On the Production of Cognitive Achievement and Gaps in Test Scores

Michael Creel and Montserrat Farell[*]

Universitat Autònoma de Barcelona

August 14, 2014

Abstract

We investigate the accumulation of cognitive achievement using an indirect production function, estimated using a flexible, dynamic econometric model and a new, rich data set. Gaps between black and white children remain constant, narrow, or disappear entirely as children grow older, depending upon the achievement measure and the family structure. Income elasticities are substantially higher for children of black families, and there are some notable differences in elasticities with respect to parents' educational levels. Within racial groups, the effects of fathers' and mothers' educational levels are different. Between children of stable two-parent families and children of mother-only families, a gap exists that is at least as important as the racial gap.

Keywords: education; cognitive achievement; test score gaps

JEL Codes: D13; I20; J15; J24

1

1 Introduction

Much has been written on children's cognitive achievement, its evolution over time, 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. 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 flexible, dynamic 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 not widen with age. For stable two parent families, gaps either narrow substantially (the LW score) or remain more or less constant (the AP score) over the ages 6-17. For mother-only families, the result is even stronger, with gaps in both scores narrowing and finally disappearing as children enter their teenage years. We calculate the elasticities of LW and AP scores with respect to conditioning variables, including 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. We also find that there is a gap in test scores between children of stable two-parent families and children of mother-only families. This gap is especially important for white children, and in general, the instability gap is at least as important as is the racial gap. This adds evidence to previous results that parental absences have a detrimental effect on child's academic performance (Lyle, 2006).

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

---

[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.

some variables is likely to be of concern. We also take into serious consideration the issue of the functional 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.

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 rely at least in part on 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 data set is based upon two waves of survey data, so that current and historical information is available for each child. Among other benefits, this allows for conditioning on previous measures of achievement. To our knowledge, this is the first paper that uses this data to estimate an educational achievement production function.

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. Blau (1999) also gives a useful outline of the issues.

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 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 problem is the issue of the observability of inputs. Some inputs, both endogenous and exogenous, are not observable, at least to the econometrician. While the number of inputs that can affect cognitive achievement is no doubt large, surveys can gather only a limited amount of information. 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 refer 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 that is self-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 and x_t^u. To simplify the exposition and notation, we assume 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 childrens' 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 (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:

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

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 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 parentally-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 parentally-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. Instrumental variables estimation can be used to deal with the possible endogeneity of the history of achievement.

Finally, the indirect production function provides a simpler, clearer framework within which to analyze policies directed to improve children's 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. Our simple presentation above assumes that these variables are fixed over time, but it is a simple extention to allow them to vary or to be multi-dimensional. Policy changes that affect exogenous factors such as school characteristics can also be analyzed using the indirect production function. On the other hand, full knowledge of the direct production function would not in itself be enough to allow for policy analysis, since one would still need to know the effects of policies on the levels of the direct inputs. 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 analysis and policy formation.

3 Data

As noted in the introduction, we use the Child Development Supplement (CDS, Mainieri, 2006) to the Panel Study for Income Dynamics (PSID). 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 information from the PSID and the two waves of the CDS, we obtain two time series observations on several measures of children's cognitive achievement, as well as covariates such as family income, parental education, race, time use, *etc.*

To limit the heterogeneity that we expect the econometric model to deal with, we consider two subsamples. Our main sample (970 observations) is restricted to children who lived with the same two parents in both CDS waves. The secondary sample (766 observations) is made up of children who lived with only their mother during both waves of the sample. Our results, discussed below, show that the evolution of cognitive achievement in these subsamples is substantially different, suggesting that one can speak of an "instability gap". In this paper, we make no attempt to use information about children who experienced a change in family structure during the period between the two CDS waves.

Turning to the variables, we use the Letter-Word (LW) and Applied Problems (AP) 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 use the years of education of the mother and the father as measures of the parents' endowments (μ_p). For the two-parent subsample we use both parents' educational level, but for the mother-only subsample, we are forced to use only the mother's educational level, since the father's level not available. 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 that we report, using the chosen measure of income. The observable exogenous variables (the Z_t^o of the previous section) are race, sex, and age of the child. The racial classifications we use are "white", "black" and "other". In the two parent data set, we have 608 white children, 241 black children,

7

and 121 children in the other group. In the sample of children living with only the mother the breakdown is 158 white children, 571 black children, and 37 children in the other classification.

The data set contains information on school characteristics for children who attend public schools, but this information is not available otherwise. For this reason, we do not include measures of school characteristics. This is a limitation of our analysis which we hope to address in future work.

We did some exploratory work with the time-use diary data that the CDS contains, looking at daily hours parents' spent interacting with their children, and the 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. Nevertheless, we did explore the effect of including time-use variables in the model. We were somewhat surprise to find that they did not have any significant impact and their exclusion was not rejected by formal statistical tests. Based upon the above considerations, we do not use the time-use variables in the model.

Work by other authors, for example, Todd and Wolpin (2004) and Fryer and Levitt (2004) has included parentally-chosen inputs such as number of books in the household, and they have appeared as significant regressors. 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 to have a significant effect when the variables that appear only after substituting in the optimal solutions (those related to the budget and information processing abilities of the parents) 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 through 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 ad-

8

dressed (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 vary freely, even at a single arbitrary point of evaluation. Our econometric model of the indirect production function allows for nonlinear and interaction effects, along the lines of the flexible functional form literature (Caves and Christensen, 1980). We find that nonlinearities and interaction effects are important, since restrictions that supress them are strongly rejected. This suggests that results in the literature based upon simple linear models, or very limited extensions, may suffer from biases.

Since our model includes the lagged test score and we have only two waves of data, the sample that is used to estimate the model 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. For this reason, 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), as discussed above, 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 must assume that the entire history of achievement Q_{t-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 can be written as

$$q \;\; = \;\; q(\mathbf{x}) + \eta(\mathbf{x}, Z^u, \epsilon) \tag{3}$$

We define $\eta \equiv \eta(\mathbf{x}, Z^u, \epsilon)$, and treat it as an econometric error term.

Comparing the theoretical version of the indirect production function in equation 2 with the econometric model, in equation 3, we see the correspondences $Z^0 = AGE, SEX, B, W$, $\mu_p = ME, FE$, $Q_{-1} = q_{-1}$, which makes clear the connection between the theoretical development and the econometric model. The possible dependence of the error term on the variables in $\mathbf{x}$ and Z^u emphasizes the possibility of endogeneity and heteroscedasticity. Even if η does not depend on $\mathbf{x}$, it is still likely to be heteroscedastic due to its capturing the effect of the unobserved exogenous variables, Z^u. As these variables are likely to be collinear with the observed exogenous variables,

---

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

there is a distinct possibility that the conditional variance will not be constant. However, we do not claim that heteroscedasticity is necessarily present, we merely note that it is a possibility, in our view, a quite likely possibility. Regarding exogeneity, we assume that all variables in $\mathbf{x}$ except q_{-1} are exogenous. It is possible that parents could obtain additional years of education in response to their assessment of their childs' endowment, and it is also possible that they could seek to alter family income in response to the same factor, so exogeneity of ME, FE and M may not be a certainly. However, we believe that such effects are likely to be small if they exist at all, so we assume exogeneity of these variables . For the variables in Z^o, we believe that exogeneity requires no discussion. However, endogeneity is to be expected in the case of q_{-1}, since the unobserved endowment of the child, μ, affects both the current and lagged scores.

We specify a quadratic parametric model for $q(\mathbf{x})$, so our econometric specification is

$$q \quad = \quad \alpha + \mathbf{x}'\beta + \frac{1}{2}\mathbf{x}'\Gamma\mathbf{x} + \eta \qquad (4)$$

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

$$q \quad = \quad \mathbf{z}'\theta + \eta. \qquad (5)$$

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. 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}. The elements of $\mathbf{z}$ that do not involve q_{-1} are weakly exogenous by the above assumptions. The remaining instruments are the first lags of these instruments. Under our assumptions, these variables are clearly not correlated with the error η, and they are also clearly correlated with the potentially endogenous variable, q_{-1}, following

equation 5. We also estimate using OLS.

Plots of the GIV and OLS residuals strongly suggest that the errors are heteroscedastic, as is to be expected. White's test for homoscedasticity (White, 1980), using **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 η in equation 4 is modeled as $V(\eta) = \exp(\alpha_1 + \alpha_2 \log AGE)$, which we expect is only an approximation to the true conditional variance.[4] We continue to apply a heteroscedastic-consistent covariance matrix estimator, to allow for residual heteroscedasticity that is not captured by this simple approximating model of the error variance. Most applied work on cognitive achievement makes no effort to control for heteroscedasticity. Given that it is to be expected on theoretical grounds, and based upon the empirical results that strongly confirm its presence, we believe that this is unfortunate, because of the loss of efficiency in estimation.

The theoretical development of the last section suggests that the regressors in **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 almost certainly inefficient, given the low p-values of the White tests (Tables 1 and 2, rows 1), and after the simple correction, any remaining unmodelled heteroscedasticity or nonnormality of the errors would 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 exogeneity 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

---

[4]Separate models are used for LW and AP scores.

11

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, 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 through 9 report results for Wald tests of several hypotheses. The hypotheses tested, and the corresponding rows in the Tables are:

- (row 6) The model can be reduced to a simple linear specification, without interactions or nonlinearities in the variables. This hypothesis implies that Γ in equation 4 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, and 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 reasonably convincingly, at least at the 10% level. Thus, we do not impose any restrictions upon the general quadratic specification (equation 4). 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 report their estimated values[5]. Instead, we present plots of predicted LW and AP scores, and elasticities of both predicted scores with respect to the explanatory variables, along with two standard error bars. Because estimated elasticities are

---

[5]Parameter estimates are available upon request.

nonlinear functions of the estimated parameters of the models, the delta method[6] was used to calculate the estimated standard errors of the elasticities. We present results only for black and white children, because the small sample size of the "other" group limits the precision of the results.

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. In Tables 3 and 4 we give the descriptive statistics for raw LW and AP scores, by age and race, without controlling for 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 results after controlling for covariates though the econometric model, we first present fitted scores and elasticities, plotted as functions of the child's age. Figures 1 and 2 present the fit and elasticities for the AP score, for white and black children, respectively, while Figures 3 and 4 do the same for the LW score. In Figures 1-4, we present results using the data for children living in stable two-parent families. It is to be emphasized that the use of this selected sample means that the results will no longer be directly comparable with the results discussed in the last paragraph, which use the entire sample. We evaluate elasticities and fit at the sample means of the explanatory variables, given the age of the child.[7] 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.[8]

In panels (a) of Figures 1 and 2, we see that AP score generally rises with age, except for some non-monotonic behavior in the teenage years, which is accompanied by an increase in the breadth of the 2 standard error bars. The pattern is the same for white and black children, and a black-white gap in scores conditional upon covariates is apparent (we return to this in more detail below).

In panels (b) we see that for both racial groups 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 1 for a person who had arrived to a stable level of AP score, since the

---

[6] See, for example, Cameron and Trivedi (2005), pg. 231-2.
[7] In the case of family income, which has a quite skewed distribution, we use the sample median, conditional on age.
[8] Older children have older parents, who have higher incomes and more education, on average.

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, with a value of 0.75 for both blacks and whites. The fact that the elasticity is different from one at age 17 means that the AP score is still malleable at this age. This is also apparent in panels (a), which show that the AP score is still generally increasing in the teenage years. This result coincides with previous work (*e.g.*, Hansen *et al.*, 2004) that shows that cognitive ability does not reach a fixed level in early childhood, as Hernstain and Murray (1994) argue.

Panels (c) of Figures 1 and 2 show that the years of education of the mother has a similar impact for both black and white children. The elasticities are positive and significantly different from zero at all ages. Both curves have a similar shape, with that for black children lying about 0.02 points above that for white children. To facilitate interpretation, note that a mother with a college degree has approximately 33% more years of education than does a mother with a high school degree. An elasticity of 0.1 implies that such a change in the mother's educational level would lead to approximately a 3.3% increase in the AP score. For 10 year old children, a 3.3% increase in the AP score would be about a whole point.

In Figures 1 and 2, panels (d), we have the elasticities of AP score with respect to years of education of the father. For white children, the elasticity is close to 0.1 except for the youngest children, and it is significantly different from zero between ages 9-16, inclusive. Quite interestingly, it is of greater magnitude than is the elasticity of AP score with respect to the mother's education. This result is new in the economic literature on children's acquisition of cognitive ability, to our knowledge. For black children, the elasticity with respect to the father's education is of a lower magnitude than is the elasticity with respect to the mother's education, and it is not significantly different from zero at any age of the child.

Panels (e) of Figures 1 and 2 show the elasticities of the AP score with respect to family income. For black children, this elasticity is positive and significantly different from zero from ages 6-16 inclusive, while for white children, it is closer to zero, even negative for some ages, and it is not significantly different from zero at any age. Overall, this elasticity is 3-4 times higher for blacks than for whites, at comparable ages. Though the income elasticity is small even for black children, it is important to remember that the variation in incomes is large. It is reasonable to study the effect of a 100% change in income, since the sample includes much larger differences in incomes. A 100% increase in income with an elasticity of 0.05 would lead to approximately a 5% change in AP score. For 10 year old children, the impact would be close to two points. At age 10, the gap in the AP score between black and white children is approximately 2 points (compare Figures 1

14

and 2, panels (a)). Also, the effects of income and other variables at young ages are transmitted forward due to the effect of the lagged score variable, so that impacts accumulate as children age. For these reasons, we find that the difference in the income elasticity across black and white families is quite interesting. Blau (1999) presents a careful recent study of the effect of income on cognitive achievement. It is worth noting that Blau's model, which is linear in the variables and which allows only the constant to vary by racial group, through dummy variables, does not pick up the difference in income elasticities that we find here. Recall that we test and reject the hypothesis that our model can be collapsed to a simple linear model with dummy variables (lines 6, Tables 2 and 1).

Panels (f) of Figures 1 and 2 show the elasticity of the AP score with respect to the child's age, plotted by age. This panel requires a careful interpretation, since the plot is made holding lagged score constant at the sample mean for children of each given age. However, as we can see in panels (a), scores, and consequently lagged scores too, generally rise with age. A partial derivative with respect to age, holding lagged score constant, corresponds to a child growing older without realizing the normal improvement in cognitive ability, which could be interpreted as a learning problem. This would imply a lower predicted score, and consequently a negative elasticity, as is observed in the plot.

We now turn to discussion of the results for the Letter-Word (LW) score. Figures 3 and 4 present plots that are analogous to those for the AP score that we have seen in Figures 1 and 2. Figure 3 shows plots of fit and elasticities for white children, and Figure 4 gives the same results for black children.

Beginning with panels (a), which plot the fitted LW score as a function of age, we see a similar shape for both black and white children. The LW score increases with age, except for some non-monotonic behavior for the youngest children and for teenagers. For the 6 year-olds, the standard error bars are wide enough to suggest that the non-monotonicity may just be an artifact. The rate of increase of LW score is moderate from ages 7 to 9, and then accelerates for 10 - 12 year olds. From ages 13 to 14 there is an additional moderate increase, which is followed by more erratic behavior.

Early gains in LW score persist, as is reflected in panels (b), where the elasticity of current LW with respect to lagged LW is seen to be positive at all ages. This elasticity is significantly different from zero and from one at all ages. It trends upward, and finally reaching values of 0.56 and 0.66 for whites and blacks, respectively. Considering that the comparable elasticity for the AP score is 0.76 for both groups, this suggests that LW scores remain more malleable at age 17 than are AP scores.

Years of education of the mother, in panels (c), has a significant impact at all ages for both white and black children. The elasticity is larger for younger children, and for black children. The plot of the elasticity of LW score with respect to the father's years of education, in panels (d), is very similar for white and black children. It is not significantly different from zero at any age for black children, and it is barely significant at a few ages for white children. Comparing these results with those for the AP score, we see that the father's role is quite clearly important in the case of the AP score for whites, but otherwise the evidence suggests that the mother's educational level is a more important factor.

Panels (e) show that the LW score of children of black families is more elastic with respect to income than is the case for children of white families, at all ages, though income elasticities are generally low in magnitude. Again, these results need to be interpreted with some care. The LW score 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 consequently, when they are older they will also have somewhat higher values of the lagged LW score. However, the plots in Figures 3 and 4 set the value of the lagged score to the overall sample mean, which ignores the impact of persistently higher income on the lagged LW score. 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 the elasticities in the Figures suggest, due to the dynamic effect that is transmitted forward through the lagged score.

Panels (f) show the elasticity of the LW score with respect to the child's age, plotted by age. The interpretation of the negative elasticity is the same as that given above for the AP score.

In Figure 5 we plot the fitted AP scores for black and white children using the primary sample of stable two parent families[9], along with the fitted AP scores for black and white children using the secondary sample of children who live with their mothers only. This plot allows examination of possible racial and instability gaps in AP scores. The model for the secondary sample is identical to the model for the primary sample, except that the father's education is not available as an explanatory variable. Considering the stable two parent family results, there is a racial gap that remains more or less constant at 2-2.5 points. It is widest from ages 9-14, and it narrows slightly for children of 16 and 17 years of age. It is clearly not increasing with age. Covariates explain much of the gap in raw scores, as can be seen comparing the 2-2.5 point gap in Figure 5 with the 5-6 point gap in raw scores, in Table 4. Looking at the results for the mother only families, the racial gap in AP scores decreases in magnitude for children 14 and older, and the gap even reverses its sign for 17 year old children. However, two standard error bars (not plotted,

---

[9]These fitted scores already appear in panels (a) of Figures 1 and 2.

to avoid over-cluttering the Figure) make it clear that the sign reversal is not statistically significant. Nevertheless, there is no evidence of a widening racial gap in AP scores. It is interesting to observe that the gap between scores of children from two-parent families compared to children of mother-only families (an "instability gap"), for both black and white children, is of similar or greater magnitude than is the gap between black and white children, for a given type of family structure. The instability gap for white children is in general larger than the instability gap for black children, and is in general larger than the racial gap.

Figure 6 is analogous to Figure 5, except that results for the fitted LW score are plotted. 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. It also declines substantially as children grow older, narrowing from 2.5-3 points down to roughly 1 point in the case of stable families, and to no difference in the case of mother-only families. The instability gap, for both black and white children, is larger than is the racial gap between children of families with a given structure.

6 Conclusions

The theoretical structure we put forth regarding the specification of a model of an indirect cognitive achievement production function leads to a relatively parsimonious econometric model with less severe problems of endogeneity than are likely to exist if a direct model is estimated. The results that we report above show that the indirect approach, conditioning on past achievement and using a flexible specification, leads to interesting empirical results that have not previously been seen in studies that use direct or mixed direct/indirect production functions. Racial gaps in test scores exist and are important, but they do not widen with age after controlling for covariates, as appears commonly in the literature. This holds for both the mathematical and verbal achievements. The result is stronger in the case of mother-only families, where gaps reduce to zero or even change sign. We also have found some interesting variations in elasticities across racial groups, with parents' educational levels and income having substantially different effects. Furthermore, the existence of information on years of education of both the mother and the father allows us to observe the important role of fathers' education in the production fuction, which has not been taken into account in most of this literature. It is also observable that the instability gap in test scores of children of mother-only families compared to children of stable two-parent families is at least as important as is the racial gap.

Several features of our model and methodology can account for these findings. First, the model is more highly parameterized than are most models that have been used to analyze similar

data, including quadradic terms and interaction terms. The added flexibility allows the model to better capture the dynamics of cognitive achievement. Because various parametric restrictions that remove some of this flexibility are very strongly rejected, we conclude that the model captures features of the dependent variable that more limited models cannot. The better fit to the dependent variable afforded by the nonlinear in variables model allows us to observe curvature and nonlinear effects apparent in Figures 1-4 which would not be revealed by an excessively restrictive linear in the variables model. Secondly, we include the lagged score as a regressor, to control for the information that parents have when they make their choices regarding direct inputs. This regressor is highly significant, which means that its omission in all probability lead to a model that suffers from dynamic misspecification and missing variable bias. A third factor is the use of a data set that has not been extensively explored. The results of this paper make it clear that conclusions can change depending on the model and the data set. Finally, we estimate separate models for stable two parent families and for children who live with only their mother. Given the importance of the stability gap that we have observed, it seems that it is important to use and econometric model that deals with stability differences in family structure in a flexible way, rather than just including a dummy variable. An important question for future work would to be to estimate a variety of models, direct and indirect, and possibly fully structural models, using a variety of data sets, to attempt to identify more precisely the sources of the different results that are obtained.

References

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

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

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

[4] Bureau of Labor Statistics (2001). *NLS Handbook 2001*, U.S. Department of Labor (Washington, D.C.).

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

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

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

[8] Creel, M. (2004) "Modified Hausman Tests for Inefficient Estimators." *Applied Economics*, *36*(21), pp. 237-76.

[9] Fryer, Jr., R. G. and Levitt, S,D. (2004) "Understanding the black-white Test Score Gap in the First Two Years of School." *The Review of Economics and Statistics*, *86*, pp. 447-464.

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

[11] Haveman, R. and Wolfe, B. (1995) "The Determinants of Children's Attainments: A Review of Methods and Findings." *Journal of Economic Literature*, *33*, pp. 1829-1878.

[12] Herstein and Murray (1994) The Bell curve: Intelligence and Class Structure in American Life, The Free Press (New York).

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

[14] Leibowitz, A. (1974) "Education and Home Production", *American Economic Review, 64*, pp. 243-250.

[15] Lyle, David S. (2006) "Using Military Deployments and Job Assignments to Estimate the Effects of Parental Absences and Household Relocations on Children's Academic Achievement", *Journal of Labor Economics, vol. 24, nž2*, pp. 319-350.

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

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

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

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

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

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

[22] Woodcock, R. W., and Johnson, M. B. (1989) "Tests of Achievement, Standard Battery [Form B]", Riverside Publishing (Chicago).

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
9	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
6	Simple linear model	Wald	$\chi^2(32)$	87.72	0.000
7	black=white=Other	Wald	$\chi^2(14)$	66.09	0.000
8	black=Other	Wald	$\chi^2(7)$	12.20	0.094
9	boys=girls	Wald	$\chi^2(8)$	13.79	0.088

Table 3: Letter-Word (LW) score by age and race

age	mean (overall)	mean (blacks)	mean (whites)	gap	gap/st.dev.
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: Applied Problem (AP) score by age and race

age	mean (overall)	mean (blacks)	mean (whites)	gap	gap/st.dev.
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: Applied Problem (AP) Score, Fitted Score and Elastisticities by Age (with 2 standard error bars) - White Children
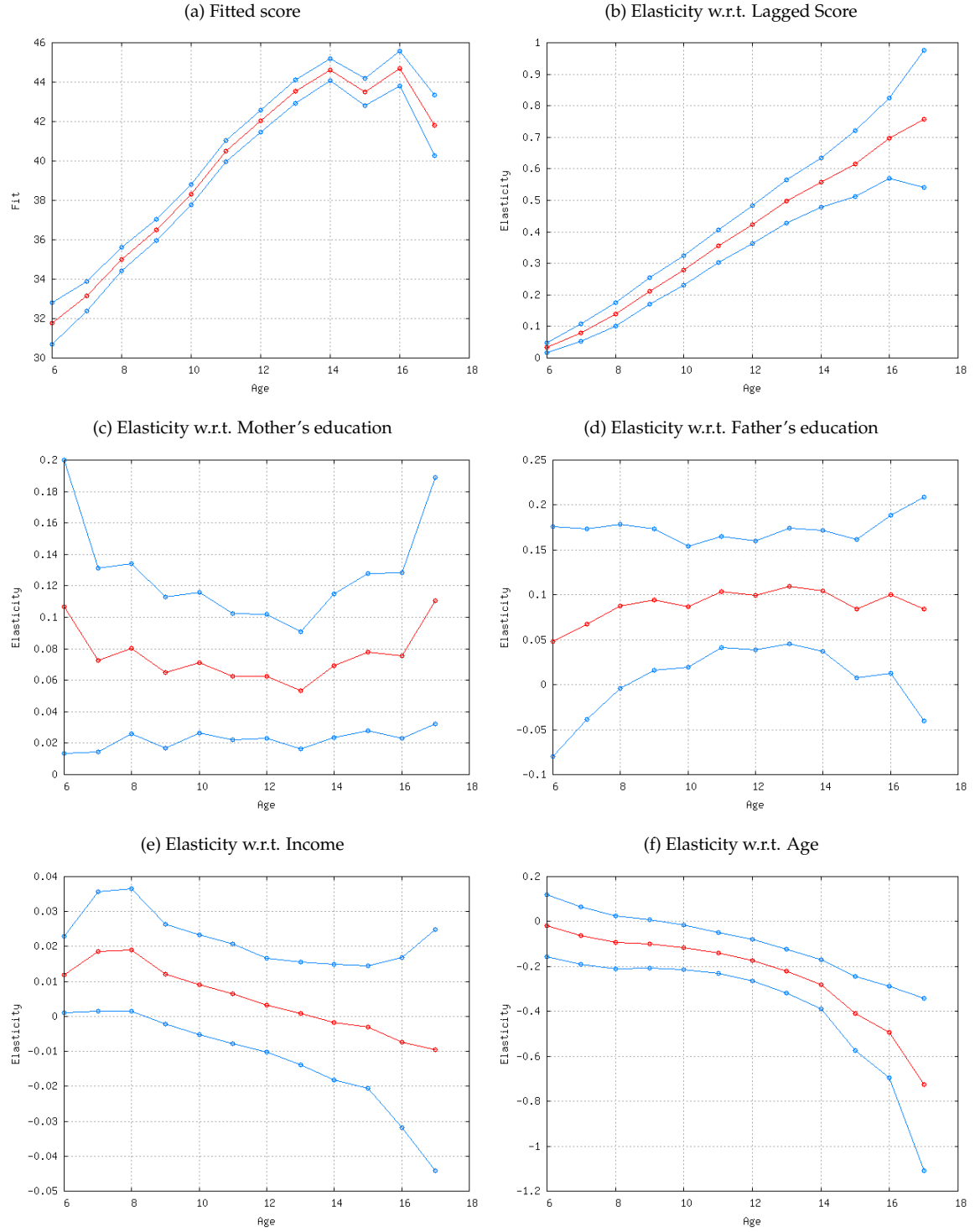

(a) Fitted score

(b) Elasticity w.r.t. Lagged Score

(c) Elasticity w.r.t. Mother's education

(d) Elasticity w.r.t. Father's education

(e) Elasticity w.r.t. Income

(f) Elasticity w.r.t. Age

Figure 2: Applied Problem (AP) Score, Fitted Score and Elastisticities by Age (with 2 standard error bars) - Black Children

(a) Fitted score

(b) Elasticity w.r.t. Lagged Score

(c) Elasticity w.r.t. Mother's education

(d) Elasticity w.r.t. Father's education

(e) Elasticity w.r.t. Income

(f) Elasticity w.r.t. Age

24
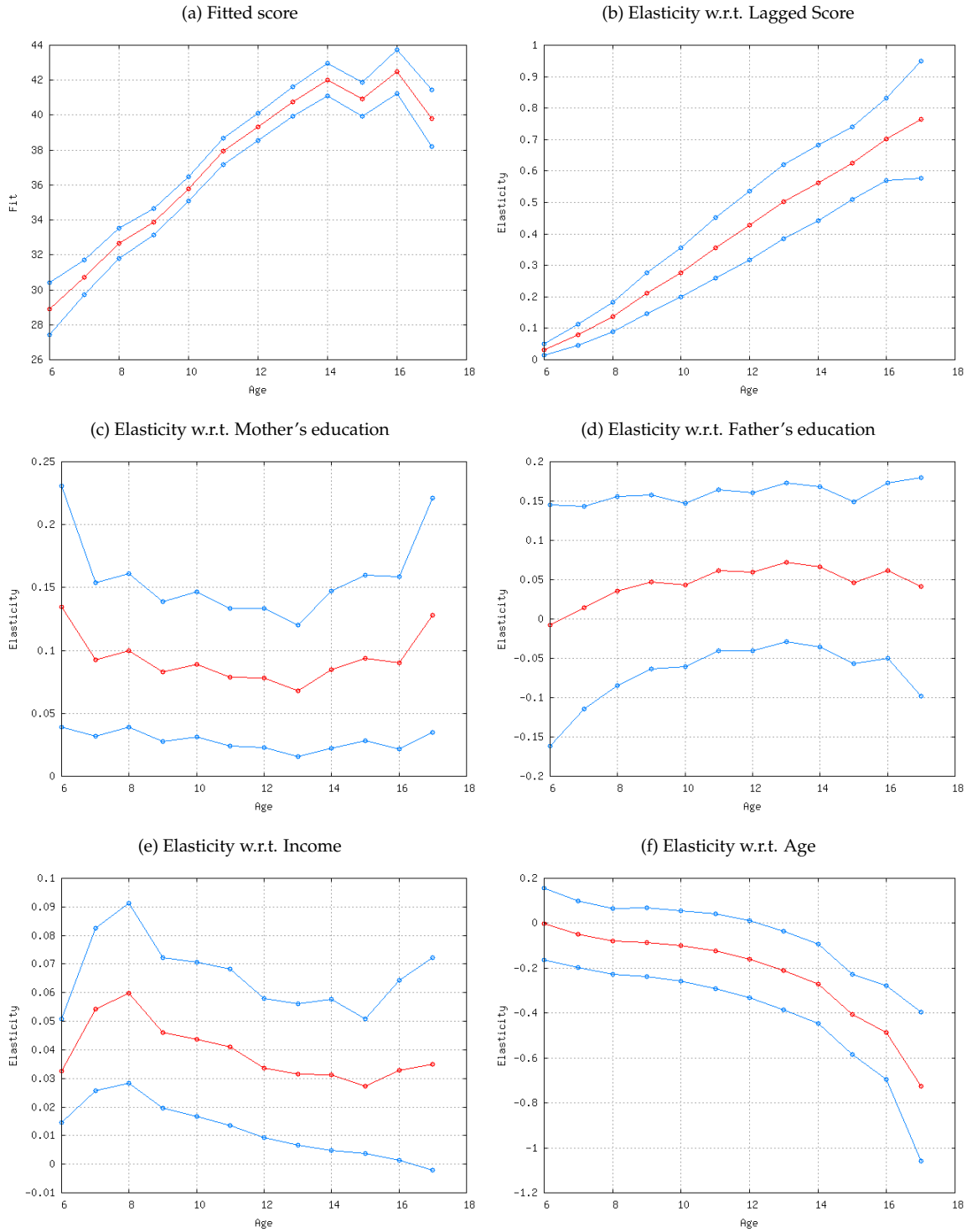
Figure 3: Letter-Word (LW) Score, Fitted Score and Elastisticities by Age (with 2 standard error bars) - White Children

(a) Fitted score

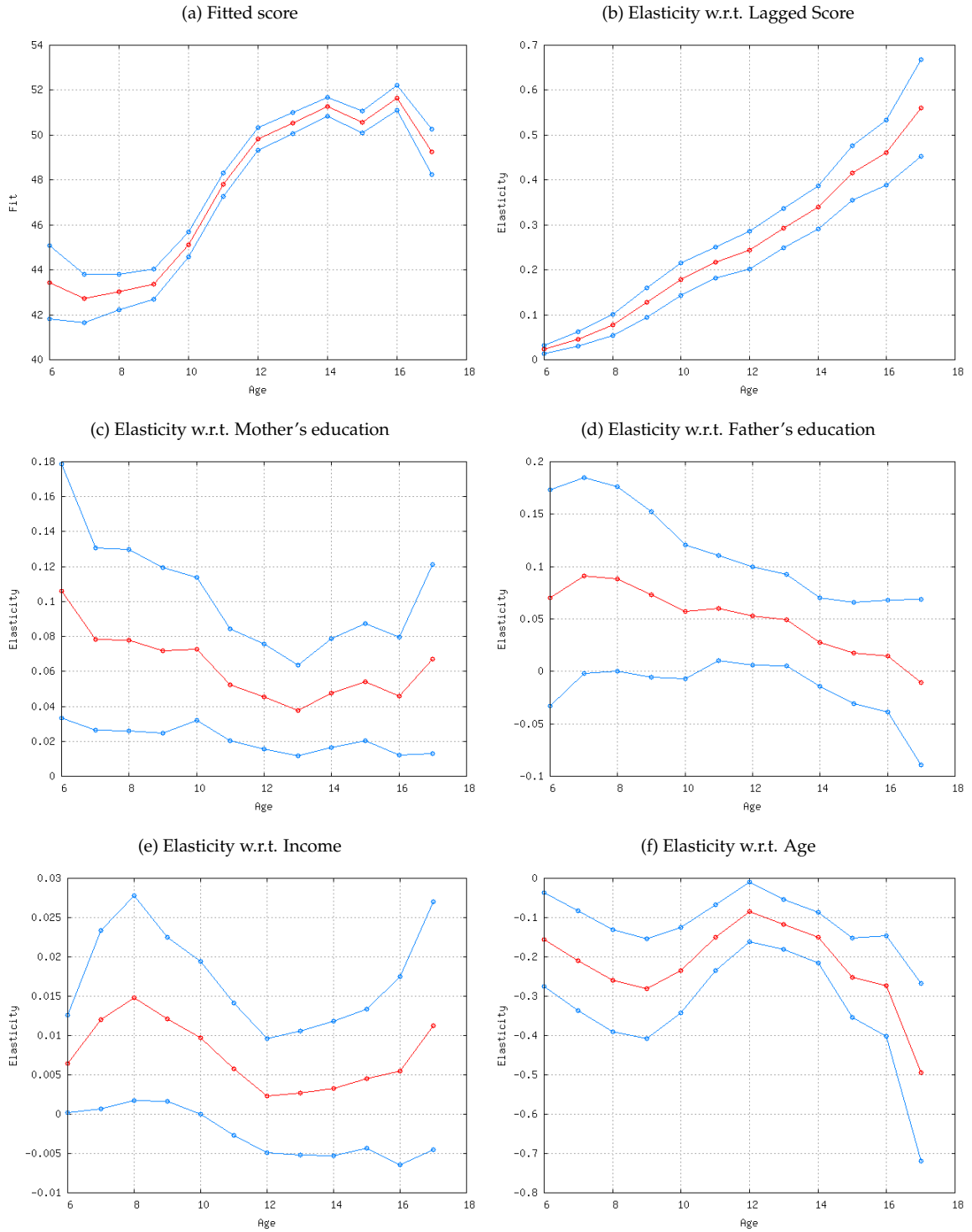(b) Elasticity w.r.t. Lagged Score

(c) Elasticity w.r.t. Mother's education

(d) Elasticity w.r.t. Father's education

(e) Elasticity w.r.t. Income

(f) Elasticity w.r.t. Age

Figure 4: Letter-Word (LW) Score, Fitted Score and Elastisticities by Age (with 2 standard error bars) - Black Children

(a) Fitted score

(b) Elasticity w.r.t. Lagged Score

(c) Elasticity w.r.t. Mother's education

(d) Elasticity w.r.t. Father's education

(e) Elasticity w.r.t. Income
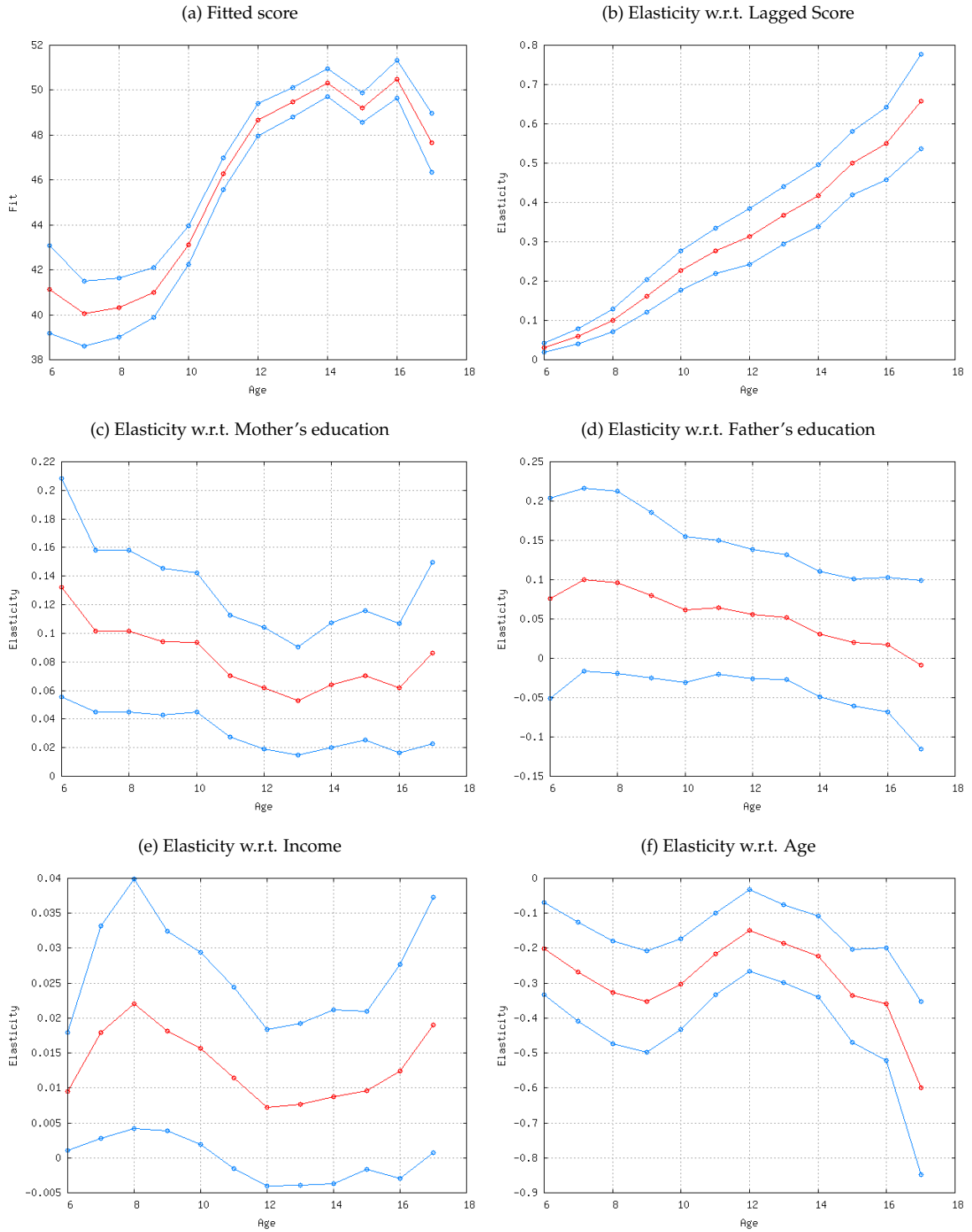
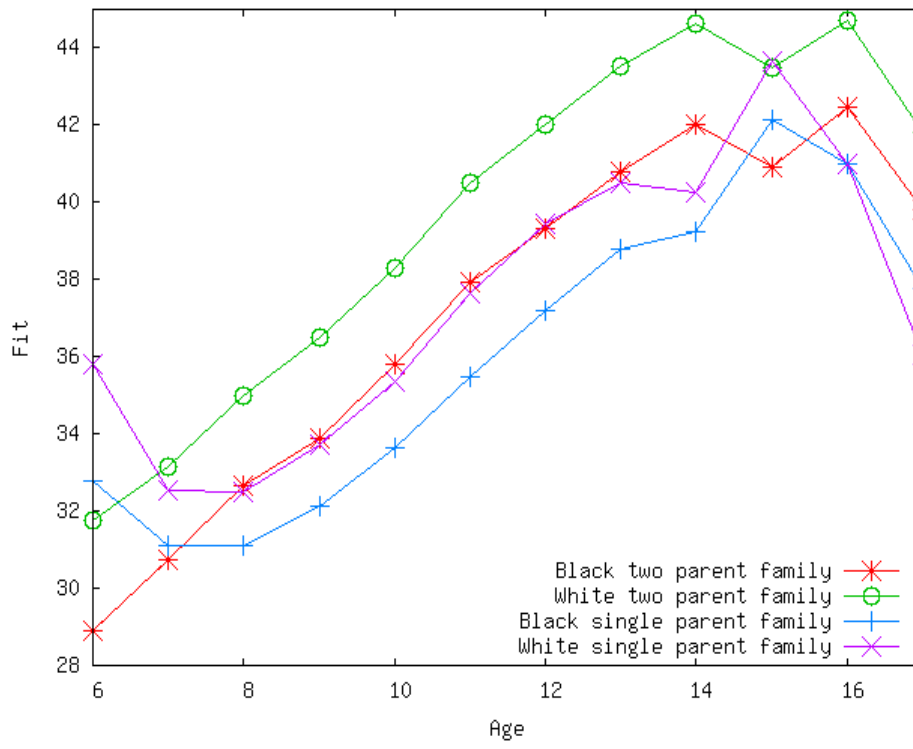(f) Elasticity w.r.t. Age

Figure 5: AP score, black-white differences

Figure 6: LW score, black-white differences