The black-white test score gap widens with age?

June 20, 2006

Abstract

We re-examine the theoretical concept of a production function for cognitive achievement, and argue that an indirect production function that depends upon the variables that constrain parents' choices is both more tractable from an econometric point of view, and more interesting from an economic point of view than is a direct production function that depends upon a detailed list of direct inputs such as number of books in the household. We estimate flexible econometric models of indirect production functions for two achievement measures from the Woodcock-Johnson Revised battery, using data from two waves of the Child Development Supplement to the PSID. Elasticities of achievement measures with respect to family income and parents' educational levels are positive and significant. Gaps between scores of black and white children narrow or remain constant as children grow older, a result that differs from previous findings in the literature. The elasticities of achievement scores with respect to family income are substantially higher for children of black families, and there are some notable differences in elasticities with respect to parents' educational levels.

Keywords: education; cognitive achievement; test score gaps

JEL Codes: D13; I20; J15; J24

1 Introduction

Much has been written on the development of children's cognitive achievement and its determinants[1]. This is hardly surprising, given the magnitude of society's investment in education, and the fundamental role of learning in the future course of a child's life. Both individually and collectively, few issues are equally important in terms of long term welfare.

Much research has looked at the gap between the test scores of black and white children. Some recent contributions that reference and summarize previous findings are Carneiro *et al.* (2005), Fryer and Levitt (2004), and Todd and Wolpin (2004). These papers find evidence that gaps widen as children grow older[2]. Fryer and Levitt find that controlling for covariates substantially explains gaps in scores for children entering kindergarten, but that they subsequently increase with age. Todd and Wolpin note that gaps in raw scores, without controlling for different levels of covariates, increase with age. They find that the magnitude of the gaps decreases when covariates are equalized, but they do not look at how covariate-adjusted gaps evolve as children grow older.

Using a new data set and a more flexible econometric model, we examine two measures of cognitive achievement, the Letter-Word (LW) and Applied Problems (AP) tests from the Woodcock-Johnson Revised Tests of Achievement (Woodcock and Johnson, 1989). We find that gaps in covariate-controlled LW and AP test scores do exist, but that they either narrow substantially (the LW score) or remain more or less constant (the AP score) over the ages 6-17. We calculate the elasticities of LW and AP scores with respect to parental education and family income, and look at how these elasticities evolve over the course of childhood. The scores of black children respond more strongly to changes in family income and mother's education than do the scores of white children. These findings may

[1] Havemann and Wolfe (1995) offer a general survey of the literature.

[2] The results concerning gaps that widen with age are Carneiro *et al.* (2005), page 7, Fryer and Levitt (2004), page 455, and Todd and Wolpin (2004) pp. 23-24 and Figures 3a and 3b.

shed some light on the possible sources of the existence of the gap in scores, and can inform policy intended to reduce it.

We begin by re-examining the theoretical underpinnings of the econometric approach to data on cognitive achievement, from the perspective of the production function literature (Ben-Porath, 1967; Leibowitz, 1974; Todd and Wolpin, 2003). This helps us to more carefully select which variables to include in the econometric model, and makes it clear that endogeneity of at least some variables is likely to be of concern. We also take into serious consideration the issue of the form of the cognitive achievement production function, in contrast to much of the literature which assumes a simple linear form. A simple linear model is strongly rejected by statistical tests, and a more richly parameterized model is needed to explain the data. The richness of the econometric model allows us to uncover dynamics in the evolution of test scores and elasticities that may be hidden by more restrictive econometric models that impose stronger forms of parameter constancy across groups.

As noted, we make use of a new data set. Much of the literature on the evolution of cognitive achievement in economics has made use of the National Longitudinal Survey of Youth (NLSY) and the associated Children of the NLSY (CNLSY) data (Bureau of Labor Statistics, 2001). Korenman *et al.* (1995), Neal and Johnson (1996), Blau (1999), Hansen *et al.* (2004), Todd and Wolpin (2004) and Carneiro *et al.* (2005) are examples of papers that draw at least some of their conclusions using this data. Other data sets have also been used, to a lesser extent. For example, Fryer and Levitt (2004) use the Early Childhood Longitudinal Survey (ECLS, National Center of Education Statistics, 2002). Our results are based upon the Child Development Supplement (CDS) to the Panel Study of Income Dynamics (PSID) data (Mainieri, 2006). This is a rich data set, and to our knowledge, this is the first paper by economists that uses it to estimate an educational achievement production function. It is important to check the robustness of findings across different data sets to control for possible biases due to the specific way survey information is gathered. Our findings show that the previous conclusions regarding widening gaps are called into question when this new data set is used.

The next section considers the educational production function and the econometric issues faced when attempting to estimate it. Section 3 discusses the data. Section 4 presents the econometric model in detail and gives results related to the choice of the final specification and the estimation method. Section 5 presents the principal findings, and Section 6 gives conclusions and discusses directions for future work.

2 The production of cognitive achievement

Children's performance on achievement tests may be modeled as the output of a production function, the inputs of which are determined by the family, the school environment, and other factors. Todd and Wolpin (2003, henceforth, TW) provide a detailed discussion, which we build upon. We assume that choices are made in discrete time. Notationally, let a vector (indicated by lower case) indexed by t indicate the current period values of a set of variables, and let a matrix (indicated by upper case) represent the entire history up to the time of the index. For example, $A_{t-1} = (a_0, a_1, ..., a_{t-1})$. Let q_t be a child's achievement at time t. There is a time-constant genetic endowment, μ, which is not directly observable. We postulate that current achievement q_t depends on the endowment μ, as well as on current and lagged inputs, from the time the child is born ($t = 0$) to the present.

Some inputs are chosen by the parents, either directly or indirectly, and others are determined beyond the control of the parents. When discussing the theoretical model, we will refer to parentally chosen inputs as endogenous, and externally chosen inputs as exogenous. The daily time that parents read to a child or the number of books in the household are examples of directly chosen endogenous inputs. One can easily think of a great number of such inputs. The number of hours a child spends in regular school classes is in some cases indirectly chosen by the family, through the choice of the school the child attends. The occurrence of a serious illness that affects cognitive achievement or government-mandated characteristics of curricula are examples of exogenous inputs.

Another issue is the issue of the observability of inputs. Some inputs, both endogenous and exogenous, are not observable, at least to the econometrician. Surveys can gather only a limited amount of information, and the number of inputs that can affect cognitive achievement is no doubt large. Parents are also likely to find reporting information about the way they raise their children to be a sensitive topic, and this may induce substantial measurement error. We define $X_t^o = \left(x_0^o, x_1^o, ..., x_t^o\right)$ to be the matrix that holds the complete history of the observable endogenous inputs up to time t. Likewise, X_t^u is the corresponding matrix of unobservable endogenous inputs. We define Z_t^o and Z_t^u to be the complete histories of the observable and unobservable exogenous inputs, respectively.

The direct achievement production function is

$$q_t = f(X_t^o, X_t^u, Z_t^o, Z_t^u, \mu, \epsilon_t) \tag{1}$$

which we take to be the true technology that generates achievement. This equation is conceptually similar to TW's equation 3, which they referred to as the cumulative specification. Here, we explicitly recognize that many inputs, both endogenous and exogenous, may not be observed, and we interpret the error, ϵ_t, as due to both possible measurement error and to random variations in performance across individuals. TW deal with methods that can be used to estimate the model when the endowment μ is not observed, assuming that all unobservable inputs (X_t^u, Z_t^u) are uncorrelated with the observed inputs X_t^o, Z_t^o. We find this assumption to be quite implausible. Many of the inputs that are not observed by the econometrician will be endogenous, chosen by the parents jointly with the observed endogenous inputs. Since the two sets of inputs, X_t^o and X_t^u, are chosen by the parents, simultaneously and subject to the same constraints, it is difficult to believe that they will be uncorrelated with one another. Measurement error due to the use of possibly sensitive information reported by parents is likely to reinforce an existing correlation.

We assume that parents choose the current period values of endogenous inputs to maximize some form of discounted expected family utility function. This maximization will be subject to a budget constraint. Expectations will be formed consistently with the currently available information (we define timing to be such that current period values of exogenous variables are known when current period endogenous variables are chosen), and will also depend upon the information processing capabilities of the parents, which we refer to as the parents' endowments, and denote by the vector μ_P. The histories of the inputs, $X_{t-1}^o, X_{t-1}^u, Z_t^o, Z_t^u$, will affect the choices parents make regarding the current period endogenous inputs, x_t^o, x_t^u. To simplify the exposition and notation, we make the simplifying assumption that the family's income (M) and parents' endowments (μ_P) are all known at the outset, and are constant over time. It is important to recognize that parents learn about their children's endowments over time, and that they will adjust input levels accordingly. For simplicity, assume that the history of achievement, Q_{t-1}, is the full set of information that informs parents' learning about their child's endowment. Then the optimal levels of the endogenous inputs in each period t will be the vector-valued functions $x_t^o(M, \mu_p, Z_t^o, Z_t^u, Q_{t-1})$ and $x_t^u(M, \mu_p, Z_t^o, Z_t^u, Q_{t-1})$. These optimal solutions do not depend upon the previous histories of the endogenous variables, because those variables were in turn chosen optimally in the past, as functions of the same arguments, with shorter histories. Since these histories are already subsumed in the longer histories that are the arguments of the current period optimal levels, there is no need to write them again as additional arguments. If the optimal levels of the endogenous inputs are substituted (recursively) back into the direct production function of equation 1, we obtain an indirect production function that depends upon the parents' endowments, family income, the history of exogenous factors, and the history of achievement test scores:

$$\begin{aligned} q_t &= f(X_t^o(M,\mu_p,Z_t^o,Z_t^u,Q_{t-1}),X_t^u(M,\mu_p,Z_t^o,Z_t^u,Q_{t-1}),Z_t^o,Z_t^u,\mu,\epsilon_t) \\ &\equiv g(M,\mu_p,Z_t^o,Z_t^u,\mu,Q_{t-1},\epsilon_t) \end{aligned} \quad (2)$$

The indirect production function depends on previous achievement, as does the well-known value-added model (Todd and Wolpin, 2003).

From the point of view of econometric estimation, the indirect production function has several advantages, compared to the direct production function:

First, it depends upon many fewer arguments. The endogenous inputs of the direct production function, the vectors X_t^o and X_t^u, include the full histories of observed and unobserved items. The sheer number of direct inputs makes their inclusion in a model problematic, due to both due to problems of missing data and to severe collinearity between the variables that can be included. The use of any particular index function to reduce the number of inputs to include will be debatable. The number of variables in the indirect production function is much smaller, since the endogenous inputs disappear, and the number of added variables, related to the restrictions to utility maximization, is small in comparison.

Second, there are less severe problems of endogeneity (in the econometric sense) in the case of the indirect production function. For the direct production function, the fact that observable and unobservable family-chosen inputs are chosen jointly in response to common factors implies that observed inputs are almost certainly correlated with unobserved inputs, as was noted above. If the unobserved inputs go into the econometric error term, there will be a problem of econometric endogeneity in the case of the included family-chosen direct inputs. The fact that both current and lagged achievement depend upon the child's endowment μ, does lead to concern about the endogeneity of Q_{t-1}, as is noted by Todd and Wolpin (2003, pp. F20-F22). We believe that this is less problematic than is the endemic endogeneity of the family-chosen inputs in a model of the direct production function. Well-known econometric methods can be used to deal with the possible endogeneity of the history of achievement.

Third, the indirect production function provides a simpler, clearer framework within which to analyze policies directed to improve children's acquisition of cognitive achievement. Policy might affect family income in the short run, and parents' endowments in the longer run, and the indirect function depends upon these variables. Policy changes that affect exogenous factors such as school characteristics can also be analyzed using the indirect production function. The effect of policies on the inputs of the direct function would be many, diverse, and difficult to predict. The direct function requires information at a level of detail that is difficult to supply during the estimation phase, and it supplies information that is difficult to interpret and use at the stage of policy formation.

Finally, though knowledge of the direct production function is no doubt interesting, it is not fundamentally an economic issue. Other areas of science, and methods other than econometric inference, are likely to tell us more about the detailed process through which children's cognitive achievement changes in response to direct inputs. If experimental or observational methods were used, the issue of endogeneity of inputs could be dealt with more effectively and convincingly than by applying instrumental variables estimation methods to survey data.

3 Data

As noted in the introduction, we use two data sets linked to one another. The oldest and most well-known is the Panel Study for Income Dynamics (PSID). The PSID is a longitudinal survey containing socio-economic information of a representative sample of families of the U.S., with data at the level of the individual. The second source of data is the Child Development Supplement (CDS) to the PSID (Mainieri, 2006). The CDS contains detailed information about cognitive achievement, health status, time use at home and at school, and information about schools, for children from PSID families. There are two waves of CDS data, CDS-I and CDS-II, the first gathered in 1997 and the second in 2002-2003. The CDS-II wave is based on interviews of 91% of the families that participated in CDS-I. Combining the PSID and the two waves of the CDS, we obtain time series data on several measures of children's cognitive achievement, as well as covariates such as family income, parental education, race, time use, *etc.*

As measures of cognitive achievement, we use the Letter-Word (LW) and Applied Problems (AP) subscales of the Woodcock-Johnson Revised Tests of Achievement (Woodcock and Johnson, 1989) as measures of verbal and mathematical achievement, respectively. Both scores range from 0 to 60. We use the scores from both CDS waves, to obtain the current and historical scores, q_t and Q_t, of the previous section. We restrict our sample to children who lived with the same two parents in both CDS waves. Dealing with heterogeneous and changing family structures is left for future work. We use the years of education of the mother and the father as measures of the parents' endowments (μ_p). Our measure of income (M) is family income per family member, at the time of the second wave. We also explored using average family income across the two CDS waves and a more flexible specification where total family income appears as a regressor and the number of siblings as another. Such specifications give results that are very similar to those we report, using our chosen measure of income. The observable exogenous variables (the Z_t^o of the previous section) are race, sex, and age of the child. For the variables we use, the available sample size is 983 children.

In the available data, the only usable characteristic of schools attended by the children in the sample were the average class size, since missing data rendered the sample too small if other measures were used. Average class size had no significant impact on the results, so we do not use this variable in the results we report.

We also did some exploratory work with the time-use diary data that the CDS contains, looking at daily hours parents' spent actively interacting with their children, and total amount of time children spent watching television. These variables may be thought of as direct endogenous inputs to the achievement production function. Following our theoretical presentation of the last section, we would argue that a proper indirect production function should not include direct endogenous inputs as arguments. If a direct input is included, it should not contribute significantly to the explanatory power of a model of the indirect production function, supposing that the income and parental endowment variables introduced though the substitution of the optimal solutions into the direct model are well measured, and the functional form of the model is adequate. When the time-use diary variables were included in our model, they did not have any significant impact and their exclusion was not rejected by formal statistical tests. Based upon the above considerations, we interpret this insignificance as evidence in favor of the specification of our model, which is explained in detail in the next section.

Work by other authors has included parentally-chosen inputs such as number of books in the household, and they have appeared as significant (for example, Todd and Wolpin (2004); Fryer and Levitt (2004)). However, Todd and Wolpin's paper is an explicit attempt to model a direct production function, and they purposely do not include any measure of family income or parental education in their model. It is not surprising that a small number of direct inputs appear significantly when the variables that would appear after substituting in the optimal solutions are excluded from the model. Fryer and Levitt's model might be interpreted as a mixed direct/indirect production function, since it has both direct inputs and variables that enter though constraints. They make use of a single socioeconomic status index, which is a composite of family income, parental education and other factors. Perhaps the significance of the direct input might be in part due to the inability of this single index to adequately account for the separate effects of income and parental endowments.

4 The econometric model

Much of the econometric literature on estimation of the relationship between educational achievement and conditioning factors assumes a simple linear relationship between the inputs and the output, and issues of functional form and possible nonlinearities have only seldom been addressed (Baker, 2001). It is important to recognize that a simple linear model imposes strong restrictions on the production function. The marginal effects of all variables are constant, and elasticities cannot freely vary, even at a single arbitrary point of evaluation. Our econometric model of the indirect production function allows for nonlinear and cross-variable effects, along the lines of the flexible functional form literature (Caves and Christensen, 1980). We find that nonlinearities and interaction effects are important, and are overlooked when one estimates a simple linear model of the type commonly encountered.

Our data is a cross section of individuals, each observed in the second wave of the CDS at a specific age between 6 and 17 years old. Because we model only the LW and AP scores at the time of the second wave of the CDS, we drop the t subscript that was used previously in the general treatment. Since we do not assume that achievement is constant with respect to age, we use age (AGE) as an observable exogenous variable (one of the components of Z^o). The other observable exogenous variables we use are dummy variables for gender (SEX) and ethnic background (the groupings are black (B), white (W) and other, which is the default, absorbed in the constant).[3] As measurements of the parents' endowments (μ_P), we use the mother's (ME) and the father's (FE) years of education. Income (M) is measured as total family income divided by the number of family members. We only have data on a single lagged achievement score, so we assume that the entire history of achievement Q_{-1} can be approximated by the single lag, q_{-1}. We collect these eight variables in the vector $\mathbf{x} = (ME, FE, q_{-1}, AGE, M, SEX, B, W)$. We assume that the indirect production function is additively separable in the observable and unobservable arguments. For a representative individual, we write

$$q \quad = \quad q(\mathbf{x}) + \eta(Z^u, \epsilon)$$

We define $\eta \equiv \eta(Z^u, \epsilon)$, and treat it as an econometric error term. We specify a quadratic parametric model for $q(\mathbf{x})$, so our econometric specification is

$$q \quad = \quad \alpha + \mathbf{x}'\beta + \frac{1}{2}\mathbf{x}'\Gamma\mathbf{x} + \eta \tag{3}$$

To simplify notation, we note that the quadratic model may be written, with appropriate definitions, as

$$\alpha + \mathbf{x}'\beta + \frac{1}{2}\mathbf{x}'\Gamma\mathbf{x} = \mathbf{z}(\mathbf{x})'\theta,$$

so

$$q \quad = \quad \mathbf{z}'\theta + \eta. \tag{4}$$

[3]The available sample did not allow for more detailed treatment of ethnic groups.

The vector-valued function $\mathbf{z} = \mathbf{z}(\mathbf{x})$ contains the original vector $\mathbf{x}$ as well as a constant term and the squares and cross-products of the elements of $\mathbf{x}$, and the vector θ contains all the free parameters in α,β and Γ. To identify the Γ matrix, we restrict it to be symmetric, and the coefficients of the squared dummy variables and interactions between ethnic group dummies are restricted to be zero.

We anticipate that there may be endogeneity of q_{-1}, as is discussed above, but we assume that the other components of $\mathbf{x}$ are exogenous. In our quadratic model, the possible endogeneity of q_{-1} spreads to a number of the components of $\mathbf{z} = \mathbf{z}(\mathbf{x})$. To address this, we perform generalized instrumental variables (GIV) estimation. Our instruments are the elements of $\mathbf{z}$ that do not involve q_{-1}, as well as the elements of $\mathbf{z}_{-1}$ (obtained from the first wave (CDS-I) data) that do not involve q_{-2}. We also estimate using OLS.

Plots of the GIV and OLS residuals strongly suggest that the errors are heteroscedastic. White's test for homoscedasticity (White, 1980), using $\mathbf{z}$ as the variables that explain the squared GIV residual in the artificial regression, strongly rejects homoscedasticity, for both the LW and AP scores (Tables 1 and 2, row 1). To improve efficiency in estimation, we henceforth use a partial correction for heteroscedasticity. The variance of the error of the error is specified as $V(\eta) = \exp(\alpha_1 + \alpha_2 \log AGE)$ (separate models are used for LW and AP scores). We continue to apply a heteroscedastic-consistent covariance matrix estimator, to allow for residual heteroscedasticity that is not captured by this simple model of the error variance.

The theoretical development of the last section suggests that the regressors in $\mathbf{z}$ that depend upon q_{-1} are likely to be endogenous. To test for exogeneity, the standard Hausman test requires that one of the two estimators use to define the vector of contrasts be fully efficient under the null hypothesis of exogeneity. In our case, the least squares estimator is unlikely to be fully efficient. Without the heteroscedasticity correction it is certainly inefficient, and with the simple correction, any remaining unmodelled heteroscedasticity or nonnormality of the errors also imply that the least squares estimator is inefficient. This would cause the standard Hausman test to become invalid. Creel (2004) develops a modified version of the Hausman test that is valid when neither of the two estimators that are contrasted is efficient.

Tables 1 and 2, rows 2 and 3, present the standard and modified Hausman test statistics for the null hypothesis of exogeneity of all regressors, without the heteroscedasticity correction. In the case of the LW score (Table 1), the standard Hausman test without the GLS correction (row 2) suggests rejection of exogeneity. This result is of doubtful validity, since the Hausman test is invalid in the presence of heteroscedasticity, which almost certainly exists, given the test results reported in row 1 of the Tables. The modified test (row 3), which is valid in the presence of heteroscedasticity does not reject at any conventional significance level. Lines 4 and 5 present the standard and modified Hausman tests, using the partial GLS modelling of the variance of the error term. We see that neither test rejects exogeneity at the 10% significance level. For the standard Hausman test, the reversal of the conclusion depending upon whether or not a GLS correction is done shows the dangers of relying on this test when using inefficient estimators. The modified test gives the same result with or without the GLS correction. In the case of the AP score (Table 2, rows 2-5), neither version of the test rejects at conventional significance levels, regardless of whether or not the GLS correction is used. Overall, when a valid test is used, exogeneity is not rejected. Given this, and the greater efficiency of least squares compared to GIV, the results we present below are based upon least squares estimation using the simple heteroscedasticity correction.

The full quadratic specification, with only the necessary restrictions for identification, contains 41 free parameters. It is possible that this level of flexibility is not needed to successfully capture the features of the data. We tested some parameter restrictions in an effort to obtain a more parsimonious model. Tables 1 and 2, rows 6-9 reports results for Wald tests of several hypotheses. The hypotheses tested, and the corresponding rows the Tables are:

- (row 6) The model can be reduced to a simple linear specification, without interaction or nonlinearities in the variables. This hypothesis implies that Γ in equation 3 is a matrix of zeros. This hypothesis is strongly ($p < 0.001$) rejected for both the LW and AP scores.

- (row 7) The three racial groupings (black, white, other) can be pooled together. This hypothesis is strongly ($p < 0.001$) rejected for both LW and AP.

- (row 8) The black and other racial groups can be pooled. This hypothesis is rejected at the 10% significance level for both AP and LW.

- (row 9) Boys and girls can be pooled together. This hypothesis is rejected at the 10% level for both the LW and AP scores.

All of the hypotheses tested are rejected fairly convincingly, at at least the 10% level. Thus, we do not impose any restrictions upon the general quadratic specification (equation 3). The data exhibit nonlinearities and interactions that cannot be captured by a simple linear model.

5 Results

Our econometric model has a large number of parameters (41). Since the model includes nonlinearities and interactions between variables, the individual parameters do not have an interesting interpretation, and for this reason, we do not present parameter estimates[4]. Instead, we present plots of predicted LW and

[4]Parameter estimates are available upon request.

AP scores, and elasticities of both predicted scores with respect to some of the explanatory variables, along with 2 standard error bars. Since estimated elasticities are nonlinear functions of the estimated parameters of the models, the delta method[5] was used to calculate the estimated standard errors of the elasticities.

Figures 1 and 2 present elasticities and fits, by the age of the child. In all cases we evaluate the elasticity or fit at the overall sample means (except income, where we use the median) of the explanatory variables, given the age of the child. It is important to calculate the evaluation point conditional on age, since the distribution of the regressors is not independent of age. Lagged score is strongly dependent on age, and family income and parental education are weakly dependent[6]. These plots average over the race and sex dummy variables, so they represent an "average" child. Later we look at more specific results.

Looking first at the plots for the LW score (Figure 1), in panel (a) we see the predicted LW score as a function of age. It is strongly increasing between ages 7 to 14, after which it levels off and even falls at age 17. In panel (b) we see that the elasticity with respect to lagged score is positive, significantly different from zero at all ages, and trending upwards. Note that this elasticity would be equal to one for a person who had arrived to a stable level of LW score, since the lagged score would be equal to current score, and a one percent higher lagged score would imply a one percent higher current score. With this in mind, we see that even at age 17, the elasticity is still significantly less than one. The fact that the elasticity is different from one at age 17 means that the LW score is still malleable between ages 12 to 17, which is confirmed by the results in panel (a) that show that LW score is not constant over these ages.

In panels (c) and (d), we see the elasticities of the LW score with respect to the mother's and father's years of education, respectively. The elasticity with respect to mother's educational level is positive and significantly different from

[5]See, for example, Cameron and Trivedi (2005).

[6]Older children have older parents, who have higher incomes and more education, on average.

zero at all ages. It exhibits a U shape, with a minimum at 13 years of age. The elasticity with respect to father's educational level is smaller in magnitude for almost all ages, and it is only barely significantly different from zero between the ages 7 to 13. For other ages, it is not significantly different from zero. The overall trend is downward. We see that the impact of parents' educational level is larger early in life. It is important to remember that early gains have permanent effects, since they are incorporated into the lagged score variable, which positively affects current score (see panel (b)).

Panel (e) shows the elasticity of LW with respect to family income. This elasticity is positive and significant up to age 11, at which point it becomes insignificantly different from zero. Nevertheless, the early effects of larger family income will continue to have a positive long run effect on predicted LW, since early gains are transmitted forward by the lagged score variable.

Panel (f) shows the elasticity of LW with respect to the child's age, plotted by age. This panel requires a careful interpretation, since the partial derivative of score with respect to age, used to calculate the elasticity, holds the lagged score constant. However, as we have seen in panel (a), scores, and consequently lagged scores too, generally rise with age. Lagged score is remaining constant as age increases could be interpreted as a case of learning difficulties. This would imply a lower predicted score, and a negative elasticity, as is observed in the plot.

Next, we turn to discussion of Figure 2, which plots the fitted AP score and the elasticities of AP with respect to the explanatory variables. Panel (a) shows fitted AP, by age. The fitted value increases with age, up to 14 years of age, at which point it levels off and then declines. This is similar to the pattern observed for LW, in Figure 1, panel (a). The test scores of 17 year-olds drop considerably for some reason. We can only speculate as to the explanation for this.

In panel (b), the elasticity of AP with respect to lagged score is positive, significant, and trending upward. At age 17, the elasticity is 0.76. The comparable elasticity for LW is only 0.60. This suggests that LW scores remain more malleable at age 17 than are AP scores.

Panels (c) and (d) show the elasticities of AP with respect to mother's and father's educational levels. The elasticity with respect to the mother's educational level is positive and everywhere significant, and the elasticity with respect to the father's educational level is significantly different from zero over ages 9-14, during which period the magnitude of the elasticity exceeds that of the mother's educational level. Beyond age 14, the U shape of the elasticity with respect to mother's education is in the rising portion. The educational level of both parents is important. The importance of the father's educational level clear in the case of AP, and weak in the case of LW. For the mother's educational level, the effect is clear in the cases of both scores.

Panel (e) of Figure 2 shows the elasticity of AP with respect to family income. The shape is very similar that observed in panel (e) of Figure 1, for LW, except that the magnitude of the elasticity is higher in the case of AP. Income has a significant positive effect in early years, up to age 12. Afterwards, the elasticity is not significantly different from zero. However, it is important to remember that the early impact is transmitted forward in the form of a higher value for the lagged score variable, which positively affects AP (panel (b)).

Panel (f) shows the elasticity with respect to age. The interpretation is the same as is given above for the case of LW. The elasticity is again negative and significant at all ages.

The results reported up to now are for an "average child", with all conditioning variables, including racial and gender dummies, set to their sample mean values. There exists a considerable literature that has analyzed differences in test scores between racial groups, with most of the focus on blacks and whites, due to the larger samples that are available. Fryer and Levitt (2004) give a detailed discussion and many references. Fryer and Levitt find that a gap in raw test scores exists, but that for children entering kindergarten the gap is not significantly different from zero after controlling for factors such as the socioeconomic background of the family and the number of books in the child's home. They find that the gap widens with age, however, even after controlling for covariates (Fryer and Levitt, 2004, pg. 455). Our data allows us to examine this prediction.

In Tables 3 and 4 we give the descriptive statistics for raw LW and AP scores, by age and race, without controlling for any covariates. We see that there is a gap in raw scores, both for LW and AP. The gap in LW scores is somewhat variable as a function of age. Overall the average gap is about 3.5-4 points. The gap in raw AP scores increases up to age 11, after which it is more or less constant at about 5.5 points (a little less than 1 standard deviation). The pattern that emerges from these summary statistics is a bit difficult to interpret, due to their variability, which is a result of the small sample sizes in some of the cells (for example, the sample contains no six year old black children). But it is quite clear that a gap in raw scores exists, and that it is larger for teenagers than for 8 year olds.

Turning to the results of the econometric model that allows for controlling for the effects of covariates, we now look at plots similar to those given in Figures 1 and 2, but looking specifically at blacks and whites, by setting the racial dummy variables accordingly. The other explanatory variables remain set at their overall sample means (except income, which is set to the sample median). For clarity, we do not plot standard error bars. A good idea of the precision of the estimates may be gotten from Figures 1 2.

Figure 3 shows the fitted LW, and the elasticity of LW with respect to conditioning variables, by age of child, for blacks and whites. In panel (a), we see that a clear black-white gap exists, but it is markedly smaller than the raw gap that does not take into account the levels of covariates, in Table 3. It also declines substantially as children grow older, narrowing from 2.5-3 points down to roughly 1 point. The LW score of blacks is considerably more elastic with respect to the mother's educational level (panel c) and income (panel e) than is the LW score of whites. Race does not appear to be a significant factor in the elasticity of LW with respect to the father's educational level (panel d). For younger children, both parents' educational levels are important, with the mother's educational level of particular importance for blacks. As children grow older, the mother's educational level is of more importance than is the father's, for both races. It appears that the impact of the mother's educational level gains importance as children near high school age.

The LW score of children of black families is more elastic with respect to income than is the case of children of white families, for children of all ages (panel e), though income elasticities are low in both cases. Again, these results need to be interpreted with some care. LW responds positively to income, at all ages, for both blacks and whites. For this reason, children of higher income families will have somewhat higher scores, and they will also have somewhat higher values of the lagged LW score. However, the plots in Figure 3 set the value of the lagged score to the overall sample mean. This ignores the impact on LW of higher income in the past. The overall effect of a permanent increase in family income on the evolution of a child's LW score over ages 6-17 would be larger than Figure 3 implies, due to the dynamic effect that is transmitted forward though the lagged score. Early gains persist, as is reflected in panel (b) of the Figure, where the elasticity of current LW with respect to lagged LW is seen to be positive at all ages.

Figure 4 shows fitted AP scores and elasticities for blacks and whites. Panel (a) shows predicted AP. There is a gap that remains more or less constant at 2-2.5 points - in this case it does not narrow with age. However, it is clearly not increasing with age, as Fryer and Levitt (2004) predicted, based upon extrapolating their results for younger children. It is also less than half the gap observed in raw AP scores, as can be seen comparing the 2-2.5 point gap in Figure **??** with the 5-6 point gap in raw scores, in Table 4.

Panel (c) shows that the elasticity of the AP score with respect to the mother's educational level is larger for black children than for white children, similarly to what was found in the case of the LW score. In contrast, panel (d) shows that the elasticity of AP score with respect to the father's educational level is larger for white children. For white children, the impact of the father's years of education is more important than the mother's years of education, between 7 to 16 years of age. For black children, the impact of the mother's years of education is always larger that the impact of the father's years of education. Panel (e) shows that the elasticity of AP score with respect to income is 3 to 4 times larger for blacks than for whites. This is a similar result to the case of the LW score, but in the present case the difference is even more dramatic.

6 Conclusions

Like the previous literature, we have found that there are substantial gaps in raw achievement scores between black and white children (Tables 3 and 4). The gaps narrow considerably when an econometric model is used to control for the effect of covariates (Figures 3 and 4). After accounting for covariates, the gap narrows with age in the case of the LW (verbal/reading) score and remains constant in the case of the AP (math) score. As noted in the introduction to this paper, there is a fairly large amount of literature that documents gaps that widen with age, even after accounting for covariates. Much of this work uses the NLSY and CNLSY data. Also, many of these studies use econometric models that impose a stronger degree of parameter constancy across sample subgroups than does our model. We find that such restrictions on our econometric model are rejected by formal statistical tests. Our work has two distinguishing factors: the CDS data that we use, and a more flexible econometric model than those of most similar studies. To determine whether our result on gaps that do not widen is due to the new data or to the more flexible econometric model, it would be interesting to estimate an econometric model similar to ours using other data sets.

Another result of our work is the observation that elasticities of achievement scores with respect to family income and parents' educational levels differ across black and white children. The fact that elasticities differ, in some cases quite markedly, may be indicative of certain cultural differences between black and white families. Such factors could explain at least in part the simple existence of the observed gaps. Differences in elasticities could conceivably be exploited by policies intended to address the score gap, though the political feasibility of such measures would be questionable.

An important factor that we have not taken into account in this work are possible differences in the qualities of the schools attended by children. The basic CDS data is not particularly well suited to the use of school characteristics, since the usable sample size is greatly diminished when such variables are included in the analysis. Given that the effect of including covariates in the econometric model is to contribute to the explanation of the gap in test scores, we expect that the most important effect of omitting school characteristics is to leave a larger residual gap than would be the case if such variables could be included. We hope to use other data sources to extend our work in this direction in the future.

References

[1] Baker, B. D. ”Can flexible non-linear modeling tell us anything new about educational productivity?”, *Economics of Education Review*, 2001, *20*, pp. 81-92.

[2] Ben-Porath, Y. ”The production of Human Capital and the Live-cycle of Earnings,” *Journal of Political Economy,* 1967, *75*(4), pp. 352-365.

[3] Blau, D. M. ”The Effect of Income on Child Development.” *The Review of Economics and Statistics*, 1999, *81*, pp. 261-276.

[4] Carneiro, P., Heckman, J.J. and Masterov, D.V. ”Labor Market Discriminations and Racial Differences in Premarket Factors.” *Journal of Law and Economics,* 2005, *48*(1), pp.1-49.

[5] Cameron, A.C. and Trivedi, P.K. *Microeconometrics*, 2005, Cambridge University Press.

[6] Caves, D.W. and Christensen, L.R. ”Global Properties of Flexible Functional Forms.” *American Economic Review,* 1980, *70*, pp. 422- 432.

[7] Creel, M. ”Modified Haussman Tests for Inefficient Estimators.” *Applied Economics*, 2004, *36*(21), pp. 237-76.

[8] Fryer, Jr., R. G. and Levitt, S,D. ”Understanding the Black-White Test Score Gap in the First Two Years of School.” *The Review of Economics and Statistics*, 2004, *86*, pp. 447-464.

[9] Hansen, K. T., Heckman, J.J and Mullen, K.J. ”The Effect of Schooling and Ability on Achievement Test Scores”. *Journal of Econometrics*, 2004, *131*, pp. 39-98.

[10] Haveman, R. and Wolfe, B. ”The Determinants of Children’s Attainments: A Review of Methods and Findings.” *Journal of Economic Literature*, 1995, *33*, pp. 1829-1878.

[11] Korenman, S. D., Miller, J. and Sjaastad, J. E. Long-Term Poverty and Child Development in the United States: Results from the NLSY, Children and Youth Services Review 17,1/2 (Special Issue 1995): pp 127-155.

[12] Leibowitz, A. ”Education and Home Production”, *American Economic Review,* 1974, 64, pp 243-250.

[13] Mainieri, T. ”The Panel Study of Income Dynamics Child Development Supplement, User Guide for CDS-II.” Institute for Social Research, University of Michigan, 2006, `http://psidonline.isr.umich.edu/CDS/`.

[14] National Center of Educational Statistics, User’s Guide to the Kindergarten-First Grade Public Use Data File, NCES-2002-149 (2002).

[15] Neal, Derek A. and Johnson, William R. ”The Role of Premarket Factors in Black-White Wage Differences.” *Journal of Political Economy*, 1996, *104*(5), pp. 869-95

[16] Todd, P. E. and Wolpin, K. I. ”On the Specification and Estimation of the Production Function for Cognitive Achievement.” *The Economic Journal*, 2003, *113*, F3-F33

[17] Todd, P. E. and Wolpin, K. I. ”The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps. PIER Working Paper 04-019, 2004.

[18] White, H. ”A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica,* 1980*, 48*(4), pp. 817-838.

[19] Woodcock, R. W., and Johnson, M. B. ”Tests of Achievement, Standard Battery [Form B]”. Chicago, IL: Riverside Publishing (Houghton Mifflin), 1989.

Tables

Table 1: Hypothesis test results, LW score

Row	Null hypothesis	Test	Distribution under H_0	Test value	p-value
1	Homosced.	White	$\chi^2(40)$	251.82	0.000
2	Exog. (no GLS)	Hausman	nominally, $\chi^2(9)$	31.50	0.000
3	Exog. (no GLS)	Hausman (mod.)	$\chi^2(9)$	8.31	0.503
4	Exog. (GLS)	Hausman	nominally, $\chi^2(9)$	6.37	0.703
5	Exog. (GLS)	Hausman (mod.)	$\chi^2(9)$	13.65	0.135
6	Simple linear model	Wald	$\chi^2(32)$	152.41	0.000
7	Black=White=Other	Wald	$\chi^2(14)$	47.00	0.000
8	Black=Other	Wald	$\chi^2(7)$	12.63	0.082
7	boys=girls	Wald	$\chi^2(8)$	14.26	0.075

Table 2: Hypothesis test results, AP score

Row	Null hypothesis	Test	Distribution under H_0	Test value	p-value
1	Homosced.	White	$\chi^2(40)$	151.35	0.000
2	Exog. (no GLS)	Hausman	nominally, $\chi^2(9)$	4.94	0.839
3	Exog. (no GLS)	Hausman (mod.)	$\chi^2(9)$	8.29	0.505
4	Exog. (GLS)	Hausman	nominally, $\chi^2(9)$	6.85	0.653
5	Exog. (GLS)	Hausman (mod.)	$\chi^2(9)$	10.44	0.316
4	Simple linear model	Wald	$\chi^2(32)$	87.72	0.000
5	Black=White=Other	Wald	$\chi^2(14)$	66.09	0.000
6	Black=Other	Wald	$\chi^2(7)$	12.20	0.094
7	boys=girls	Wald	$\chi^2(8)$	13.79	0.088

Table 3: LW score by age and race

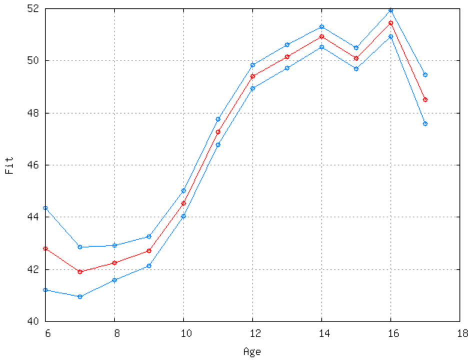
age	mean (overall)	mean (blacks)	mean (whites)	gap	gap/se
6.000	40.333	NA	40.250	NA	NA
7.000	39.250	36.474	40.106	-3.633	-0.577
8.000	43.655	41.926	44.623	-2.697	-0.568
9.000	44.011	42.053	45.527	-3.475	-0.598
10.000	46.170	44.571	47.613	-3.041	-0.568
11.000	46.385	43.769	47.629	-3.860	-0.599
12.000	48.524	45.680	49.746	-4.066	-0.807
13.000	49.450	47.174	50.470	-3.296	-0.642
14.000	50.560	49.222	51.388	-2.166	-0.471
15.000	49.770	46.324	51.782	-5.458	-1.091
16.000	51.986	51.000	52.362	-1.362	-0.429
17.000	48.688	44.750	52.286	-7.536	-1.097

Table 4: AP score by age and race

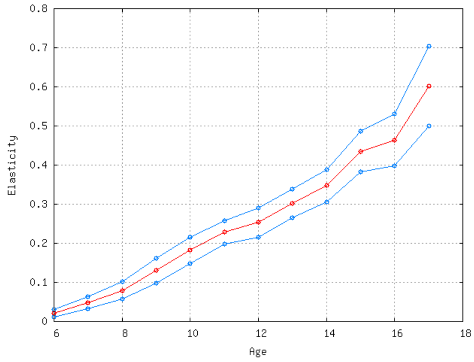
age	mean (overall)	mean (blacks)	mean (whites)	gap	gap/se
6.000	31.500	NA	32.500	NA	NA
7.000	32.486	29.158	34.000	-4.842	-1.033
8.000	35.370	33.111	36.416	-3.304	-0.793
9.000	36.682	34.105	38.127	-4.022	-0.922
10.000	39.000	36.333	40.452	-4.118	-0.812
11.000	40.010	36.269	41.919	-5.650	-0.949
12.000	40.981	37.240	42.716	-5.476	-0.985
13.000	43.290	39.348	45.227	-5.879	-0.923
14.000	44.211	39.630	46.522	-6.893	-0.977
15.000	42.880	39.382	45.182	-5.799	-0.870
16.000	44.757	41.333	46.447	-5.113	-0.803
17.000	42.750	39.625	45.571	-5.946	-0.915

Figures

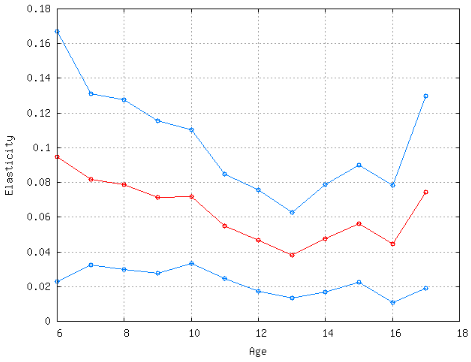
Figure 1: LW score, Fit and Elasticities, by age of child, with 2 standard error bars



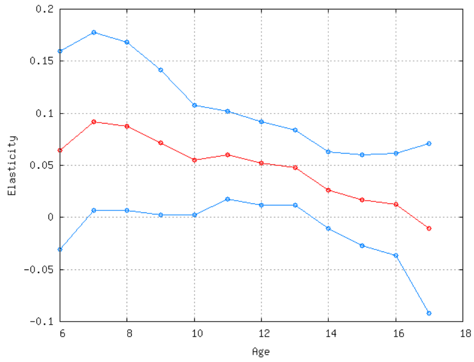
(a) Fit



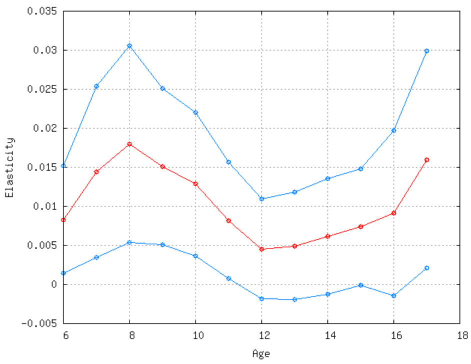
(b) Lagged score



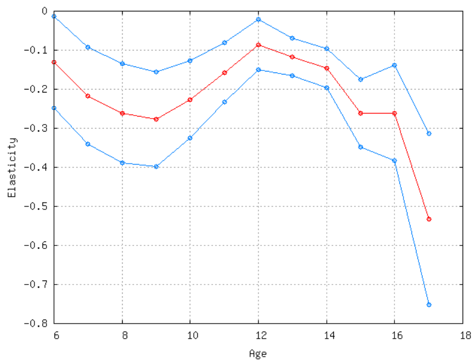
(c) Mother's education



(d) Father's education

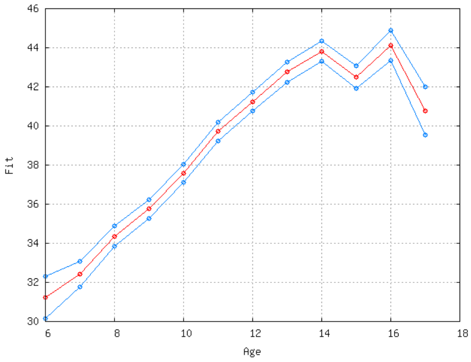


(e) Income

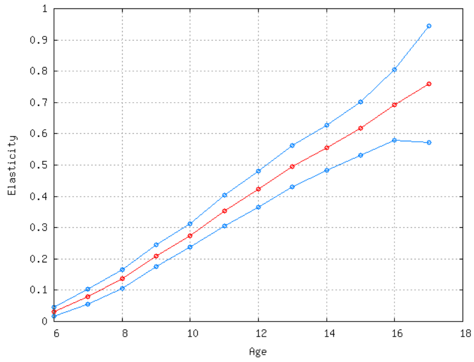


(f) Age

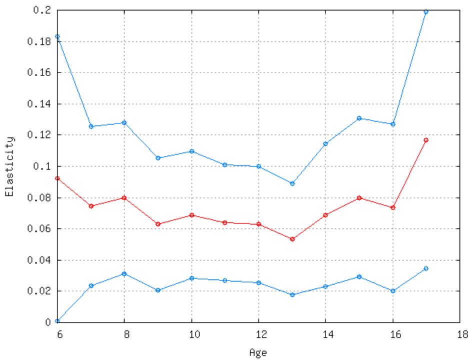
Figure 2: AP score, Fit and Elasticities, by age of child, with 2 standard error bars



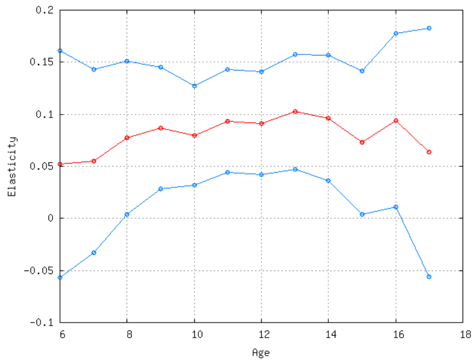
(a) Fit



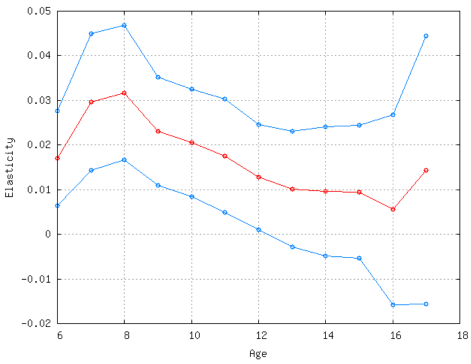
(b) Lagged score



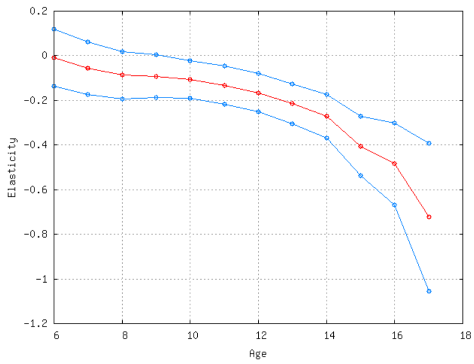
(c) Mother's education



(d) Father's education

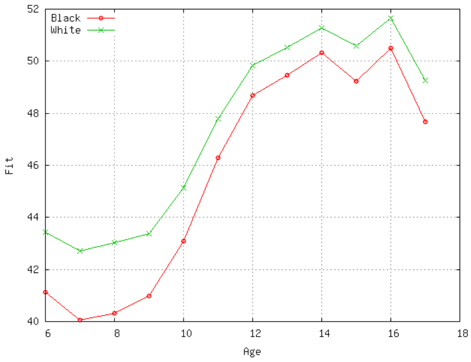


(e) Income

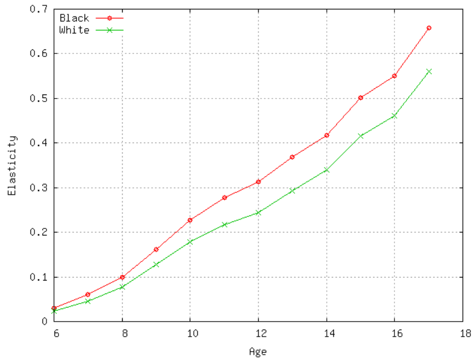


(f) Age

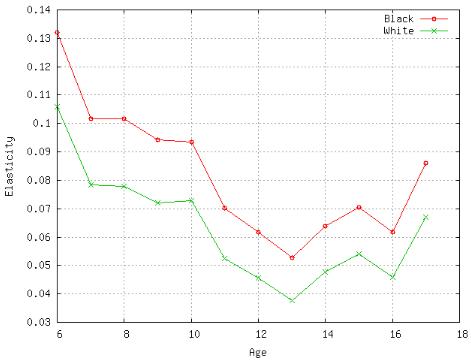
Figure 3: LW score, Black-White differences, Fit and Elastisticities



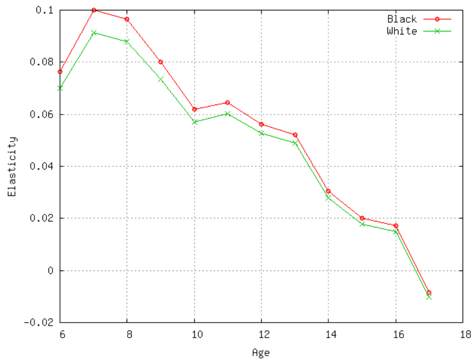
(a) Fit



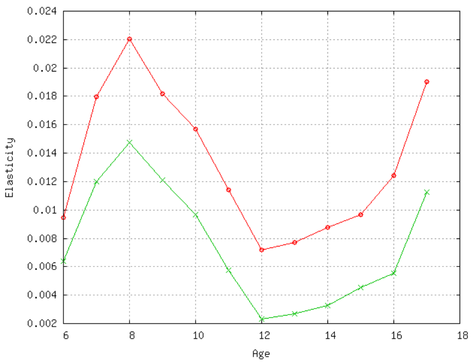
(b) Lagged score



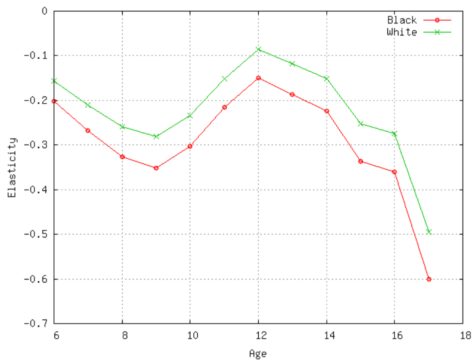
(c) Mother's education



(d) Father's education

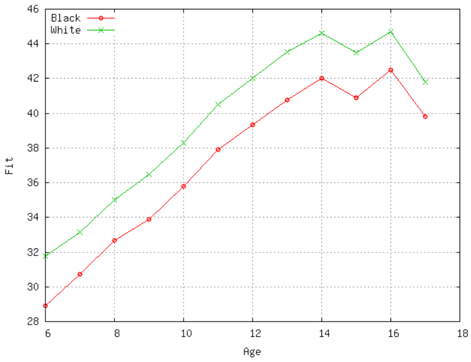


(e) Income

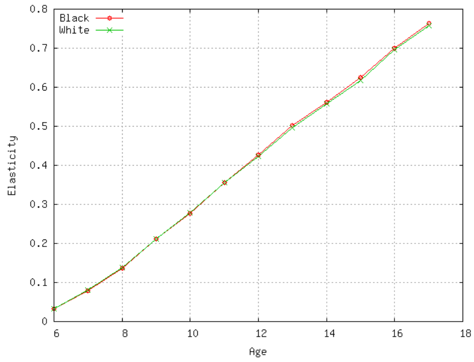


(f) Age

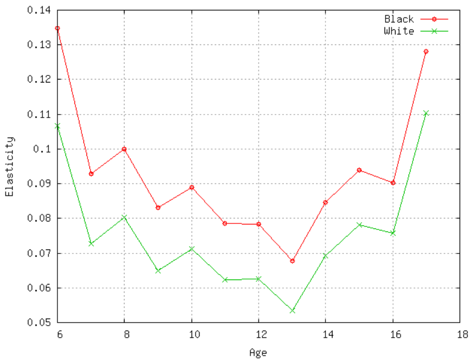
Figure 4: AP score, Black-White differences, Fit and Elastisticities



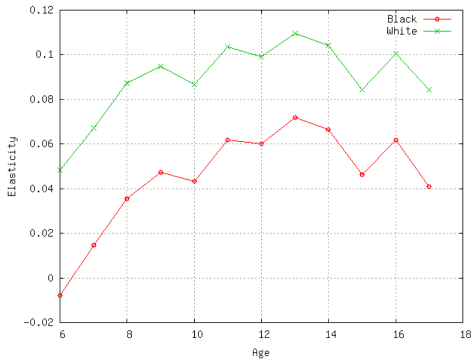
(a) Fit



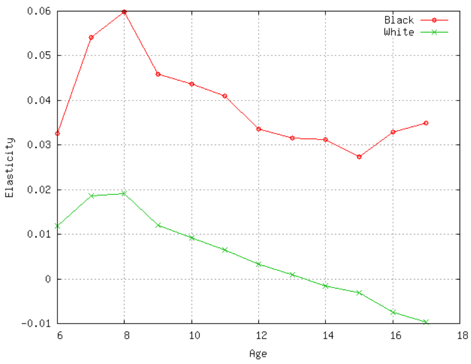
(b) Lagged score



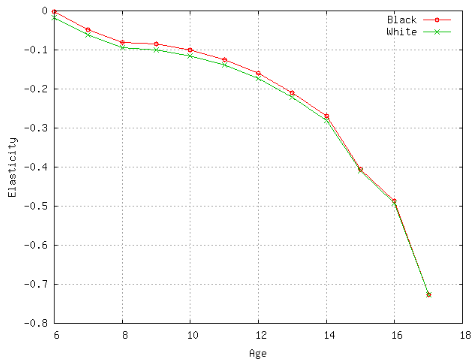
(c) Mother's education



(d) Father's education



(e) Income



(f) Age