnancy.

Comparisons were also made across the two studies. Compared with the children first seen in 1986, the children first tested in 1997 are more likely to be later in birth order. And their mothers are more likely to be older, to be married when their child was born, and to be better educated. It is not possible, however, to know whether these differences reflect more the nature of the two samples or changes in the population at large.

The stage of analyses that employed generalized propensity score methods followed the previous stage where locally linear regression had been used to depict the association between birth weight and cognitive ability when gestational age was restricted to the normal range. The results from locally linear regression showed notable differences between whites (or non-black-non-Hispanics) and blacks, but gender differences within each racial group were found to be trivial. This stage of analyses was therefore conducted separately for the whites (or the non-black-non-Hispanics) and for the blacks whose gestational ages are normal, without further subdivision into male and female subgroups. The birth weight distribution of each racial sample was divided into three adjacently positioned categories as demonstrated in Figure 5.3. A balancing score was estimated from ordinal logit regression for each individual. And each of the three birth weight categories was further stratified into five strata according to the quintiles of the balancing score.
Figure 5.3 Plots to show how the three birth weight categories were created for each subsample.

Figure 5.4 compares the fitted locally linear regression surfaces relating cognitive composite score to birth weight and balancing score between the NLSY79-C and the PSID-CDS. The plot on the upper left is the fitted surface for the non-black-non-Hispanic children in the NLSY79-C whose gestational ages are normal. The surface shows that after limiting their balancing scores to the region of overlap, the association of birth weight with cognitive composite score is upside-down U-shaped at any fixed level of gestational age, their cognitive composite scores increase gradually with birth weights until an optimum weight around 3600 grams, after that their composite scores begin to fall. The plot on the upper right is the fitted surface for the white children in the PSID-CDS whose gestational ages are normal. The surface shows that for those white children whose balancing scores are beyond 0.7, the association of birth weight with cognitive composite score is upside-down U-shaped, their cognitive composite scores increase gradually with birth weights until an optimum weight around 4000 grams, after that their composite scores begin to fall. But this upside-down U shape doesn’t persist across the
entire distribution of balancing score, for those white children whose balancing scores are between 0.55 and 0.7, the association is a continuously increasing one.

The plot on the lower left is the fitted surface for the blacks in the NLSY79-C whose gestational ages are normal. The surface shows that, after limiting their balancing scores to the region of overlap, the association of birth weight with cognitive composite score is upside-down U-shaped at any fixed level of gestational age, their cognitive composite scores increase gradually with birth weights until an optimum weight around 3300 grams, after that their cognitive composite scores begin to fall. The plot on the lower right is the fitted surface for the blacks in the PSID-CDS whose gestational ages are normal. The surface shows that for most of those black children, the association of birth weight with cognitive composite score is a continuously increasing one, their cognitive composite scores increase gradually with birth weights. But this continuous increase is not consistent across the entire distribution of balancing score. For those black children with extremely low balancing scores, the association has an upside-down U shape.

The shape of the fitted surfaces are similar between the whites (or the non-black-non-Hispanics) and the blacks within each study. The fitted surfaces are upside-down U-shaped both for the non-black-non-Hispanics and for the blacks in the NLSY79-C, and are continuously increasing both for the whites and for the blacks in the PSID-CDS, though the continuous increase is not consistent across the entire distribution of balancing score. But the fitted surfaces for the whites (or non-black-non-Hispanics) have steeper slopes than those for the blacks to indicate that when the effects of multiple covariates
that account for prenatal differences are controlled, the association is still stronger among the whites (or the non-black-non-Hispanics) than among the blacks.

![Graphs showing comparisons of fitted locally linear regression surfaces relating cognitive composite score to birth weight and balancing score between the NLSY79-C and the PSID-CDS.](image)

**Figure 5.4** Comparisons of fitted locally linear regression surfaces relating cognitive composite score to birth weight and balancing score between the NLSY79-C and the PSID-CDS when gestational age was restricted to be normal and balancing score was restricted to the region of overlap.

Table 5.3 lists the observed mean scores for each birth weight category and their 95% confidence interval, together with their estimated mean scores after stratifying on the quintiles of the balancing score and their 95% confidence interval, first for the non-black-non-Hispanics in the NLSY79-C and the whites in the PSID-CDS whose gestational ages are normal and whose balancing scores are in the region of overlap, and
then for the blacks in the NLSY79-C and the blacks in the PSID-CDS whose gestational ages are normal and whose balancing scores are in the region of overlap.

As demonstrated by their observed means, the whites (or non-black-non-Hispanics) whose birth weights are around the average have the highest mean score, they are followed by the whites whose birth weights are lower than them and then by those whose birth weights are higher. After differences in covariates were adjusted for using generalized propensity score methods, this upside-down U-shaped association was strengthened for the non-black-non-Hispanics in the NLSY79-C, but it was converted to a continuously increasing one for the whites in the PSID-CDS with the decrease at the highest birth weights changed to a rough increase.

The results for the blacks in the NLSY79-C do not confirm with the results for the blacks in the PSID-CDS, again at the highest birth weights. In the NLSY79-C, as demonstrated by their observed means, the blacks whose birth weights are around the average have the highest mean score, they are followed by the blacks whose birth weights are lower than them, and then by those whose birth weights are higher. This upside-down U-shaped association was strengthened after differences in covariates were adjusted for using generalized propensity score methods. In the PSID-CDS, however, the blacks whose birth weights are the highest have the highest mean score, they are followed by the blacks whose birth weights are around the average, and then those whose birth weights are the lowest. This rough increase, particularly at the highest birth weights was strengthened after differences in covariates were adjusted for using generalized propensity score methods.
Table 5.3

Comparisons of Observed Mean Scores for Each Birth Weight Category with Their Estimated Mean Scores after Adjusting for the Quintiles of the Balancing Score

<table>
<thead>
<tr>
<th>Birth Weight Category</th>
<th>Non-black-Hispanics in NLSY79-C</th>
<th>Whites in PSID-CDS</th>
<th>Blacks in NLSY79-C</th>
<th>Blacks in PSID-CDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>observed means</td>
<td>observed means</td>
<td>observed means</td>
<td>observed means</td>
</tr>
<tr>
<td>lower birth weight</td>
<td>0.15[-0.06,0.3]</td>
<td>0.32[0.2,0.44]</td>
<td>-0.25[-0.43,-0.07]</td>
<td>-0.4[-0.5,0.3]</td>
</tr>
<tr>
<td>middle birth weight</td>
<td>0.25[0.17,0.33]</td>
<td>0.37[0.29,0.45]</td>
<td>-0.16[-0.27,-0.05]</td>
<td>-0.34[-0.42,-0.26]</td>
</tr>
<tr>
<td>estimated means (95%CIs)</td>
<td>0.02[-0.23,0.27]</td>
<td>0.29[0.08,0.5]</td>
<td>-0.55[-0.83,-0.27]</td>
<td>-0.3[-0.57,-0.03]</td>
</tr>
<tr>
<td>higher birth weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adjusted means</td>
<td>0.15[-0.1,0.4]</td>
<td>0.37[0.28,0.46]</td>
<td>-0.2[-0.3,-0.1]</td>
<td>-0.39[-0.55,-0.23]</td>
</tr>
<tr>
<td></td>
<td>0.26[0.16,0.36]</td>
<td></td>
<td>-0.42[-0.65,-0.19]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.11[-0.47,0.25]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

The first research question, “What is the nature of the predictive association of birth weight and gestational age for cognitive ability in childhood?”, was not answered in exactly the same way using data for the two samples using locally linear regression. One could argue that there is insufficient evidence to support either an association or no association between birth weight and cognitive ability or between gestational age and cognitive ability, especially for the full ranges of these variables. Two reasons can possibly explain this finding. One is that both birth weight and gestational age variables had wide ranges, with a number of outliers at both extremes. So the three-dimensional loess surface tends to be distorted by these outliers. In addition, gestational age was estimated in weeks in the NLSY79-C and therefore majority of these children have a gestational age of 38 weeks. The associations particularly those between gestational age
and cognitive composite score, at least for this data set, could be disguised by lack of variance in the data.

More revealing analyses resulted from restricting attention to children whose gestational age was normal (37-42 weeks gestational age) or who were born at term. After placing the “normal” restriction on gestational age, birth weight ranges from 1500 grams to 5000 grams for whites (or non-Black-non-Hispanics), and from 1000 grams to 5000 grams for blacks. Therefore, the second research question that further analyses were intended to answer is actually asking about what the predictive association of birth weight for cognitive ability in childhood is like among children who were born at term and who were born with normal birth weight. If any associations were observed, the observed associations cannot be driven by those with extremely low birth weight since they were excluded.

The results of the loess regressions of cognitive composite score on birth weight show an increase in cognitive scores with increasing birth weights among lower birth weights until the reach of an optimum weight. A decrease after the optimum weight was observed for both the non-black-non-Hispanics and the blacks in the NLSY79-C. But this decrease among higher birth weights is not evident either for the whites or for the blacks in the PSID-CDS. The results for whites did show a decrease at the highest birth weights beyond 4300 grams; however, there are only a few such cases. In other words, the association of birth weight with cognitive score has a more or less upside-down U shape among the children in the NLSY79-C, but the association is roughly a continuously increasing one among the children in the PSID-CDS.
Disparity between the two studies in terms of the shape of the association between birth weight and cognitive score were probably caused by factors beyond control of the current study. After all, they are two studies that were conducted in different time periods with two different samples, variations between the two studies in multiple aspects coexist with similarities. It is therefore not surprising to see differences in the results. For instance, the two studies used different achievement tests to measure cognitive capabilities in childhood, and validity as well as reliability of these tests, such as internal consistency, was not reported for the sample recruited by any of the two studies. Although in the current study, special attention was paid to select the scores from two subtests in each test that assess similar constructs to be combined as cognitive composite score, the difference in the shape of the association between birth weight and cognitive score, particularly as what we see among higher birth weights, could still be caused by the different formats the two tests have or by the different ways in which children, particularly those with higher birth weight, respond to the two tests.

Another potential cause for the difference in the shape of the association is the ten-year time span between the two studies. If any other variations are assumed to be ignorable, it can be concluded that, based on the results from the two studies, the disadvantages the children born with higher birth weight had in their cognitive ability in 1986 hardly existed ten years later in 1997 among the children who were also born with higher birth weight. Interpretation of the difference in this way is more interesting and promising, but whether it is trustworthy has to be verified in future work.

Despite of this notable difference, results from the two studies display considerable consistency. It is shown that within each study, the shape of the association
of birth weight with cognitive ability is similar between whites (or non-black-non-Hispanics) and blacks. But the association is stronger for whites (or non-black-non-Hispanics) than for blacks, and at any fixed level of birth weight, whites (or non-black-non-Hispanics) have higher cognitive score than blacks. Racial differences in terms of cognitive ability in childhood are consistent across the two studies, although the non-black-non-Hispanic sample in the NLSY79-C includes whites as well as non-black-non-Hispanic minorities. This finding is also consistent with that of previous studies such as Rowe (2002), Kiweon (1992), and Hardy & Mellits (1977). It is surprising to find that the mechanism underlying racial differences in the association of birth weight with cognitive ability has not been elaborated very much in the literature. There have been attempts to identify causes related to children’s social environments. Compared to blacks, whites generally are better-educated, economically richer, and less likely to be subject to racial discrimination. This can partially explain why the association of birth weight with cognitive score is stronger among whites, and why whites have higher cognitive scores than blacks with comparable birth weights. However, the social environment explanation still lacks empirical evidence. In some studies (e.g., Brooks-Gunn, Klebanov, & Duncan, 1996; Rowe, 2002), the differences remained intact after social class, a proxy measure of social environment, was statistically adjusted for. There have also been attempts to contribute these differences to genetics (Rowe, 2002; 2005). But there has been an unsettled debate over the genetic explanation. Opponents (such as Sternberg, Grigorenko, & Kidd, 2005; Cooper, 2005) argue race is a social construction imposed by history, so any attempts to genetically link race to intelligence are not scientifically grounded. Even
advocates (such as Rowe) admit that the specific genetic mediators are largely unknown at this time.

Gender differences were examined separately for whites (or non-black-non-Hispanics) and for blacks. Gender differences in the association of birth weight with cognitive score are trivial within each racial sample. The results from the two studies fail to support a stronger association either for males or for females, which is contradictory to the findings from those previous studies (e.g., Jefferis et al., 2002; Matte et al., 2001; Shenkin et al., 2001). It can be argued that, of course, the small sample size we have after further breaking each racial group into males and females may not be sufficient to detect any associations. It can also be argued that racial differences are confounded with gender differences and adjusting for racial differences could diminish differences between the two genders. The findings from the two samples indicate that, on average, white (or non-black-non-Hispanic) males have slightly higher cognitive score than white (or non-black-non-Hispanic) females, while black females have slightly higher cognitive score than black males, although the advantages enjoyed by white males and by black females are not consistent across the entire distribution of birth weight. Boys and girls exhibit different fetal growth rates (Forsen et al., 1999) and thus may respond differently to prenatal insults (Hutchison, 1997). This argument can be used to justify sex differences in either direction.

Confounding of multiple covariates was adjusted for using generalized propensity score methods, which included sex, age, birth order, and covariates that account for prenatal differences. Notable differences were observed between whites and blacks in the covariates that measure SES of the family and in those that are closely related. For
example, the mean family net income of the whites is notably higher than that of the blacks. There are much more white mothers who were married when their child was born, but much less of them received financial support from government and public agencies when they were pregnant. Although the differences were restricted to be prenatal, this can be the evidence that partially supports the social environmental approach to explain the differences between whites and blacks in the association of their birth weight with their cognitive score. Whites have better prenatal social environment and possibly better postnatal environment as well.

When multiple covariates were adjusted for, the upside-down U-shaped association observed both for the non-black-non-Hispanics and for the blacks in the NLSY79-C persisted and was slightly strengthened. The decrease in the cognitive score at the highest birth weights around and beyond 4000 grams became even more severe. The continuously increasing association observed both for the whites and for the blacks in the PSID-CDS persisted and was also strengthened, with the decrease observed for the whites at the highest birth weights beyond 4000 grams changed to a rough increase and with the slight increase observed for the blacks around 4000 grams and beyond becoming even more evident.

The finding that the association persisted and was slightly strengthened after adjustment for covariates is contradictory to that of most of the studies previously conducted. For example, Deary, Der, and Shenkin (2005) also analyzed the data from the NLSY79-C, and it was found that the association of birth weight with cognitive ability was attenuated after taking into account mother’s IQ. Deary et al., (2005) used random effects models that assumed the association between birth weight and cognitive ability is
simply a linear one, but the graphic results from the current study consistently show the association is upside-down U shaped among the children in the NLSY79-C. When covariates, such as mother’s IQ, is taken into the model, the regression coefficient of the linear effect of birth weight could diminish in magnitude but the upside-down U shape could actually become stronger. And Deary et al., (2005) only controlled for the confounding effects of a small number of covariates, including race, gestational age, the child’s age at the time of testing, and mother’s IQ. It is hard to tell if the association would still be attenuated when multiple covariates were controlled as what was done in the current study. A number of previous studies that showed an attenuation in the magnitude of the association, such as Gorman (2002), had controlled for the effects of postnatal environment, while the current study controlled for the confounding effects of genetics and prenatal environment with postnatal environment left out. The results from the current study imply that prenatal environment is not very important as the association of birth weight with cognitive ability persisted. But postnatal environment could have a much stronger influence and controlling for postnatal environment could attenuate or enhance the associations.

With regard to the decrease in cognitive scores at the highest birth weights, Richards et al., (2002) think it may have been due to these births occurring later in birth order who are more likely to have higher birth weight but lower cognitive scores. In their study, there was no longer a decrease in cognitive ability at birth weights above 5000 grams when birth order was corrected. But this point of view can hardly hold in the current study. There were more births later in birth order in 1997 than in 1986. We should have expected a more severe decrease among the children with higher birth weight in the
PSID-CDS, which is not the case according to the results. But the results from the PSID-CDS do show that, after birth order and any other covariates were adjusted for, the slight decrease in the score at the highest birth weights disappeared. Another possible contributor to the decrease at the highest birth weights is maternal illnesses which are not accounted for in the current study. Large babies are more likely to be born to mothers with diabetes, and these babies are known to be at higher risk for perinatal complications, however, little has been known about their long-term outcomes, including cognitive ability (Shenkin, Starr, & Deary, 2004).

**Limitations and Recommendations for Future Research**

In the current study, loess regression and propensity score methods were chosen over traditional methods because of their strengths, but their limitations also should be noted, as described below. The special feature of loess regression of fitting weighted regression models to localized subsets of data demands parameters from multiple models to be estimated, and the substantial number of calculations involved requires a large sample size and the results are hard to summarize with a few numerical statistics. In the current study, the size of the two samples is smaller than could be desirable, and this is especially true for subsamples defined by race and sex. The graphic representations of the results from loess regression did not ameliorate the problem of small sample size. The fitted loess curves and the three-dimensional surfaces are less trustworthy particularly at the both ends of the birth weight distribution or of the gestational age distribution where there are sparse data points.

Furthermore, the methods used this study had not been used previously in this context, and it is not easy to compare the graphic results presented in the current study
with the numeric results reported in the previously existing literature that had included conditional means, correlation coefficients, regression parameter estimates, and inferential statistics to determine whether the association is statistically “significant”. The primary interest of the current study was to examine the shape of the association between birth weight and cognitive ability, or what the association is like, before and after prenatal differences are controlled, rather than in whether the association is statistically “significant”. To improve the generalizability of the results, instead of using inferential methods such as constructing a confidence band around the fitted curve, the current study analyzed two datasets and the results were compared across. But there is a ten-year difference between the data collection times for these two datasets, so the two datasets were based on two populations that are different in a way largely unknown. Differences were indeed detected in the results across the two data sets. In this case, conclusions regarding the association between birth weight and cognitive ability in childhood can not be unified unless the results are verified by a third sample.

The associations depicted graphically in the current study are generally weak. Because so many other factors beyond birth weight, especially genetic and after birth environmental ones, have roles in determining cognitive performance, this result is not surprising. Furthermore, as previously noted, current sample sizes, and especially subsample sizes, are relatively small.

The use of propensity score methods to control the effects of confounders is also novel to research on the association between birth weight and cognitive ability. Propensity score methods may be able to estimate the gap in cognitive ability that can be contributed to differences in birth weight, while any relevant variables other than birth
weight and cognitive ability are treated as confounders and their contributions to cognitive ability are disguised. However, treating any other variables as confounders can be oversimplified when studying a complex phenomenon such as cognitive ability. It is clear that birth weight explains only a small proportion of the variance in cognitive ability, and the large proportion of the variance is certainly contributed by the variables that are treated as confounders in propensity score methods and by the variables that are not controlled for. The relative importance of these variables themselves and their interrelationships with one another, with reference to the causal chain from birth weight to cognitive ability, can be very interesting and research on such variables could provide insights about mechanisms that drive the associations. If this is the only research goal, propensity score methods should not be used because they control for but do not estimate the effects of “confounders”. Statistical techniques more common in psychological research, such as path analysis and structural equation modeling as done in Silva, Metha, & O’Callaghan (2006), might be useful for such purposes.

The confounders whose effects were controlled for in the current study were strictly confined to those whose quantity is a constant such as birth order and maternal IQ and those that are prenatal and are unaffected by birth weight. The “prenatal” restriction on the confounders leaves out postnatal social and economic factors that have far more important influences on children’s cognitive ability than prenatal ones. It is therefore likely that the associations observed in the current study were the partly due to residual confounding, that is, the associations were affected by postnatal social and economic factors that were not included in the analyses. The current study has only examined the importance of prenatal environment but future research is needed to assess the
importance of postnatal environment and the interactions of variables reflecting prenatal influences with early and late postnatal experiences.

Loess regression has been extended for up to four predictors. Thus, theoretically, the simple and two-predictor loess regression methods used in the current study could be extended in future research to account for both prenatal and postnatal differences. If it could be assumed that prenatal differences have been summarized in the balancing score estimated by propensity score methods, and the balancing score has been used together with birth weight in loess models as predictors of cognitive ability, the variables that describe postnatal experiences could be added to the models as third or even fourth predictors. Postnatal environment has been classified into within-family effects and between-family school and neighborhood effects. Models that use sibling data (e.g., Gorman, 2002b; Yang et al., 2008) have been acknowledged as the models that can, to some degree, estimate and differentiate within-family effects and between-family effects. Similar nonparametric models that use sibling data can be also developed. Using Yang et al. (2008)’s model as an example, mean birth weight can be calculated for each family as an indicator of between-family effects and a deviation from the family mean can be calculated for each sibling as an indicator of within-family effects. In similar loess regression models, the mean birth weight of each family could serve as one predictor, and the deviation from the family mean of each sibling could serve as a second predictor; then a balancing score estimated by propensity score methods could become a third predictor if prenatal environment based on differences among siblings. These models are only conceptual at this time, and experience would be needed to learn whether or how much such approaches might be effective, even if large samples were to be available.
Future research can also be pursued by separating cognitive ability in math from the ability in reading. In the current study, the scores from a math subtest and the scores from a reading subtest were combined to generate a composite score as the sole indicator of cognitive ability. But birth weight could affect the math ability of the general population in a way different from how it affects the reading ability, and further, the causal path from birth weight to math ability might be different from the causal path to reading ability in subpopulations defined by race and sex. The same analyses can be repeated separately on the math score and on the reading score. The results could be helpful to those people who intend to facilitate cognitive development of the general population by providing targeted interventions.
References


Landry, S. H., Smith, K.E., Miller-Loncar, C. L., & Swank, P. R. (1997). Predicting cognitive-language and social growth curves from early maternal behaviors in
children at varying degrees of biological risk. Developmental Psychology, 33, 1040-1053.


McKeown, T., & Gibson, J. R. (1951). 1951 observations on all births (23,970) in Birmingham, 1947, II. British Journal of Social Medicine, 5, 98-112.


Appendix

R Code for Major Analyses

1. **Fitted surface for locally linear regression (span=0.85) of cognitive composite score on birth weight and gestational age**

   birth = seq(min(birth.weight), max(birth.weight), len=25)
   gest = seq(min(gestational.age), max(gestational.age), len=25)
   mod.lo = loess(cognitive.composite.score ~ birth.weight + gestational.age, span=0.85, degree=1)
   newdata = expand.grid(birth.weight=birth, gestational.age=gest)
   fit = matrix(predict(mod.lo, newdata), 25, 25)
   persp(birth, gest, fit, theta=30, phi=30, ticktype='detailed', xlab='birth weight in grams', ylab='gestational age in days', zlab='cognitive composite score', zlim=c(-1.3, 1), expand=2/3, shade=0.5)
   persp(birth, gest, fit, theta=60, phi=30, ticktype='detailed', xlab='birth weight in grams', ylab='gestational age in days', zlab='cognitive composite score', zlim=c(-1.3, 1), expand=2/3, shade=0.5)

2. **2D scatterplots with loess curves showing birth weight in relation to cognitive composite score by race**

   plot(birth.weight[race==1], cognitive.composite.score[race==1], xlab="birth.weight in grams", ylab="cognitive composite score", pch=1, col="black", lwd=2, xlim=c(1474, 6038), ylim=c(-3.67, 3.46))
   points(birth.weight[race==2], cognitive.composite.score[race==2], lwd=2, pch=1, col="lightgrey")
rug(birth.weight[race==1], ticksize = 0.02, side = 1, col = "black")

rug(cognitive.composite.score[race==1], ticksize = 0.02, side = 2, col = "black")

rug(birth.weight[race==2], ticksize = 0.02, side = 3, col = "lightgrey")

rug(cognitive.composite.score[race == 2], ticksize = 0.02, side = 4, col = "lightgrey")

a=birth.weight[race==1]
b=cognitive.composite.score[race==1]
o.a=a[order(a)]
o.b=b[order(a)]

lines(o.a,predict(loess(o.b~o.a,span=0.85,degree=1)),col= "black",lwd=3,lty=2)

c=birth.weight[race==2]
d=cognitive.composite.score[race==2]
o.c=c[order(c)]
o.d=d[order(c)]

lines(o.c,predict(loess(o.d~o.c,span=0.85,degree=1)),col= "darkgrey", lwd=3,lty=1)

temp <- legend("topleft", legend = c(" ", " "),

text.width = strwidth("Whites,n=654"),

lty = c(2,1),lwd=2,col=c("black","darkgrey"))

text(temp$rect$left + temp$rect$w,temp$text$y,

c("Whites,n=834", "Blacks,n=741"), pos=2)

abline(h=mean(norm.white[,12]),lty=3,lwd=3,col="black")

abline(h=mean(norm.black[,12]),lty=3,lwd=3,col="black")

fit1=loess(cognitive.composite.score[race==1]~birth.weight[race==1])
plot(birth.weight[race==1], predict(fit1), xlab="birth weight in grams", ylab="predicted cognitive composite score", type="n", pch=1, col="black", lwd=3, xlim=c(1474,6038), ylim=c(min(predict(fit1))-1, max(predict(fit1))+0.5))
fit2=loess(cognitive.composite.score[race==2]~birth.weight[race==2])
points(birth.weight[race==2], predict(fit2), pch=1, lwd=2, type="n", col="darkgrey", bg="darkgrey")
a=birth.weight[race==1]
b=cognitive.composite.score[race==1]
o.a=a[order(a)]
o.b=b[order(a)]
lines(o.a, predict(loess(o.b~o.a, span=0.85, degree=1)), col="black", lwd=4, lty=2)
c=birth.weight[race==2]
d=cognitive.composite.score[race==2]
o.c=c[order(c)]
o.d=d[order(c)]
lines(o.c, predict(loess(o.d~o.c, span=0.85, degree=1)), col="darkgrey", lwd=4, lty=1)
temp <- legend("topleft", legend = c(" ", " "),
  text.width = strwidth("Whites,n=654"),
  lty = c(2,1), lwd=2, col=c("black","darkgrey"))
text(temp$rect$left + temp$rect$w, temp$text$y, c("Whites,n=834", "Blacks,n=741"), pos=2)
abline(h=mean(norm.white[,12]), lty=3, lwd=4, col="black")
abline(h=mean(norm.black[,12]), lty=3, lwd=4, col="black")
3. Generalized propensity score analysis

3.1. Estimation of the balancing score

data$BIRTH[birth.weight <= 3200] <- 1
data$BIRTH[birth.weight > 4200] <- 3

model=lrm(BIRTH~
  sex+poly(ageby1997,2)+mothers.age+birth.order.to.mother+marital.status.of.mother+mothers.education+fathers.age+birth.order.to.father+private.insurance+MEDICAID+WIC+food.stamp+ADC.AFDC+sex*poly(ageby1997,2))

3.2. Construction of strata (stratification)

fit=predict.lrm(model, type="fitted")
propensity=fit[,1]
data=cbind(data, propensity)

cutpoint=as.data.frame(quantile(propensity, probs=seq(0,1,0.2)))
norm.white$strata2[propensity<=cutpoint[2,]]=1
norm.white$strata2[propensity>cutpoint[2,] & propensity<=cutpoint[3,]]=2
norm.white$strata2[propensity>cutpoint[3,] & propensity<=cutpoint[4,]]=3
norm.white$strata2[propensity>cutpoint[4,] & propensity<=cutpoint[5,]]=4
norm.white$strata2[propensity>cutpoint[5,]]=5
3.3. Histograms to show overlapping of the balancing score distributions

```r
par(mfrow=c(3,1))

hist(propensity[BIRTH==1],xlim=c(min(propensity),max(propensity)),xlab="",main="histogram of balancing score for birth weights <= 3200 grams")

hist(propensity[BIRTH==2],xlim=c(min(propensity),max(propensity)),xlab="",main="histogram of balancing score for birth weights > 3200 grams and <= 4200 grams")

hist(propensity[BIRTH==3],xlim=c(min(propensity),max(propensity)),xlab="",main="histogram of balancing score for birth weights > 4200 grams")
```

3.4. `cbal.psa` plot checking covariate balance

```r
data$treatment[birth.weight <= 3200] <- 1
data$treatment[birth.weight > 3200] <- 2

cbal.psa(covariates, treatment, propensity, strata=5)
```

3.5. Fitted surface for locally linear regression (span=0.85) of cognitive composite score on birth weight and balancing score

```r
birth=seq(min(birth.weight),max(birth.weight),len=25)

prop=seq(min(propensity),max(propensity),len=25)

mod.lo=loess(cognitive.composite.score~birth.weight+propensity,span=0.85,degree=1)

newdata=expand.grid(birth.weight=birth, propensity=prop)

fit=matrix(predict(mod.lo,newdata),25,25)

persp(birth,prop,fit,theta=60,phi=30,ticktype='detailed',xlab='birth weight in grams',ylab='balancing score', zlab='cognitive composite score',expand=2/3,shade=0.5)

persp(birth,prop,fit,theta=180,phi=30,ticktype='detailed',xlab='birth weight in grams',ylab='balancing score', zlab='cognitive composite score',expand=2/3,shade=0.5)
```