Essays on Human Capital Formation

by

Victor Andrea Checchia Ronda

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
July, 2017

Abstract

This dissertation consists of three essays on the economics of human capital formation. In these essays, I explore how parents determine the skills developed in children, how these skills lead to important economic outcomes and the issues involved in the measurement of these skills in children.

In the first chapter, "The Effect of Maternal Psychological Distress on Children's Cognitive Development", I study the relationship between maternal mental health and children's cognitive skills. I develop a model that allows me to separate the different mechanisms that relate maternal mental health to children's cognition. In order to identify the causal effect of maternal mental health, I exploit variation among U.S. states in mental health insurance coverage laws, which improved access to mental health care services. I find that maternal mental health problems mainly affect children through a decrease in the productivity (quality) of maternal time investments.

The second chapter, "The Economic Value of *Breaking Bad*: Misbehavior, Schooling and the Labor Market", studies the relationship between childhood misbehavior, schooling and labor market outcomes. We show that externalizing behavior (linked, for example, to aggression), reduces educational attainment yet increases earnings. This finding illustrates our main point that, different than cognition and health, non-cognitive skills can be productive in some economic contexts and counter-productive in others. As a result, policies designed to promote human capital accumulation could have mixed effects or even negative economic consequences, especially for policies targeting non-cognitive skill formation for children aimed solely at improving educational outcomes.

In the third and final chapter, " When Mothers and Teachers Disagree: Observer Reports and Children's Noncognitive Skills", I explore the methodological challenges involved in measuring noncognitive skills in children. The usual approach involves asking parents or teachers about different child behaviors. This is problematic, as I show that mothers and teachers rarely agree when reporting on these behaviors. More importantly, I show that maternal and teacher reports are measuring different aspects of child development. While teacher reports mainly measure child misbehavior associated with adult risky behaviors, maternal reports also measure behaviors related to the child's mental health.

Primary Reader: Nicholas W. Papageorge

Secondary Reader: Robert A. Moffitt

Acknowledgement

I would like to thank my PhD advisors, Professor Nicholas Papageorge and Robert Moffitt, for their support and encouragement these past six years. I would like to express my sincere gratitude to Nick for being a tremendous mentor since my second year in the program. Nick's constant encouragement and support helped me grow as a researcher. His advice on both research as well as on my career have been invaluable. I would also like to thank Professor Robert Moffitt for his insightful comments and encouragement. His advice, questions and comments helped improve my research and my understanding of economics as a whole. I would also like to thank Richard Spady for his constant encouragement and for always helping me question the methods and assumptions and view issues from multiple perspectives. I also benefited from the helpful comments and discussion with Matthew Wiswall, Yingyao Hu and Jorge Balat.

I would also like to thank my friends and colleagues: Emmanuel Garcia-Morales, Alanna Bjorklund-Young, Sohini Mahapatra and Daniel Garcia. They made the long PhD process fun and enjoyable. This journey would not have been possible without their friendship, encouragement, help and support.

Finally, I am especially grateful to my family for all of their love, patience and encouragement. I would like to thank my mother, Walenia Checchia. I owe everything in my life to her. I would also like to thank my grandmother, Denise Checchia, which passed away during my Ph.D, for teaching me what love and compassion are. Finally, I would like to thank Cecile Karsenty for her unconditional love, support and patience and for making every day of my life special, fun and joyful.

Contents

List of Tables

List of Figures

Chapter 1

The Effect of Maternal Psychological Distress on Children's Cognitive Development

1.1 Introduction

There is a growing interest, in a multitude of fields, towards understanding intergenerational transmission of human capital. Studies have shown that inequality in family resources is translated into inequality in children's outcomes (Heckman and Mosso, 2014; Duncan, Kalil, and Ziol-Guest, 2013; Currie and Almond, 2011; Alexander, Entwisle, and Olson, 2014). Moreover, we now know

I would like to thank Robert Moffitt, Nicholas Papageorge and Richard Spady for their helpful comments, suggestions, encouragement and guidance. I gratefully acknowledge helpful comments from Matthew Wiswall, Yingyao Hu and Jorge Balat, along with seminar participants at Johns Hopkins University, the WEAI Graduate Student Workshop, Escola de Economia de So Paulo, Insper, Universit de Montral, Aarhus University, IZA Workshop on Gender and Family Economics and SOLE 2017. All errors are my own.

that at least half of the variation in lifetime earnings is determined in childhood (Cunha, Heckman, and Navarro, 2005). Many factors have been shown to explain the transmission of poverty, including for example genetic endowments or the number of words spoken to the child in infancy. These results are often discouraging as many of these factors are not easy policy targets. This is not the case for maternal mental health, which has been shown to be highly correlated with both socioeconomic status and child outcomes (Cogill et al., 1986; Caplan et al., 1989) and proven to be malleable through policy (Earls et al., 2010; Evans and Garthwaite, 2014). However, maternal mental health has been understudied in Economics and as a result we know very little about the mechanisms through which it affects child outcomes.

The objective of this paper is to evaluate the mechanisms that explain the effect of maternal mental health on children's cognitive development. I build on many strands of the child development literature to describe the different mechanisms that relate maternal mental health to children's cognitive development. In order to estimate these different mechanisms, I estimate policy functions for different maternal investments jointly with the child's technology of skill formation. This approach allows me to assess the effect of maternal mental health on the quantity as well as on the productivity (quality) of maternal investments. I use variation in parity laws that determine mental health care access and coverage across states to estimate the causal effect of maternal mental health.

I find that maternal mental health matters for children's development. Moreover, I find that maternal psychological distress mainly affects children through a decrease in the productivity (quality) of maternal time investments. Next, I investigate policy interventions that mitigate these effects. My findings suggest large payoffs for children from mental health treatment for at-risk mothers. My findings also suggest that programs that improve maternal parenting can have large benefits for these children. Moreover, both policies are significantly more cost effective than comparable income transfers.

Different fields use different models to understand human capital formation in children. I bring together these different models by incorporating maternal mental health into a standard economic model of maternal investments. Economists understand child development through the family investment model (Becker, 1981; Becker and Tomes, 1986). In this model, parents influence children through biological endowments (genetics) and social endowments (values) as well as through time and monetary investments. In this basic framework there are at least five mechanisms that explain how maternal mental health can have an effect on children. The first mechanism comes from the idea in psychology that mental illness can be contagious, so children of distressed mothers are more likely to develop mental health problems of their own, which would impair their cognitive development (Rosenquist, Fowler, and Christakis, 2011; Currie and Stabile, 2006). The second mechanism is the idea from the family stress model in sociology that maternal mental health problems can affect the quality of mother-child interactions as it diminishes the mother's ability to be supportive and engaged with her child (McLoyd, 1990; Conger et al., 1994; Yeung, Linver, and Brooks-Gunn, 2002). The third and forth mechanisms come from the idea in psychology that mental health problems can increase the cost of spending time in productive activities, and as a result can affect the amount of time the mother spends with her child as well as her labor force participation (Blair, 2010; Frijters, Johnston, and Shields, 2014). The last mechanism come from the economics literature, which has shown that mental health problems could lead to lower labor market productivity (Chatterji, Alegria, and Takeuchi, 2011). In turn, lower earnings could translate into lower monetary investments in children.

Identifying these mechanisms is important for policy, as different mechanisms point towards different policy proposals. For example, if mental health affects the quality of maternal parenting, home visitation programs that improve the quality of mother-child interactions might be highly beneficial for children of mothers in poor mental health. Alternatively, if the effect is through a change in the mother's labor market productivity, then income supplement programs such as the earned income tax credit (EITC) might be important for these families. Identifying these mechanisms can also highlight heterogeneous effects of mental health treatment across families. For example, the benefits of treatment will be higher for children of working mothers if the mental health effect is through the mother's labor market productivity, and possibly larger for stay-at-home mothers if the effect comes through the quality of mother-child interactions.

In estimating the causal effect of maternal distress, I confront two empirical challenges: measurement error in the mental health construct and the endogeneity of mental health. In order to control for the measurement error problem, I use an item response theory (IRT) model. IRT is a common method in psychology used to identify and construct unobservable scales from a series of discrete measurements. Mental health scales, including the Kessler 6 psychological distress scale that I use in this paper, are constructed from multiple self-reported discrete responses about different psychological symptoms. The IRT approach recognizes and controls for the intrinsic measurement error in these self-reported questionnaires. Moreover, it controls for the fact that measurements differ in quality, each providing a different signal about the unobserved mental health. I find that not controlling for these problems, and instead using a simple summation score, leads to biased and unreliable estimates.

In order to address the endogeneity of maternal mental health, I use variation in state mental health parity laws. These laws require insurers in the state to provide an equal level of benefits for mental illness and physical disorders.[1] These laws are generally thought to improve access to mental health services in the state, and have been shown to increase utilization of mental health care services and contribute to improve mental health outcomes (Harris, Carpenter, and Bao, 2006; Lang, 2013). In theory, these laws only enter the model through an effect on the mother's mental health and as a result serve as exclusion restrictions that identify the model. Otherwise, I would not be able to identify the causal effect of the mother's mental health, as it is possibly correlated with unobservable investments in children.[2] [3] The literature has struggled to correct for this problem often relying on poor instruments, bounding or propensity score methods (see Frank and Meara (2009) and Dahlen (2016) as examples). Exceptions are papers that use exogenous variation in stressors that could trigger mental health illness, such as terrorist attacks (Camacho, 2008) or the death of a relative or close friend (Persson and Rossin-Slater, 2014; Frijters, Johnston,

[1]These benefits include visit limits, deductibles, copayments, and lifetime and annual limits.

[2]For example, I do not observe neighborhood characteristics, such as the crime rate or school quality, that we know are important for children's development.

[3]Another problem is reverse causation. At the same time that maternal mental health can influence labor market and child outcomes, lack of financial resources and poor child outcomes can lead to maternal mental health problems (Dohrenwend et al., 1992).

and Shields, 2014).[4]

My findings show that maternal mental health, measured by mothers' psychological distress, has large effects on children's cognitive development. The main mechanism explaining these large effects is the effect of maternal mental health on the returns of maternal time investments (quality of mother-child interactions). This channel alone explains 70% of the effect of maternal distress on children. I also find evidence of a direct effect of maternal distress on children's cognitive development, possibly explained by the contagion of mental health. I find no evidence of the other mechanisms once I control for measurement error and endogeneity of maternal distress. These findings point towards two policy interventions for children of psychologically distressed mothers. The first is mental health treatment, either with therapy or medication. I find that treatment for at-risk mothers can have huge payoffs for children that are 16 times more cost effective than comparable income transfers. The second policy would be to improve the quality of mother-child interactions, as in ,for example, home visitation programs may do. Policies that improve maternal parenting reduce the negative effect of maternal distress on the returns of maternal time investments, and as a result produce large benefits for children of distressed mothers. These programs can be thought as complementary to mental health treatment and a viable option when treatment does not work.

The remainder of the paper proceeds as follows. In Section 1.2, I describe the data and my measure of mental health. In Section 1.3, I describe the conceptual model of maternal investments and highlight the different mechanisms through

[4] One issue with these instruments, when studying postpartum mental health, is that they can directly influence children and as a result would not be valid exclusion restrictions.

which maternal mental health can influence children's development. In Section 1.4, I describe my econometric framework and estimation strategy. In Section 1.5, I describe my main findings. In Section 1.6, I discuss the policy implications of these findings. Section 1.7 concludes.

1.2 Data and Preliminary Analysis

In this section, I first provide details on the data used and on how I construct the analytic sample. Then I discuss my measure of mental health and provide background information on the measure that might be informative for some readers. Lastly, I report estimates from a preliminary econometric model relating maternal mental health with child cognition and maternal investments. In particular, I demonstrate that maternal distress is negatively correlated with both child cognition and other relevant maternal investments.

1.2.1 The Panel Study of Income Dynamics

In this paper, I use data from the *Panel Study of Income Dynamics* (PSID) and its *Child Development Supplement* (CDS). The PSID is an ongoing dynastic longitudinal survey. It started as a nationally representative sample of 18,000 individuals living in 5,000 families in 1968 in the United States. The CDS collected information on 3,563 children living in 2,394 PSID families. Information was collected in three waves: 1997, 2002 and 2007. Eligible children were between the ages of 0 and 12 in 1997, at the time of the first survey. These surveys include a broad array of developmental outcomes as well as information on the home environment of the child. The PSID-CDS is particularly well-suited for this study since it provides information about mothers' mental health together with information about the quantity of mother-child interactions and mothers' labor market outcomes. Therefore, the data set allows me to relate the mother's mental health to maternal investments in the child.

From the main PSID survey, I collect data on mothers' labor supply decision, labor income, total family income and relevant demographic variables from the year the child was born until she reaches 16 years old. This data collection goes as far as 1985 and as recent as 2013. From the CDS, I collect information on the child's cognitive ability, on the mother's mental health and on the mother's time with the child. This data is collected from the three CDS surveys in 1997, 2002 and 2007. In constructing my analytic sample, I keep respondents with valid information on the child's cognitive test score, mother's labor supply, mental health and time with her child. I drop individuals with missing information on the child's race, gender, birth-order, as well as those with missing information on the mother's education and age at the child's birth. The resulting analytic sample has information on 2,459 children and their mothers.

I measure the child's cognitive development with the Letter-Word (LW) module of the Woodcock-Johnson aptitude test. The Letter-Word Identification test assesses symbolic learning and reading identification skills. The test is ideal as it can be administered to children between the ages of 3 and 17 and as a result most children were eligible for the test in two CDS surveys.

I use the child's time diary to measure the time the mother spends with her child. This is a distinctive feature of the PSID-CDS. The CDS asks participant children, or their primary caregivers, to record a detailed, minute by minute timeline of their activities for two days of the week: one random weekday and one random weekend day. Activities were coded at a fine level of detail.

From this data, I construct a measure of maternal time investments by taking a weighted sum ($\frac{5}{7}$ for the weekday and $\frac{2}{7}$ for the weekend) of the total hours in which the mother is recorded as **actively participating** with the child in each diary activity. Active participation can be thought of as a measure of maternal engagement with the child.

1.2.2 Psychological Distress

I use the *Kessler 6 (K6) Psychological Distress Scale* (Kessler et al., 2002) to measure the mother's mental health. The K6 scale is a simple and widely used measure of general psychological distress.[5] Psychological distress is largely defined as a state of emotional suffering characterized by symptoms of depression (e.g. lost interest, sadness, hopelessness) and anxiety (e.g. restlessness, feeling tense) (Mirowsky and Ross, 2003; Drapeau, Marchand, and Beaulieu-Prévost, 2011).

The K6 scale involves asking 6 questions about the individual's emotional state in the previous four weeks. Each individual is asked 'in the last 4 weeks, about how often did you feel': 1) nervous, 2) hopeless, 3) that everything was an effort, 4) so sad that nothing could cheer you up, 5) worthless, and 6) restless or fidgety. Each question is scored on a scale of five values (0-4), where 4 indicated "All of the time" and 0 indicated "None of the time".[6].

[5]Other scales have also been developed with the intent to measure psychological distress. Other examples are the General Health Questionnaire (GHQ), the Kessler K10 scale and the Brief Symptom Inventory (BSI).

[6]I use a simple item response theory (IRT) approach to construct a continuous measures that controls for measurement error in the six responses . I discuss this approach in detail in Section 1.4.4 In comparison, the usual approach is to sum the scores on the six questions and use cut points to separate individuals in three levels of distress. As a general rule, a cut point of 13+ is used as the optimal cut point for assessing the prevalence of serious mental disorder in the national population (Kessler et al., 2010). A cut point between 5 and 8 can

The prevalence rate of psychological distress is non-trivial for the U.S. adult population. Psychological distress is usually measured on a continuous scale, but more often than not individuals are classified into three groups: those suffering from moderate psychological distress, those suffering from serious psychological distress, and those under no distress. Moderate levels of distress are very common, with a prevalence rate of 20-30% for the U.S. adult population.[7] Serious psychological distress is much rarer, with a prevalence rate of about 3% for the U.S. population.

In spite of being quite common, psychological distress can lead to serious life impairments. Individuals in serious distress report lower productivity in the home and in the labor market, and problems in interactions with friends and family members. Individuals with moderate levels of distress suffer similar impairments but at a lesser rate. For instance, 85% of individuals under serious distress report facing some work impairment, while about 60% of individuals under moderate distress report the same (Prochaska et al., 2012).

The prevalence of psychological distress is fairly constant across geographical regions, but there are important group differences (Drapeau, Marchand, and Beaulieu-Prévost, 2011). In particular, the prevalence of psychological distress is higher for women than for men, and peaks during early adulthood (18-29 years old).[8]

---

also be used to indicate a moderate mental disorder (Prochaska et al., 2012; Herrick, 2015). This separation is often used to analyze the prevalence rate of psychological distress in the population.

[7]These numbers depend on the cut-off being used.

[8]Also, there are no significant differences in prevalence across races or ethnic groups, but the prevalence is higher for immigrants (Nemeroff, Midlarsky, and Meyer, 2010; Drapeau, Marchand, and Beaulieu-Prévost, 2011).

1.2.3 Summary Statistics

In this section, I discuss the most important patterns in the data. I first demonstrate that both child cognition and maternal distress are highly correlated with family income, a result that motivates this paper. Next, I describe the other key variables: the mother's time investments in the child, hours of work and her wage offer.

This paper is motivated by the fact that maternal mental health is strongly correlated with both socioeconomic status and child cognition and, as a result, can be thought of as a mediator of the intergenerational transmission of human capital. Figure 1.1(a) plots the association between psychological distress and family income. The pattern is striking. Individuals in the lower end of the income distribution face much higher levels of distress (1 s.d. higher) than individuals with high levels of income. Moreover, this negative relationship is stronger at lower levels of income, suggesting that psychological distress is strongly related to financial strain and poverty.[9]

Similarly, child cognition is also highly correlated with family income. This can be seen in Figure 1.2(a), which plots average standardized letter-word score for different percentiles of family income. This gap in cognitive skills can be as large as one standard deviation. Moreover, this gap tends to grow over time, as can be seen in Figure 1.2(b). The gap doubles from .6 to 1.2 points of a standard deviation from age 3 to age 15. One of the goals of this paper is to

[9]Figure 1.1(b) plots the density distribution of psychological distress for the mothers in my sample. There is a very clear clustering of scores around zero. This is due to the fact that in my sample about 15% of the mothers respond "none of the time" to all 6 questions in the K6 scale. As a contrast, only 6% of the individuals respond "all of the time" to *any* question and only 4 individuals respond "all of the time" in all 6 questions.

explore the role of maternal mental health in explaining this gap.

Besides maternal mental health, family time and goods investments can also explain this gap and, as a result, are part of my model. Table 1.1 provides summary statistics on these variables. On average, mothers spend 19 hours per week engaged in activities with their children. However, as can be seen in Figure 1.3(b), there is large variation in time investments across child ages. Mothers spend more than double the amount of time with young children than with teenagers. Similar patterns can be found for maternal labor force participation. On average, mothers spend 1,209 hours working every year; their labor force participation is lower in the first years of the child's life and increases steadily as the child ages, as can be seen in Figure 1.3(c). Perhaps due to human capital accumulation and depreciation through work experience, mothers' wages decrease when children are young, when mothers take time off from the labor market, and increase steadily over time, as mothers accumulate labor market experience. This can be seen in Figure 1.3(d).

One important thing to notice in Table 1.1 is that I do not observe all variables at all ages for each child. For example, I only observe the letter-word score for 4,582 child-age observations, close to two observations per children. Similarly, I only observe 25,795 observations for mothers' labor force participation, about 10 observations per mother. However, I do observe these same variables at all ages for at least some children, as can be seen in Figures 1.3(a)-1.3(d). As a result, I can construct moments that will allow me to estimate the model proposed in the next two sections. That is, these patterns in the data motivate the method of simulated moments estimation approach described in detail in Section 1.4.5.

1.2.4 Preliminary Results

In my main econometric analysis, I control for endogeneity of maternal mental health using variation in state mental health parity laws. Moreover, I simulate maternal investments and jointly estimate these with the child's cognition production function. However, for the preliminary analysis conducted here, I assume maternal mental health is exogenous and explore its correlation with child cognitive development and its correlation with other relevant maternal investments. These preliminary results serve to illustrate important patterns in the data and to demonstrate that my main results are not driven by my estimation strategy.

I start by estimating a static model of children's cognitive development. The main outcome of interest are age-standardized logged letter-word scores. I estimate OLS regressions of the following form:

$$log(A_{it+5}) = log(H_{it})\phi_1^H + log(MT_{it})\phi_1^M + log(Inc_{it})\phi_1^I + X_{i1}\phi_1^X + \epsilon_{i1} \quad (1.1)$$

where the letter word score for individual i at time t is given by A_{it}, and H_{it}, MT_{it}, Inc_{it} refers to the mother's psychological distress, maternal time investments and family income respectively. X_{i1} is a vector of covariates and ϵ_{i1} is a normally distributed disturbance.[10]

Estimates for equation 1.1 are presented in Table 1.2 for varying sets of covariates X_{i1} and family investments. Column [1] displays the raw relationship between maternal psychological distress and children's cognition. A ten percent increase in maternal psychological distress is related to a decrease of about

[10] I use children's scores at $t + 5$ in order to capture the idea that today's investments determine future cognitive skills. The analysis with cognitive scores measured at t yield qualitatively similar results.

1.2 percentage points in children's cognitive skills. This relationship is also plotted in Figure 1.4(a). In column [2], I add family controls, such as the mother's education. Including these controls decreases the magnitude of the association with maternal psychological distress by about a half. This illustrates the strong endogeneity problem due to unobserved investments. In column [3] and [4], I include my measure of maternal time investments and family income respectively. Including these variables further decreases the magnitude of the association with the mother's distress by about 10%. In the end, it looks that maternal mental health is 70% as important in determining child cognition as family income. These results, however, should be taken as correlations since they assume investments are exogenous and ignore the dynamic nature of child development.

Given that time and goods investments are important determinants of children's skills, I would like to understand the correlation of maternal distress with the determinants of these investments. In order to do so, I estimate OLS regressions of the following form:

$$Y_{it} = log(H_{it})\phi_2^H + X_{i2}\phi_2^X + \epsilon_{i2} \tag{1.2}$$

where Y_{it} measures weekly maternal time investments, annual hours at work or log hourly wages for individual i at time t. As before, H_{it} refers to the mother's psychological distress, X_{i2} is a vector of covariates and ϵ_{i2} is a normally distributed disturbance.

Estimates for equation 1.2 are presented in Table 1.3 for varying sets of covariates X_i. Figure 1.4 also plots the relationship between family investments and psychological distress using linear polynomials. Column [1] displays the raw relationship between maternal psychological distress and weekly maternal time investments. Each percent decrease in maternal psychological distress is related to an increase of 0.825 hours in maternal time investments. These are relatively large associations as can be seen in Figure 1.4(b). However, this relationship decreases by a third when I add family controls, such as the mother's education (Column [2]). Similar effects can be seen for the effect of distress on the mother's labor supply (Columns [3] and [4] and Figure 1.4(c)). Each percentage increase in maternal distress is associated with a decrease of about 40 hours worked in a year. Perhaps related, psychological distress is also strongly correlated with labor market productivity. As can be seen in Columns [5] and [6] and Figure 1.4(d), a ten percentage increase in psychological distress is associated with a five percentage decrease in hourly wages.

The results in Tables 1.2 and 1.3 and Figure 1.4 provide preliminary evidence that maternal mental health matters for children's cognitive development. Moreover, they provide suggestive evidence that maternal mental health can affect children through its effect on other family investments. However, these results have several shortcomings. They do not control for endogeneity in the child cognition production function as a result of unobserved investments (omitted variable bias). Moreover, they ignore the fact that child development is a dynamic process and that investments interact in non-obvious ways in determining child outcomes. The model developed in the next couple of sections takes these issues seriously. Moreover, it formally describes the different channels through which maternal mental health can affect children.

1.3 A Model of Cognitive Skills Formation

In this section, I describe a standard model of maternal investments in children's cognitive development. This model allows me to distinguish the channels through which maternal mental health can affect children. I start by describing the technology of cognitive skill formation in children. I argue that the child's cognitive skills are determined over multiple periods by mother's time and goods investments as well as by the mother's mental health. I then discuss the determinants of these maternal investments. I show that maternal time and goods investments are determined by the mother's preferences, the constraints she faces and her productivity in the labor market, and that maternal mental health can affect these investments. At the end of this section, I comment in some detail on the different channels through which maternal mental health can influence the child's cognitive skills formation. Throughout this section, I treat mental health as exogenous and address the endogeneity problem later, in Section 1.4.3.

It is worth noting at this point that the ideas developed in this section are mainly used to motivate the empirical model described later in Section 1.4. In that model, I will approximate the maternal time allocation decisions with policy functions and estimate these jointly with the child's skill production function and the mother's wage offer equation, where the technology of skill formation is estimated in its structural form. I will argue that this approach provides some advantages over fully structural estimation and that this approach is closely related to what has been done in the literature (see Cunha, Heckman, and Schennach (2010) and Agostinelli and Wiswall (2016) for two examples).

1.3.1 Cognitive Skill Formation

Child cognition is determined over multiple periods ($t \in \{0, 1, ..., 16\}$). Each period is equivalent to a year in the child's life. The model starts when the child is born ($t = 0$) and ends when she reaches age 16 and can leave the household. The child's stock of cognitive skills (A_t) is determined at the beginning of every period. In the initial period, the child is born with an initial ability stock (A_0), which is determined by genetic and in-utero investments. At each subsequent period, the mother determines her child's skill evolution by allocating time (MT_t) and goods investments (G_t) for the child. This is a common assumption in the literature (see Becker and Tomes (1986) as an early example). Expanding on the literature, I also allow the mother's mental health (H_t) to influence the child's accumulation of skills. Formally, the child's skills evolve as follows:

$$A_{t+1} = f_t(A_t, G_t, MT_t, H_t, \eta_t) \tag{1.3}$$

where η_t captures shocks and unobserved inputs that affect the child's development. The technology $f_t(\cdot)$ is allowed to change as the child ages in order to capture different stages of development.

The specification of the technology of skill formation ($f_t(\cdot)$) should take into account two important features of child development: dynamic and static complementarities of investments. Dynamic complementarity suggests that the returns to current investments depend on the child's current ability ($\frac{\partial^2 A_{t+1}}{\partial A_t \partial I_t} \neq 0$). As a result, returns to current investments will depend on past investments in the child ($\frac{\partial^2 A_{t+1}}{\partial I_{t-1} \partial I_t} \neq 0$) (see (Cunha and Heckman, 2007) for a thorough discussion). Moreover, static complementarity suggests that the technology should allow for the returns to current investments to depend on other investments, e.g.

$\frac{\partial^2 A_{t+1}}{\partial MT_t \partial H_t} \neq 0$. This second feature is especially important when incorporating maternal mental health.

Maternal mental health can enter the human capital production function in two ways. First, maternal mental health can be thought as a 'direct' component of the child's human capital production function in the same way as financial investments or maternal time investments. One explanation is that children suffer from their parents' psychological distress and in turn develop psychological problems of their own (Rosenquist, Fowler, and Christakis, 2011; Eisenberg et al., 2013; Ross, 2000). In turn, psychological problems inhibit children's cognitive functions such as planning and attention leading to further developmental problems (Blair, 2010; Blair et al., 2011). Second, maternal mental health can influence the productivity of maternal time investments. The idea comes from the family stress theory in sociology, which proposes that maternal psychological distress can provoke harsh, inconsistent and low nurturing parenting (Conger et al., 1994, 2002). In a sense, the idea is that the mother's mental health is an important determinant of the quality of maternal time investments. The technology of skill formation should take into account these two different mechanisms.

In order to accommodate these features, I assume the technology of skill formation follows the translog (transcendental logarithmic) specification. Perhaps the most obvious alternative approach would be to follow Cunha, Heckman, and Schennach (2010) and assume child development is described by a CES production function. The CES is appealing because it contains both the Leontief and the Cobb-Douglas functions in the limit as the complementarity parameter approaches $-\infty$ or 0. Moreover, the CES specification allows mental health to have a ‘direct’ effect on children’s development. It also allows for maternal time investments and mental health investments to be complements in the child’s human capital production function. However, the CES is problematic because it assumes identical elasticities of substitution between all input factors. This restriction is limiting as it does not allow the estimation of the separate effects of the mother’s mental health on the productivity of her time with the child. On the other hand, the translog allows that.[11] Using the translog specification, I write the technology of skill formation as:

$$
\begin{aligned}
ln(A_{t+1}) = ln(K_t) &+ \alpha_{1t}ln(A_t) + \alpha_{2t}ln(G_t) + \alpha_{3t}ln(MT_t) + \alpha_{4t}ln(H_t) \\
&+ \alpha_{5t}ln(A_t)ln(G_t) + \alpha_{6t}ln(A_t)ln(MT_t) + \alpha_{7t}ln(A_t)ln(H_t) \qquad (1.4) \\
&+ \alpha_{8t}ln(G_t)ln(MT_t) + \alpha_{9t}ln(G_t)ln(H_t) + \alpha_{10t}ln(MT_t)ln(H_t) + \eta_t^a
\end{aligned}
$$

where K_t corresponds to the total factor productivity of investments.

The translog is a generalization of the Cobb-Douglas, which is the special case where the interaction parameters are all zero ($\alpha_{jt} = 0 \,\forall j \in \{5, 10\}$). These same interaction parameters allow for non-constant elasticity of substitution between inputs, which is not allowed in the Cobb-Douglas, and for different partial elasticities of substitution between inputs, which are restricted in the CES.[12] α_{5t},α_{6t} and α_{7t} capture the degree of dynamic complementarity (beyond the one implied by the Cobb-Douglas), where early investments are allowed to

[11] Another alternative is a Nested CES production function. In a Nested CES two inputs, maternal mental health and time investments, are combined in a CES production function, which is then nested in a further CES production function which includes goods investments and the child’s original human capital. I find qualitative similar results when I use a Nested CES production function.

[12] The translog function could be expanded to include additional terms to provide an approximation to any unknown production technology.

influence the returns of today's investments (Cunha, Heckman, and Schennach, 2010; Aizer and Cunha, 2012). For example, $\alpha_{5t} \geqslant 0$ implies that $\frac{\partial^2 A_{t+1}}{\partial A_t \partial G_t} \geqslant 0$.[13] Similarly, α_{10t} describes the elasticity of substitution between maternal mental health and maternal time investments. If $\alpha_{10t} = 0$ the elasticity of substitution between maternal mental health and maternal time investments equals the one implied by the Cobb-Douglas specification.

It is also important to note that all parameters are subscripted by t. Following Cunha, Heckman, and Schennach (2010), I assume there are two stages of development, ages 0-5 and ages 6-16. There are many reasons for this distinction. For one, at age 6, the child enters formal schooling and as a result is no longer exposed to only the home environment. Also, the interpretation of the returns of maternal time changes at age 6. Both the types of activities the mother engages with the child and the types of activities the child engages without the mother changes once the child enters formal schooling. As a result, we should expect the return of maternal time to be different across developmental stages. The same is true for the other investments.

Since the child initial ability A_0 is unobserved, I also need to make some assumptions on how it is realized. I assume A_0 is a function of the mother's and child's observed characteristics at the child's birth (X_0^a), such as the mother's education and age, and the child's race, gender and birth weight. Formally:

$$ln(A_0) = X_0^a \alpha_0^x + \eta_0^a \tag{1.5}$$

where α_0^x picks up the idea that in-utero investments and children's genetic endowments differ by family types.

[13]The Cobb-Douglas imposes dynamic complementarity of investments, so even if $\alpha_{5t} < 0$ it is possible for dynamic complementarity to be present.

1.3.2 Maternal Investment Decisions

As described in the previous section, the child's cognition is determined by mother's time (MT_t) and goods (G_t) investments. These investments are determined by the mother's time allocation decisions. That is, in every period, the mother rationally chooses the amount of hours to spend in the labor market and the amount of hours to spend with the child in the form of time investments. She takes into account how these decisions affect her own and the child's human capital accumulation. By working more hours, the mother accumulates labor market experience, which will influence her future earnings potential. Similarly, by spending quality time with her child she improves the child's human capital stock. Her decision depends on both her preferences and constraints.

Preferences

In every period t, the mother chooses $d_t = (HW_t, MT_t)$, where HW_t represents the choice for annual hours of work and MT_t represents the choice for hours engaged with the child in cognitive productive activities. A woman's preferences over the choice set is defined by her period utility function. Her period utility depends on her current mental health status H_t and observed individual characteristics X_t^u. The utility function is separable across consumption (C_t), leisure (L_t) and the child's human capital (A_t).

$$
\begin{aligned}
U(C_t, L_t, A_t; H_t, X_t^u) = & \ \lambda_c(H_t, X_t^u) f_c(C_t) \\
& + \lambda_l(H_t, X_t^u) f_l(L_t) \qquad (1.6) \\
& + \lambda_a(H_t, X_t^u) f_a(A_t)
\end{aligned}
$$

The function $\lambda_c(.)$ allows the marginal utility of consumption to vary with the mother's mental health status as well as observable characteristics such as her education and age. Mental health enters $\lambda_c(.)$ in order to capture the idea that individuals in poor mental health receive different enjoyment from consumption than individuals in a good mental health state. Similarly, mental health enters $\lambda_l(.)$ as a result of the fact that individuals suffering from mental illnesses are more likely to spend time out of the labor market and miss days of work, and thus could have a higher cost of working (Frijters, Johnston, and Shields, 2014). It is less obvious but also possible that mental health could influence how mothers value their children's human capital development $\lambda_a(.)$.

Constraints

The model assumes women face two constraints in every period, a budget constraint and a time constraint. The budget constraint is given by:

$$
C_t + G_t = Inc_t = w_t HW_t + N_t + B(\tau_{st}, w_t, HW_t, N_t) \qquad (1.7)
$$

where G_t corresponds to the income share that is spent on the child as goods investments, Inc_t is the total family income, N_t is the part of income that does not depend on the woman's labor supply and includes for example the husband's labor income if the woman is married, family transfers and gifts, and $B(\tau_t, w_t, HW_t, N_t)$ are government transfers received by the family such as food stamps, welfare benefits and earned income tax credits. Government transfers are assumed to depend on state-year welfare rule parameters τ_{st}, the wage rate w_t, hours of work HW_t and other family income N_t. [14] [15] [16]

One assumption that is commonly made in the literature and that I follow here is that all families spend an equal and fixed proportion of their income on their child in the form of goods investments. That is, $G_t = a \times Inc_t$. This assumption is necessary since many goods investments, such as the quality of the child's toys, the number of books she has access to and whether she has access to a computer, are usually unobserved or hard to quantify monetarily.

The time constraint is given by:

$$L_t = TT - HW_t - MT_t \tag{1.8}$$

where TT is the total time available for the women in a year and leisure will depend on how many hours are left after taking into account the number of hours spent in the labor market and the number of hours spent interacting with the child.

[14] As was noted by Moffitt (1983), many women who are eligible for welfare benefits based on their income do not collect them. The model explicitly ignores the welfare participation decision. This is to keep the model simple and tractable.

[15] It is important to note that these welfare rules should affect individuals differently depending on their previous welfare participation. For example, work requirements might be binding for some individuals but not for others depending on the age of their youngest child and on their previous labor force and welfare participation. The model ignores these important dynamics.

[16] Welfare rule parameters (τ_{st}) provide important exclusion restrictions as they influence the woman's decisions but do not affect her labor market productivity directly and only enter the child's human capital production function through the family income. I explain this identification argument in more detail in Section 1.4.3.

The Wage Process

The mother's labor market productivity determines the budget constraint she faces as well as the amount of monetary resources to be invested in the child for any given time allocation decision. As a result, both goods and time investments received by the child should depend on the mother's labor market productivity.

The wage process takes into account the women human capital accumulation through work experience, or learning by doing. The wage offer at each period is assumed to be determined by the woman's observable characteristics (X_t^w), which include her age, race and education. It is also assumed to depend on her experience stock at the beginning of the period (EX_t) , her employment decision in the previous period (HW_{t-1}), her mental health state in the current period (H_t) and local labor market conditions (ζ_{st}) in her state of residency (s). That is:

$$ln(w_t) = X_t^w \beta_x^w + \beta_1 H_t + \beta_2 EX_t + \beta_3 \mathbb{1}[HW_{t-1} = 0] + \zeta_{st}\beta_s^w + \eta_t^w \qquad (1.9)$$

where β_x^w allows the model to capture returns to education and possible labor market discrimination based on the woman's race, and β_1 captures the idea that mental health disorders are associated with a loss in productivity in the labor market, leading to lower wages and a higher probability of being unemployed (Ettner, Frank, and Kessler, 1997). The third term in Equation 1.9 captures the labor market returns to human capital accumulation through work experience, and β_3 captures the temporary labor market penalty for spending time out of the labor market. That is, the dynamic wage process allows for endogenous state dependence through human capital accumulation and the dependence of the current wage offer on the woman's previous work choice. β_s^w is a vector that translates labor market conditions (ζ_{st}) into offered wages.[17] Work experience accumulation is determined by the following process: $EX_{t+1} = EX_t + HW_t$.

Value Functions

The solution for the mother's time allocation decision can be derived from the value functions implied by the model. That is, let Ω_t be the state space faced by the mother that arises from her decisions made up to period t then the mother's optimal time allocation choice in period t is given by:

$$\{HW_t, MT_t\} = \arg\max\{V_t(\Omega_t)^1, ...V_t(\Omega_t)^J\}$$

Where the utility of choice j for individual i at any period is given by:

$$V_t(\Omega_t)^j = U_t^j(C_t, L_t, A_t|d_t = j, \Omega_t) + \beta\mathbb{E}[V_{t+1}(\Omega_{t+1})|d_t = j, \Omega_t]$$

where her choice ($d_t = \{HW_t, MT_t\}$) will depend on the state space she faces in period t (Ω_t) as well as on her beliefs on the state space evolution given her choices $\mathbb{E}[\Omega_{it+1}|d_t = j, \Omega_t]$.

1.3.3 Mental Health Mechanisms

The relationship between maternal mental health and child cognitive development can be represented through five key mechanisms. I discuss these different pathways below.

The first mechanism corresponds to the direct effect of maternal mental health on children's human capital accumulation. This mechanism is captured

[17] State variation in labor market conditions (ζ_{st}) are important for the identification of the empirical model described in Section 1.4.

by α_{4t} in Equation 1.4. The direct mechanism can be thought as the effect maternal mental health has on children that is not captured by the other channels. Theoretically, one possible explanation is contagion of mental health, where children suffer from their parents' psychological distress and in turn develop psychological problems of their own (Rosenquist, Fowler, and Christakis, 2011; Eisenberg et al., 2013; Ross, 2000). Higher stress inhibits planning, emotional control and attention, and as a result can lead to cognitive developmental problems (Blair, 2010; Blair et al., 2011). This channel also has dynamic implications. First, parents might increase investments in their child in the current period as a way to compensate for this decrease in human capital. Second, due to dynamic complementarity, a decrease in current human capital could affect the returns of family investments in subsequent periods.

The second mechanism corresponds to the effect mental health has on the productivity of maternal time investments. That is, this mechanism is related to a change in the quality of these investments. This idea comes from the family stress model in sociology, and suggests that a distressed mother can lose her ability to be supportive and to interact in a consistent manner with her child. This decrease in quality of mother-child interactions results, in turn, in fewer learning experiences for the child (McLoyd, 1990; Mayer, 2002). The model captures this channel with the parameter α_{10t} in Equation 1.4. This parameter captures the degree of complementarity between maternal mental health and maternal time investments, and as a result, captures how the returns to maternal time investments change with the mother's mental health status. A high degree of complementarity between these two inputs implies that the value of maternal time investments is much higher for mothers in good mental health when compared to those in poor mental health.[18]

The effect mental health on the value of leisure leads to two other mechanisms. That is, the effect of maternal mental health on the quantity of maternal time investments and on her labor force participation. In the model, this effect is described by the marginal utility parameter $\lambda_l(H_t, X_t^u)$ in Equation 1.6. The idea is that mental health problems can influence impulse, attention and emotional control, so that spending consistent time in productive activities, such as time in the labor market or engaged with the child, becomes more costly (Blair, 2010; Frijters, Johnston, and Shields, 2014). This increase in cost is captured by an increase in the marginal utility of leisure and will result in a reduction in investments in the child. An increase in leisure implies either a decrease in monetary investments due to lower labor force participation or a decrease in time investments. A reduction in labor force participation will also reduce the mother's human capital accumulation.[19]

These reductions in the woman and her child's human capital also have dynamic implications. A decrease in the mother's experience capital can lead to a decrease in her future labor market productivity, as captured by β_2 in Equation 1.9. This, in turn, leads to lower resources available in the future to be invested in the child. Similarly, a decrease in the child's human capital will influence the returns of future family investments. This comes from the idea of dynamic complementarity, where the returns of current investments depend on the amount of past investments received by the child (see Aizer and Cunha

[18] The same could be true about the complementarity between maternal mental health and family income. It is possible that financial investments in the child are more productive for mothers in good health. This would be captured by α_{9t} in Equation 1.4.

[19] Similarly, mental health can affect the value the mother places on consumption (λ_c) and on the child's human capital development (λ_a), also influencing her investment decisions.

(2012) and the discussion in Section 1.3.1).

The fifth and last mechanism corresponds to the mental health effect on the mother's productivity in the labor market (Ettner, Frank, and Kessler, 1997). This effect is captured by β_1 in Equation 1.9. A reduction in the mother's labor market productivity, conditional on hours worked, implies a reduction in resources to be invested the child. This mechanism is especially important for single women, who are the sole bread-winner in the household. This channel also has dynamic implications due to an ambiguous effect on the mother's labor force participation.

1.4 Empirical Strategy

This Section describes the estimation strategy used in the paper. I start by describing how I approximate mothers' time allocation decision rules with policy functions. I also discuss the benefits and costs of this approach. I then move to explore the main threats to estimation - measurement error and endogeneity of inputs - and how I handle these issues. At the end of the section, I describe the method of simulated moments (MSM) procedure that I use to estimate the empirical model.

1.4.1 Approximation to the Decision Rules

The empirical strategy involves approximating maternal time allocation decisions with policy functions and estimating these jointly with the child's technology of skill formation and the mother's wage offer. This approach is similar to that of other papers in the literature (see (Cunha, Heckman, and Schennach,

2010) and (Agostinelli and Wiswall, 2016) for two examples). The alternative approach would be to fully estimate the dynamic model described in the previous section. That would allow me to estimate the preferences parameters described in Equation 1.6. However, it would require me to make explicit assumptions regarding the mother's knowledge of her child's skills and of the technology of skill formation. Moreover, it would require me to make strong assumptions regarding maternal investments in other children in the household, or to restrict my sample to single child families.

As explained in Section 1.3.2, the mother's choices in time t ($d_t = \{HW_t, MT_t\}$) will depend on the whole state space she faces in period t (Ω_t) and on her beliefs on the state space evolution given her choices $\mathbb{E}[\Omega_{it+1}|d_t = j, \Omega_t]$. The specific form of the policy functions for the mother's time allocation decision will depend on how one specify the mother's preferences as well as the mother's knowledge about both her child's ability and the technology of cognitive skill formation. However, without taking a stance on these issues, we could write the policy functions for the mother's time allocation as a general function of the state space faced by the mother in period t. This approach accommodates most models of maternal behavior. That is, we can write the policy functions for the mother's time allocation decisions as:

$$HW_t = f^{hw}(\Omega_t) + \eta_t^{hw} \tag{1.10}$$

$$MT_t = f^{mt}(\Omega_t) + \eta_t^{mt} \tag{1.11}$$

where η_t^{hw} and η_t^{mt} capture shocks to the mother's decision.

This estimation approach proposed has some clear advantages and disadvantages. There are two main advantages. First, it avoids making strong assumptions about the investment process and can approximate multiple models of household behavior. For example, it avoids making assumptions on the mother's knowledge about the technology of skill production.[20] These assumptions can heavily influence policy simulation exercises. Second, it allows me to estimate the model for households with multiple children by allowing the mother's decisions to depend linearly on the family composition. This is not possible on a fully structural model. As a matter of fact, since allocation of investments across all children in a household is rarely observed in data, most papers that try to recover individual preferences have focused on one-child families (see Bernal (2008); Griffen (2012); Brilli (2014) for examples).[21]

The main disadvantage of the proposed empirical strategy is that it does not allow me to recover deep utility parameters from the model (Equation 1.6). This can be problematic in counterfactual policy analysis as I cannot estimate the effect of policies on mothers' preferences. Moreover, it does not allow me to estimate the effect of the mother's mental health on the these preferences. For example, the overall effect of mental health on labor force participation in Equation 1.10 captures both the effects of mental health on the marginal value of leisure and consumption as well as its effects on the mother's wages and the child's cognition, and how these affect the mother's labor force participation.

[20] As a matter of fact, Cunha (2013) provides evidence that mothers have biased beliefs about the production function of child skills.

[21] One exception is Del Boca, Flinn, and Wiswall (2014) which allows for both one-child and two-child families.

Linear Policy Functions

Ideally I would like to estimate the policy functions nonparametrically as shown in Equations 1.10 and 1.11. However, given the large state space, large number of parameters (100+) and the number of observations ($\sim 2,500$), for computational and identification reasons, I assume that the policy functions are linear-in-parameters.

The state space is composed of many different variables described in the conceptual model. There are three state variables that evolve endogenously in the model: labor market experience (EX_{it}), the history of hours in the labor market ($\{HW_{iz}\}_{z=0}^{t}$) and the history of maternal time investments ($\{MT_{iz}\}_{z=0}^{t}$). These variables determine the wage offer received by the mother and the child's ability in period t (see Equations 1.4 and 1.9). There are also exogenous state variables that are fixed over time or evolve exogenously from the model. These include exogenous variables that determine the child's initial ability (X_{i0}^{a}), exogenous variables that determine the wage offer (X_{it}^{w}), exogenous variables that enter the flow utility function (X_{it}^{u}) and the mother's mental health (H_{it}), see Equations 1.5, 1.9 and 1.6. Moreover, it includes state level variation in welfare rules (τ_{st}) and state variation in labor market conditions (ζ_{st}) that determine family income and hourly wages. The state space can be characterized by: $\Omega_{it} = \{EX_{it}, \{MT_{iz}\}_{z=0}^{t}, \{HW_{iz}\}_{z=0}^{t}, \chi_{it}\}$, where $\chi_{it} = \{H_{it}, X_{it}^{w}, X_{i0}^{a}, X_{it}^{u}, \zeta_{st}, \tau_{st}\}$ is the vector of exogenous state variables.

As a result, the linear-in-parameters policy functions for the mother's time allocation decision can be described by: [22]

$$
\begin{aligned}
HW_{it}^{*} =& \gamma_0^h + \gamma_1^h EX_{it} + \gamma_2^h \mathbb{1}[HW_{it-1} = 0] + \gamma_3^h HW_{it-1} + \gamma_4^h MT_{it-1} \\
& + \gamma_5^h H_{it} + X_{it}^{w}\gamma_{xw}^{h} + X_{i0}^{a}\gamma_{xa}^{h} + X_{it}^{u}\gamma_{xu}^{h} + \zeta_{st}\gamma_6^h + \tau_{st}\gamma_7^h + \eta_{it}^{hw} \qquad (1.12)
\end{aligned}
$$

$$
HW_{it} = \begin{cases} HW_{it}^{*} & \text{if} \quad HW_{it}^{*} \geqslant 0 \\ 0 & \text{if} \quad HW_{it}^{*} < 0 \end{cases}
$$

where η_{it}^{hw} captures shocks to the mother's decision. The policy function for maternal time investments is assumed to follow the exact same structure.

1.4.2 Empirical Model

The empirical strategy constitutes estimating the policy functions described above jointly with the child's technology of skill formation and the mother's wage offer. As a result, the empirical framework can be summarized by the

[22] In order to specify the expectation over the evolution of the state variables, these approach needs two assumptions. First, I assume that the mother's decision in the previous period ($\{HW_{it-1}, MT_{it-1}\}$) is a sufficient statistic for the whole history of decisions up to the last period ($\{HW_{iz}, MT_{iz}\}_{z=0}^{t}\}$). This is required for tractability. Otherwise, I would have to re-write the child human capital production function so as to reduce the state space (see (Bernal and Keane, 2010) as an example). Second, I assume that current state level variables ($\{\zeta_{st}, \tau_{st}\}$) are sufficient statistics for future changes in state level conditions.

following system of equations:

$$\begin{aligned}
ln(A_{t+1}) &= ln(K_t) + \alpha_{1t} ln(A_t) + \alpha_{2t} ln(G_t) + \alpha_{3t} ln(MT_t) + \alpha_{4t} ln(H_t) \\
&+ \alpha_{5t} ln(A_t) ln(G_t) + \alpha_{6t} ln(A_t) ln(MT_t) + \alpha_{7t} ln(A_t) ln(H_t) \\
&+ \alpha_{8t} ln(G_t) ln(MT_t) + \alpha_{9t} ln(G_t) ln(H_t) + \alpha_{10t} ln(MT_t) ln(H_t) + \eta_t^a \\
ln(A_0) &= X_0^a \alpha_0^x + \eta_0^a \\
ln(w_t) &= X_t^w \beta_x^w + \beta_1 H_t + \beta_2 EX_t + \beta_3 \mathbb{1}[HW_{t-1} = 0] + \zeta_{st} \beta_s^w + \eta_t^w \qquad (1.13) \\
HW_{it}^* &= \gamma_0^h + \gamma_1^h EX_{it} + \gamma_2^h \mathbb{1}[HW_{it-1} = 0] + \gamma_3^h HW_{it-1} + \gamma_4^h MT_{it-1} \\
&+ \gamma_5^h H_{it} + X_{it}^w \gamma_{xw}^h + X_{i0}^a \gamma_{xa}^h + X_{it}^u \gamma_{xu}^h + \zeta_{st} \gamma_6^h + \tau_{st} \gamma_7^h + \eta_{it}^{hw} \\
MT_{it}^* &= \gamma_0^{mt} + \gamma_1^{mt} EX_{it} + \gamma_2^{mt} \mathbb{1}[HW_{it-1} = 0] + \gamma_3^{mt} HW_{it-1} + \gamma_4^{mt} MT_{it-1} \\
&+ \gamma_5^{mt} H_{it} + X_{it}^w \gamma_{xw}^{mt} + X_{i0}^a \gamma_{xa}^{mt} + X_{it}^u \gamma_{xu}^{mt} + \zeta_{st} \gamma_6^{mt} + \tau_{st} \gamma_7^{mt} + \eta_{it}^{mt}
\end{aligned}$$

where goods investments are assumed to be determined by a fixed proportion of family income ($G_{it} = a \times Inc_{it}$), and as a result I substitute family income for goods investments in the empirical model. Moreover, family income is assumed to be determined by: $Inc_{it} = HW_{it} w_{it} + N_{it} + B_{it}$, where Inc_t is the total family income, N_{it} is the part of income that does not depend on the woman's labor supply and B_{it} are government transfers received by the family.

This empirical model allows me to capture most of the mechanisms described in Section 1.3.3. α_{4t} captures the direct effect the mother's mental health has on children. One explanation is that it picks up contagion of mental illnesses

(Rosenquist, Fowler, and Christakis, 2011). The effect of mental health on the productivity of maternal time investments is captured by the complementarity parameter α_{10t}. Similarly, α_{7t} and α_{9t} captures possible complementarities between the mother's mental health and the child's skill and family income respectively. Moreover, β_1 captures the effect mental health has on labor market productivity, as discussed in Section 1.3.2. All these parameters are capturing the deep parameters in the model — the structural effects of mental health.

On the other hand, γ_5^h and γ_5^m are 'reduced form' parameters. These capture the overall effect of the mother's mental health on her labor supply and time investment decisions. These parameters are reduced form because they capture multiple effects. They capture the effect mental health has on the marginal utility of leisure and consumption (λs in Equation 1.6), as well as the 'indirect' effect through its effect on the wage offer and on the child's ability, and how these affect her time allocation decision.

1.4.3 Endogeneity and Identification

One important issue for estimation is the endogeneity of investments in the production of children's cognitive skills. So far I have avoided any discussion about endogeneity and identification. In this section, I discuss the source of the endogeneity — unobserved investments — and how I address this issue — using time invariant family types and exclusion restrictions.

The main source of endogeneity has to do with unobserved investments that affect child outcomes and which may be correlated with observed investments. I allow family income, the mother's time and the mother's mental health to influence the technology of cognitive skill formation. By doing so I ignored many other investments that have been shown to be important for children's development. For example, children differ in whether they attended preschool and in the quality of instruction they receive in school (preschool and compulsory). These investments are key for children's development and are ignored in the technology of skill production in this paper (assuming they are not picked up by family income). Moreover, these schooling investments are correlated with both family income and maternal time investments. For example, mothers spend less time with children that attend preschool. Similarly, I have ignored investments made by the father of the child (e.g. the father's time with the child). Again, it is possible that fathers compensate by spending more time with the child when the mother is absent or is suffering from a mental health condition.

Another endogeneity problem arises when estimating the effect of mental health on the mother's time allocation and on her productivity in the labor market. Here, I worry about reverse causation. For example, just as poor mental health can lead to lower labor market productivity, lower wages can lead to financial strain and higher mental health problems (Dohrenwend et al., 1992).

I control for the endogeneity of mental health in two ways. First, I model the correlation in unobserved shocks across equations with time invariant family types. Second, I use exclusion restrictions derived from the model to identify the causal effect of the mother's mental health, family income, and maternal time investments in children.

Mental Health Function

In order to control for the endogeneity of mental health, I need to specify how mental health is determined. I assume a reduced form specification for the mother's mental health. That is, I assume the mental health function is a log-linear function of the mother's observed characteristics and state variation in mental health parity laws, which are described in detail in Section 1.4.3.

Despite a large literature describing the production function of physical health, there is surprisingly very little work in economics discussing the production function of mental health. Psychologists describe that psychological distress, and mental health illnesses in general, develop from the inability of the individual to cope effectively with stressors and emotional turmoil (Horwitz, 2007; Ridner, 2004; Drapeau, Marchand, and Beaulieu-Prévost, 2011). As a result, mental health can be thought as a function of these different stressors as well as protective factors. Some stressors are economic in nature, and as such are considered to be endogenous, such as poverty and economic strain (Conger et al., 1994, 2002). Other are not, and are usually thought to be exogenous, such as the death of a relative (Persson and Rossin-Slater, 2014), or exposure to stressful events such as terrorist attacks (Camacho, 2008).

Protective factors can be thought as conditions that help the individual cope with the stressful event. For example, Evans and Garthwaite (2014) shows that government programs such as the EITC, which is thought to alleviate financial strain, can lead to reductions in maternal depression. Moreover, access to mental health services in the form of therapy and medication can alleviate and treat the symptoms related to mental disorders. In general, policies that improve the access to mental health services are expected to lead to improvements in mental health.

Following these ideas, I assume, psychological distress is a function of the mother's observable characteristics, such as her education and marriage status, as well as the state level variation in mental health parity laws. Observable characteristics capture the fact that certain groups are more likely to be exposed to stressful events than others. Similarly, it captures the idea that certain social groups have more resources to cope with stress than others (Drapeau, Marchand, and Beaulieu-Prévost, 2011). On the other hand, parity laws capture variation in access and coverage to mental health services across states. These services can be thought as helping the mother cope with the different stressors. Formally, the mental health function can be described by:

$$ln(H_{it}) = X_{it}^h \delta_x + \omega_{st} \delta_s + \eta_{it}^{mh} \tag{1.14}$$

where X_{it}^h are observable characteristics of the mother, ω_{st} is a dummy for whether state s has passed a mental health parity law by year t and η_{it}^{mh} is a shock to the mother's psychological distress.

Unobserved Types

In order to control for the endogeneity of investments, I allow for the unobserved shocks (η_{it}^{mh}, η_{it}^{a}, η_{it}^{hw}, η_{it}^{mt} and η_{it}^{w}) to be correlated across equations. I assume these unobserved shocks have two components: a time invariant component that is common to all shocks, and a time variant component that is assumed to be independently distributed over time and across equations.

These time invariant family types capture the idea that families differ in similar but unobservable ways. For example, it is possible that some families are more likely to send their children to preschool (unobservable in the model), and as a result these children develop at a fast pace even though we observe that they faced lower mother-child interactions. The time-invariant types are assumed to capture these important unobserved differences across families and as a result allow me to model the endogeneity in the empirical model.

Formally, I assume each unobserved shock (η_{it}) has two components. One that is time-invariant and common to all shocks (κ_i) and another that is independent and identically distributed over time and across equations (ϵ_{it}). I further assume that the time invariant component (κ_i) follows a discrete distribution with K types, so that we can write the unobservable shocks as:

$$\eta_{it}^J = \sum_{l=2}^{K} \rho_l^J \mathbb{1}[\kappa_i = l] + \epsilon_{it}^J \qquad \forall J \in \{a, hw, mt, w, mh\} \tag{1.15}$$

Moreover, I allow for the distribution of these different family types to differ across the population. I do so in order to account for differences in in-utero investments and genetic endowments across family types. That is, I allow for the probability of mother i to belong to family type k to be a function of her educational attainment at the time the child is born as well as for her mental health status before the child's birth. The hope is that educational attainment and early mental health conditions capture maternal skills and mental health endowments that are unobservable by the econometrician. Moreover, these endowments are correlated with in-utero investments and genetic endowments transmitted to the child, which are also unobserved.

Formally, the probability that individual i belongs to family group k is given by:

$$\pi_{ik} = \frac{exp(\theta_{0k} + \theta_{1k} S_i + \theta_{2k} D_i)}{1 + \sum_{l=2}^{K} exp(\theta_{0l} + \theta_{1l} S_i + \theta_{2l} D_i)} \quad \forall k \in 1, ..., K \tag{1.16}$$

where $\theta_{01} = 0$, $\theta_{11} = 0$ and $\theta_{21} = 0$, S_i correspond to educational attainment of woman's i and D_i is a dummy for whether she experiences depression before age 17.

In theory, the number of family types (K) can be as large as the number of individuals in the sample or as low as one. A piori, there is no theoretical reason to choose one number over another. The usual practice is to increase the number of types sequentially until the probability of a given type becomes "small enough".[23] For example, in my preferred empirical specification, I assume there are three family types since the estimated probability of belonging to the forth type was small (< 0.06) for most individuals when I allowed a forth type.[24]

Exclusion Restrictions

There are three endogenous variables in the model: the mother's mental health, monetary investments measured by family income, and maternal time investments. I use exclusion restrictions to identify their causal effects.

In order to identify the causal effect of the mother's mental health, I use variation in state mental health parity laws. In order to estimate this effect, I need a factor that affects the mother's mental health but does not enter anywhere else in the model. That is, something that does not directly influence her labor market productivity, her time allocation decisions, or child cognitive skills.

[23] This is arbitrary since it is up for each researcher to decide what number is considered "small enough".

[24] Moreover, when I allowed for a forth type, I did not observe any qualitatively changes in my results.

Finding such variation is not easy and the literature has struggled with this issue. Here, I use variation in mental health care access and coverage across states and over time. This variation comes from mental health parity laws passed by states in the 1990s. These laws are described by ω_{st} in Equation 1.14. I discuss these laws in more detail below.

In order to estimate the causal effect of family income (goods investments) for children's cognitive development, I use variation in both labor market conditions (ζ_{st}) and welfare rules (τ_{st}). In order to estimate this effect, I need a factor that affects family income but does not influence children's cognitive development directly. Both variation in labor market conditions and welfare rules serve this purpose. Labor market conditions determine the wage offer received by the mother (ζ_{st} in Equation 1.9), and as a result influence family income indirectly. Variation in welfare rules determine government benefits received by the mother (τ_{st} in Equation 1.7), and as a result influence family income directly. I use the same variation (labor market conditions and welfare rules) to identify the effect of maternal time investments for children's cognitive development. Both of these variables change the budget constraint faced by the mother, and as a result, influence the mother's time allocation decision (labor supply and time with the child). I describe these variables in more detailed below.

Mental Health Parity Laws. One long standing feature of the U.S. health system has been the unequal coverage by insurance plans of mental health care in comparison to general medical care. Until recently, with the passage of the Paul Wellstone and Pete Domenici Mental Health Parity and Addiction Equity

Act of 2008, [25] federal law provided few restrictions on this disparity.[26] In order to counter this lack of legislation, beginning in the 1970s and more aggressively in the 1990s, states passed a series of mandates requiring employers and insurers to regulate mental health benefits in their offered plans.

These laws varied significantly across states. Some states required insurance plans to provide mental health coverage in all offered plans. Moreover, they required that these benefits, including those for substance abuse, to be equal to the benefits for general physical conditions. This is the strongest type of mental health law that was approved. These laws are considered 'full parity' laws. Other states passed milder versions. Some only required insurance plans to offer mental health care coverage but left the purchase decision to the individual buyer. These laws are generally called 'mandate offering' laws. Other states passed weaker laws requiring parity in benefits only if a mental health plan was offered - 'mandate if offered' laws. Besides these distinctions, there were also significant variation across states on which mental health conditions were covered by the law and whether it excluded some important groups. For example, some laws did not apply to individual plans, while others excluded plans offered by companies with less than 50 employees. This variation in the 'quality' of these laws makes it tricky to separate states into parity and non-parity states.

[25] In 2008 Congress passed the Paul Wellstone and Pete Domenici Mental Health Parity and Addiction Equity Act, a law that prohibits financial requirements, treatment limitations and benefits for mental health and substance use disorders to be more restrictive than medical and surgical benefits.

[26] One exception, is a 1996 mandate established by the congress that prohibited discrimination with respect to annual benefit limits on employer plans that chose to offer mental health coverage. However, besides annual limits employers were free to discriminate or not offer any mental health benefits.

In order to address this issue, I use information provided by the National Alliance on Mental Illness (NAMI), which separated parity laws into two groups: 'comprehensive' and 'limited' laws. Limited laws excluded some important mental health condition or group from the parity restrictions. Following their definition, I assigned a state as having passed a parity law if they passed a 'comprehensive' full parity or mandate offering law. That is, I define that, at year t, state s has a parity law (ω_{st}) in place if by time t it had passed a 'full parity' or 'mandate offering' law that did not exclude individual or group plans and did not excluded important mental health conditions.[27]

I argue that these laws only enter the model through their effect on the mother's mental health status. However, one possible threat to identification would be if these laws also improved access to mental health services for children. In that case, these laws could improve child outcomes directly. I argue this is probably not the case in two ways. First, mental health coverage is less of an issue for children since they generally have higher rates of coverage from Medicaid and the Children's Health Insurance Program (CHIP) since 1997. As an evidence, pediatricians are less likely than other caregivers to report not providing outpatient mental health services because of lack of or inadequate coverage (Cunningham, 2009). Second, previous research provides evidence that state parity laws did not affect the likelihood of a child receiving outpatient mental health services (Barry and Busch, 2008) or receiving needed mental

[27]In order to construct these laws, I follow information collected by the National Alliance on Mental Illness (NAMI) (see: http://www.kantorlaw.net/documents/articles-and-information/2010-IAEDP/Mental-Illness-State-Mental-Health-Parity-Laws.pdf). Whenever needed, I supplemented this information with results in Lang (2013) and information provided by the National Conference of State Legislature NCLS (see http://www.ncsl.org/research/health/mental-health-benefits-state-mandates.aspx).

health care (Barry and Busch, 2007).[28] In contrast, previous research does show evidence that parity laws improved utilization of mental health care services in adults (Harris, Carpenter, and Bao, 2006).[29]

Labor Market Conditions. I use two variables to capture variation in labor market conditions. The median wage rate in the state for workers in the service sector and the share of the population in the state that works in the service sector. These variables were measured at the state and year level using data from the current population survey (CPS). These variables are commonly used in the literature as exogenous variation in the wage rate and are described in more detail in Table 1.4.

A possible threat to identification in using labor market conditions is that they are possibly related to the father's labor supply decision. The fact that they affect the father's wage is not a problem since I control for family income in the model. However, they might also influence the amount of time the father spends with the child, which is treated as an unobservable in the technology of skill production function.

Welfare Rules. I use the large variation in welfare rules across states and over time in the U.S. as exclusion restrictions in the model. Welfare rules have been shown to significantly affect the labor supply of single mothers (Moffitt, 1992). Moreover, these rules have been used in previous work to identify the effect of maternal work decision on child outcomes (Bernal and Keane, 2010,

[28] Although there is evidence that these laws reduced children's annual out-of-pocket health care spending exceeding $1,000 (Barry and Busch, 2007).

[29] Another threat to identification would be if these laws were correlated with other state level conditions. For example, these laws could be correlated with state level labor market conditions. I cannot rule out this possibility, however, I find that these laws are only weakly correlated (< 0.2) with other state level conditions I use in this paper, such as state level unemployment rate and welfare rules.

2011). I use state variation in waivers and requirements under the Temporary Aid to Needy Families (TANF) program after 1997 and state variation in benefits and income requirements under the Aid to Families with Dependent Children (AFDC) program before 1997. In addition, I supplement these with state and time variation in the shape of the earned income tax credit (EITC) schedule for a family of three. This variation is important as it also affects the labor supply of married women (Eissa and Hoynes, 2006). These variables are described in more detail in Table 1.4.

One issue with these welfare policies is that there are too many of them (18 variables in total), each having a small effect on women's labor supply decision. This is problematic for estimation as it creates unnecessary computational burden. Preferably, I would like to have a smaller set of variables with a stronger predictive power. In order to do that, I follow the approach proposed in Bernal and Keane (2011). I summarize the information contained in these 18 variables into two scores via factor analysis. These scores are estimated using the principal factor method and the varimax rotation. These scores have two important properties. First, these factors are linear functions of the original policy variables, and as a result, are also valid exclusion restrictions. Second, these scores have a much stronger predictive power than each policy variable separately.

These rules are commonly used as instruments for maternal investments in children. However, I should still mention possible threats to identification. One important threat is the fact that these laws changed significantly in 1997 with the introduction of the TANF program. However, also in 1997, the federal government introduced the State Children's Health Insurance Program (SCHIP). This program largely expanded health insurance coverage for children, and as a result is arguably correlated with child outcomes. I hope that by using variation in welfare rules from 1983 to 2013 and the variation in the EITC schedule this becomes less of a problem.

1.4.4 Measurement Error

Another issue that can lead to biased estimates is measurement error. Most concerning for this paper is the measurement error in the mental health construct. The Kessler 6 psychological distress scale used in this paper suffers from both the intrinsic measurement error in these self-reported questionnaires as well as measurement error from aggregating information from different measurements - the scale consists of six different questions. In order to control for this problem, I use an item response theory (IRT) approach.

The Kessler 6 psychological distress scale is composed by 6 questions scored on a scale of five values (0-4). The usual approach in the literature is to sum the answers to the 6 questions to end up with a score ranging from 0 to 24. There are, however, many issues with this simple approach. If we think that each question is measured with some noise and that the variance in the noise is different across questions, then summing up the scores on each question will provide a very unreliable and noisy measure of the underlying mental health.

Moreover, each question provides different information about the underlying psychological distress that it is measuring. There is no reason to believe that a score of 4 in one of the measures imply the same level of psychological distress as a score of 4 in another measure. For example, feeling nervous “all of the time” might indicate something different than feeling restless or fidgety “all of the time”. This is evident as the prevalence rates of scores are different across questions. Similarly, we have no reason to believe that different changes in scores within a measure provide the same information about the change in the underlying psychological distress. For example, answering 4 versus 3 might imply a greater increase in psychological distress than answering 2 versus 1 in one of the questions. Summing up the scores, again, ignore these issues.

A better approach, common in the psychological literature, is to use an item response theory (IRT) model to control for the measurement error in the measurements as well as this difference in information across questions. Many different IRT models have been proposed in the literature. Here, I use the grade response model proposed by Samejima (1969), which is appropriate for multidimensional ordinal items. Formally, let $M_{ij} = k$ correspond to the answer to question j by individual i, which can take 5 different values $k = \{0, 1, 2, 3, 4\}$. The IRT model is interested in estimating the probability of observing answer k or higher for question j and individual i given the underlying psychological distress level θ_i. This probability is assumed to be given by:

$$Pr(M_{ij} \geqslant k|\theta_i) = \frac{exp(a_j ln(\theta_i) + b_{jk})}{1 + exp(a_j ln(\theta_i) + b_{jk})} \tag{1.17}$$

where a_j captures the information value of question j and b_{jk} is the kth cutpoint for question j and is usually understood as the difficulty in answering k or higher in item j. Alternatively, the probability of observing outcome k is given by:

$$Pr(M_{ij} = k|\theta_i) = Pr(M_{ij} \geqslant k|\theta_i) - Pr(M_{ij} \geqslant k + 1|\theta_i) \tag{1.18}$$

where $Pr(M_{ij} \geqslant 0|\theta_i) = 1$ and $Pr(M_{ij} \geqslant 5|\theta_i) = 0$.

I compute these probabilities outside the main model estimation. This part of the model is computed by simulated maximum likelihood. Let k_{ij} be the observed answer to question j by individual i, then the likelihood for individual i is given by:

$$L_i = \int_{-\infty}^{\infty} \prod_{j=1}^{6} Pr(M_{ij} = k_{ij}|\theta_i, a_j, b_{jk}) f(\theta_i) d\theta_i \tag{1.19}$$

where $ln(\theta_i)$ is assumed to be normal distributed with mean 0 and variance 1 and the model is estimated by simulated maximum likelihood.

The value of the unobserved psychological distress for each individual is estimated in a second step by the empirical Bayes method. The value is estimated by the empirical mean and is determined by:

$$ln(\hat{\theta_i}) = \int_{-\infty}^{\infty} ln(\theta_i) \frac{\prod_{j=1}^{6} Pr(M_{ij} = k_{ij}|\theta_i, \hat{a_j}, \hat{b_{jk}}) f(\theta_i)}{\int_{-\infty}^{\infty} \prod_{j=1}^{6} Pr(M_{ij} = k_{ij}|\theta_i, \hat{a_j}, \hat{b_{jk}}) f(\theta_i) d\theta_i} d\theta_i \tag{1.20}$$

where $\hat{a_j}$ and $\hat{b_{jk}}$ are the estimated parameters in the first step and $f(\theta_i)$ is the prior distribution of theta. The estimated $\hat{\theta_i}$ is the main measure for the individual psychological distress scale.

1.4.5 Estimation: Method of Simulated Moments

I estimate the parameters of the model using the method of simulated moments (MSM). The estimation method follows an iterative process. First, I calculate the moments from the data. Then, given an initial guess of the parameter vector, I simulate 10 paths for each woman and her child. That is, I first simulate the path for the mother's psychological distress for the 17 periods (ages 0-16). Then, I simulate the hours of work and the time investment decisions at each period using the structure described in Section 1.4. Following that, I simulate the wage offer received by the mother at each period following the structure described in Section 1.3.2, and path for the child's cognitive ability described in Section 1.3.1.

Once I have these, I can calculate the moments from the simulated data and the weighted distance between the sample moments and the simulated moments from the data. The iterative process continues until this distance is minimized.

More formally, let Ω denote the parameter vector, $M_S(\Omega)$ denote the vector of moments from the simulated data and M_O the moments from the observed data. Then, the estimated parameter vector $\hat{\Omega}$ solves the following objective function:

$$\hat{\Omega} = \arg\min_{\Omega}(M_O - M_S(\Omega))'W(M_O - M_S(\Omega)) \tag{1.21}$$

where W is a symmetric, positive-definite weighting matrix. I construct W to be the inverse of the covariance matrix of M_O estimated by bootstrap with 500 replications. That is, I compute the vector of moments M_O^q for each of the Q resamples from the original N data points, which leads to the following covariance matrix for M_O:

$$W = \left(Q^{-1}\left(M_O^q - Q^{-1}\sum_q M_O^q\right)'\left(M_O^q - Q^{-1}\sum_q M_O^q\right)\right)^{-1} \tag{1.22}$$

The moments that form M_O and consequently M_S include the mean and standard deviation of the child's cognition for each of the child's age. They also include the mean of the child's cognition by different maternal characteristics such as maternal education. I also include the mean of the child's cognition by different percentile levels of maternal investments five years prior to the estimated cognition. Moreover, I include the mean and standard deviation of maternal work hours, maternal time with the child, observed wages and the mother's psychological distress, as well as the mean of each variable by the percentiles of maternal and the child observed characteristics. I also include the correlations between the observed wage rate, hours worked and the maternal time with the child, correlations between the contemporaneous wage rate and lagged work hours, and between the two contemporaneous time allocation choices and the two lagged time allocation choices.

1.5 Empirical Results

I start by describing the main estimation parameters and how these compare with estimates from a static model and a model that does not control for the endogeneity of mental health. These comparisons highlight the importance of controlling for endogeneity and allowing for dynamic effects. Next, I describe the overall effect of maternal psychological distress on children's cognitive development, and the relative importance of each proposed mechanism in explaining this effect. I will argue that maternal mental matters since a 1% increase in maternal distress in all periods results in a 0.17% decrease in children's cognitive scores at age 16. I will also show that the effect of maternal distress on the productivity of maternal investments explains 70% of the effect of maternal mental health on children.

1.5.1 Human Capital Production Function

Table 1.8 presents the estimated parameters for the main outcome of interest, the child human capital production function described in Equation 1.4. The estimates show some interesting patterns. The first evident pattern is that the total factor productivity (K) is about 50 percent higher in the second developmental period than in the first. This suggests that inputs in the production function explain a much larger share of cognitive development in the first period than in the second. This finding is similar to previous research that highlight the higher return of investments early in life (see Heckman and Mosso (2014)). The estimated parameters also show that the relative self-productivity of children's cognitive skills (captured by α_1) is much greater in the second developmental stage than in the first stage. The high self-productivity parameter in the second stage also highlights the importance of investing in children early in the life cycle.

Given these results is perhaps not surprising that I find that the relative productivity of family income (α_2) and maternal time investments (α_3) are significantly higher in the first developmental period than in the second. However, I also find that the productivity (or penalty) for the mother psychological distress (α_4) is similar across both periods. This finding could be explained by the idea that the contagion of mental health does not depend on the child's age. It also underlines possible benefits of mental health interventions at later stages in the child's development.

Parameters α_5, α_6 and α_7 in Table 1.8 are not significantly different than zero. This implies that my model rejects evidence of dynamic complementarities beyond the what is already implied by the Cobb-Douglas. This result is not that different from other papers in the literature. For instance, Cunha, Heckman, and Schennach (2010) find similar evidence under some specifications.[30] I also do not find evidence of static complementarity between family income and the maternal time investments beyond the one implied by the Cobb-Douglas function (α_8 in Table 1.8). Moreover, I find economic large but statistically insignificant

[30]Cunha, Heckman, and Schennach (2010) cannot reject the Cobb-Douglas formulation when they estimate the production function using only cognitive skills.

static complementarity between maternal mental health and family income (α_9 in Table 1.8). Also, this complementarity has an opposite sign in the two developmental stages.

In contrast, I find a large and significant static complementarity between maternal mental health and maternal time investments in both developmental stages. This is one of the key findings in the paper. It suggests that the returns to maternal time investments are highly dependent on the mother's mental health. It suggests that the value of maternal time investments are very high when the mother is in a good mental health state. Moreover, it suggests that the value of maternal time investments can be negative for children of mothers in poor mental health.

1.5.2 Time Allocation

Tables 1.10 and 1.11 present the estimated parameters for the two time allocation decisions as described by Equation 1.12. As I discussed in Section 1.2.2, the direction of the effect of maternal psychological distress on the time allocation decisions is uncertain since the estimated linear effect captures many different channels. For instance, it captures the effect of maternal distress on the marginal utility of leisure, which is expected to be positive. It also captures the effect of maternal distress on maternal wages, which in turn changes the budget constraint. Moreover, it captures the effect of maternal distress on the child's human capital, which is part of the woman's utility function. One would expect an increase in the utility of leisure to decrease the time spent in either the labor market or interacting with the child. However, a decrease in her productivity in the labor market could have an ambiguous effect due to substitution and income effects. The same thing is true for a decrease in the productivity of maternal time investments.

My findings point to small and economically insignificant effects for the work decision and a negative but small effects for the maternal time investments decision. For instance, I find that a one standard deviation increase in maternal distress increases labor force participation by only 16 hours per year (see Table 1.10). As a comparison, in the preliminary analysis, I ignored both endogeneity and dynamic issues and found an effect on labor force participation that was four times larger and of a different sign (column [4] in Table 1.3). This change in sign is also present in a model that allows for dynamic interactions but ignores the endogeneity of mental health. Parameter estimates for this model can be seen in Table A.4 in A.1.[31] This results is different than other papers in the literature that have estimated a negative effect of mental health on labor force participation (Frijters, Johnston, and Shields, 2014; Ettner, Frank, and Kessler, 1997). Moreover, they highlight the importance of modeling the dynamics of the mother's labor force participation.

Similarly, I find that a one standard deviation increase in maternal distress decreases maternal time with the child by only 0.06 hours per week (see Table 1.11). As a comparison, this estimated effect is around ten times smaller than in models that do not control for the endogneity of mental health (see column [2] in Table 1.3 and results in Table A.5).

The remainder results in Tables 1.10 and 1.11 are unsurprising. I find that highly educated mothers spend both more time in the labor market and more

[31] For the results in A.1, I re-estimate my main empirical model under the assumptions that the shocks in each equation are uncorrelated across equations. In other words, I assume that there are no unobserved family types

time with their children. This is a well known result from the child development literature (Kalil, Ryan, and Corey, 2012). I also find that mothers spend more time with girls than boys and that higher non-maternal labor income is related to a lower time spent with the child. Finally, I also find that the two welfare rules factors are significantly predictive of the two time allocation decisions.

1.5.3 Wage Offer

Table 1.12 presents the estimated parameters for the hourly wage equation described in Equation 1.9. According to the intuition described in Section 1.2.2, one would expect maternal psychological distress to be negative related to the productivity in the labor market. However, I do not find that to be the case once I control for the endogeneity of maternal distress. A one percent increase in psychological distress causes a 0.002 percent decrease in maternal hourly wages, which is economically insignificant.

Before controlling for the endogeneity of mental health the estimated relationship between distress and wages was between ten to twenty times larger. These can be seen in column [6] in Table 1.3 and in Table A.6). Moreover, the estimated results are much smaller than the reported numbers in the rest of the literature (see (Ettner, Frank, and Kessler, 1997) as an example). These results highlight the importance of properly accounting for the selection into employment and for the endogeneity of mental health.

The other parameters in the wage equation follow standard economic theory. I find that maternal years of education is positively related to labor market productivity, that the offered wage increases as the woman ages, and that a stronger labor market, as measured by median service sector wages, is positive related to hourly wages. Moreover, I find a positive relationship between labor market experience and wages and a strong penalty for spending a period outside the labor market.

1.5.4 Decomposition

When taking in account all of the different mechanisms I find that maternal mental health matters for children's cognitive development. This is one of the key findings in this paper. I find that on average a 30% decrease in mothers' psychological distress result in a 5.09% increase in children's cognitive skills at age 16. This effect is large and similar to, for example, the effect of a $350 per week increase in family income.[32] This effect is reported in the first row of Table 1.5. This effect can be seen graphically in Figure 1.5. The solid line plots the simulated change in cognitive scores at age 16 for the median child for different levels of maternal distress.

Two mechanisms are key in explaining the effect of maternal distress on children's development. The most important mechanism is the effect of maternal mental health on the productivity of maternal time and goods investments. I call these 'complementarity effects'. As can be seen in Table 1.5, this mechanism alone explains about 70% of the overall effect. That is, this mechanism alone implies that a 30% decrease in mothers' psychological distress would result in

[32] These results on family income are comparable to other research. For example, using the same data Del Boca, Flinn, and Wiswall (2014) finds that a $250 weekly increase in child goods lead to 4.6% increase in child quality at age 16.

a 3.72% increase in children's cognitive skill at age 16.[33] [34] The importance of this mechanism can also be seen graphically in Figure 1.5. Once I control for these complementarities changes in maternal distress have a significant smaller effect on the child's cognition.

The importance of this mechanism is explained by two parameter estimates. First, it is explained by the large negative complementarity between maternal time investments and maternal distress in the technology of cognitive skill formation (α_{10} in Table 1.8). It is also explained by the finding that maternal distress does not influence the quantity of maternal time investments (see Table 1.11). These results together, imply that mothers in poor mental health spend the same amount of time engaged with their children, when compared to mothers in good mental health, even though their time investment is significantly less productive (and sometimes harmful) for their children. The fact that some mothers spend time with their children even when it is not productive (or harmful) to do so highlights the benefits of policy interventions.

The second important mechanism is the direct effect of maternal mental health on children's cognitive development (α_4 in Table 1.8). This mechanism captures all the ways mental health affects children that are not captured by the other mechanisms in my model. For example, it captures mental health

[33]Here is a brief description on how I compute the separate effect of each mechanism. First, I simulate children's cognition scores at age 16 without any changes into the model. Then, I create a new measure for the mother's distress that is 30% smaller. Next, I substitute this new measure for the old one in the maternal labor force participation equation, and compare the new simulated children's cognition scores at age 16 with the old score. This allows me to compute the percentage change in children's scores due to the effect of maternal distress on the mother's labor force participation alone. Then, I allow the new measure of maternal distress to enter the model through the other mechanisms one at a time. This allows me to compute the change in children's scores due to each of the other mechanisms.

[34]The order of the mechanisms can change the estimated contribution of each mechanism. However, the results are very similar independent of the order used.

contagion, the idea that the child develops mental health problems of their own by being exposed to the mother's mental health problems. The direct effect explains about 30% of the overall effect, as I describe in Table 1.5. This mechanism alone implies that a 30% decrease in mothers' psychological distress would result in a 1.38% increase in children's cognitive skill at age 16. I also depicted this mechanism graphically in Figure 1.5. The relative importance of the direct effect is represented by the second dotted line. It is also important to notice that, differently than family income (α_2) and time investments (α_3), the effect of maternal distress (α_4) is high in both developmental stages. This is also true for the complementarity effect (α_{10}). This means that policies that focus on improving maternal mental health are especially important when targeting older children.

I do not find strong evidence that the other mechanisms play an important role for children's cognitive development. These other mechanisms correspond to the proposed effect of the mother's mental health on her labor force participation, on the time spent with her child, and on her labor market productivity. This is perhaps surprisingly given that I do find evidence of these mechanisms when looking at static effects, as I demonstrated in Section 1.2.4. However, evidence of these effects disappear once I allow for dynamics in the mothers' labor force participation decision, and, more importantly, when I control for the endogeneity of mental health, as I described in the beginning of this section. These results highlight the importance of constructing a model that allows for dynamic interactions and endogeneity.

1.6 Policy Analysis

The results discussed in the previous section suggest two avenues for policy intervention. Two mechanisms explain the effect of maternal psychological distress on children's development: the direct effect of maternal distress and the negative complementarity between maternal distress and maternal time investments. Policies that aim to improve the cognitive development of children of at-risk mothers should target these two mechanisms. One would be to treat at-risk mothers for their mental illness. Another policy would be to improve maternal parenting, as in home visitation programs.

1.6.1 Mental Health Treatment

The most obvious policy intervention would be to screen mothers for psychological problems and treat all mothers at risk of developing a mental illness. This approach is currently recommended by the American Academy of Pediatrics regarding maternal postpartum depression. The Academy recommends that pediatricians screen mothers for postpartum depression at the infant's 1, 2, and 4 month visits (Earls et al., 2010). Similarly, we could screen and treat mothers for mental health problems at later stages in the child's life.

There are many different approaches to treat mental health problems such as distress, depression and anxiety. Two are the most common: anti-depression medication and psychological interventions, such as cognitive behavioral therapy. Both methods have been shown to be important avenues to treat depression and anxiety disorders in adults. For instance, cognitive behavioral therapy has been shown to increase the probability of remission from depression in adults when compared to usual care, and to decrease the levels of depression by up to 1 standard deviation (Churchill et al., 2013). Psychological therapies have also been shown to effectively reduced generalised anxiety disorder in adults (Hunot et al., 2007). Moreover, both antidepressants and psychological interventions have been shown to reduce symptoms and levels of postnatal depression. For instance, mothers treated with antidepressants were between 43% to 79% more likely than those treated with a placebo to show signs of remission from depression (Molyneaux et al., 2014).[35]

Given the similarities between psychological distress and both depression and anxiety disorders, these results suggest that medication and therapy can significantly reduce maternal distress. Moreover, given these results it does not seem unreasonable to assume that mental health treatment could lead to a 30% decrease in the level of psychological distress. I should mention that coming up with a value for the treatment is inevitably arbitrary given that the efficacy of treatment vary with the type of treatment and also across individuals in the population.

Table 1.6 shows the effect of different policies on children's cognitive scores at age 16. On average, mental health treatment, represented by a 30% decrease in mothers' psychological distress, results in a 5.09% increase in children's cognitive skills at age 16. This effect was computed by re-simulating the model and calculating children's cognitive skills under the 30% reduction in maternal levels of psychological distress. By comparison, an increase in family income by the median TANF benefit, which was $379 per month in 2,000, only increases

[35] Most of these studies focus on Sertraline, which is an antidepressant of the selective serotonin re-uptake inhibitor class.

children's scores by 1.2%.[36] Unsurprisingly, the returns to mental health treatment are larger for children of mothers in poor mental health. This can be seen in Figure 1.6 that plots average percentage change in children's cognitive scores at age 16 as a result of mental health treatment for different percentiles of maternal psychological distress.[37]

We can use a simple back-of-the-envelope calculation to compare the value of mental health treatments to equivalent income transfers. Russell et al. (1999) estimated the average cost of mental health treatment for adults, including outpatient visits and hospitalizations, to be around $1,200 per year. Assuming the 30% decrease in maternal psychological distress, maternal mental health treatment improves children's cognitive scores by 5.09 percentage points, on average. In comparison, a $1,200 permanent annual increase in family income improves children's cognitive socres by mere 0.32 percentage points. These results suggest that, on average, investments in mother's mental health are 16 times more valuable for improving children's cognitive skills than comparable income transfers.

[36]One important caveat from computing the effect of income transfers is that the model does not allow increases in family income to affect the mother's mental health. Higher family income should lead to better mental health as it alleviates financial strain and increases the resources to be invested in the mother's mental health.

[37]In Figure 1.6, the effect of cash transfers are larger for children of mothers in poor mental health. This is surprisingly, since dynamic complementarity implies that the return of family income would be larger for mothers in good mental health. However, I estimated a positive complementarity between family income and maternal distress in the second developmental period (α_9 in Table 1.8), which explains why I find the opposite result. As a matter of fact, if I look at younger children, where the complementarity parameter α_9 is negative, the effect of cash transfer are much larger for children of mothers in good health.

1.6.2 Improving Maternal Parenting

Another avenue for policy intervention would be to invest in programs that improve mother-child relationships for mothers with high levels of psychological distress, as in for example home visitation programs. In traditional home visitation programs, a nurse or social worker provides educational training to mothers during frequent visits to the family's home. The training focuses on many areas, including parenting skills, maternal health, and infant nutrition. These programs have become increasingly popular and include well know programs, such as the Nurse-Family Partnership (NFP), Healthy Families America, and the Infant Health and Development Program (Howard and Brooks-Gunn, 2009; Ammerman et al., 2010). These programs have been evaluated by control trials, and shown to improve children's achievement in school, decrease children's behavioral problems and reduce criminality during adolescence (Howard and Brooks-Gunn, 2009).

One of the many ways through which these programs improve child outcomes is by improving maternal parenting skills. As a result, one way to model home visitation programs would be to assume they increase the productivity of maternal time investments $\left(\frac{\partial A_{t+1}}{\partial MT_t}\right)$ for mothers in poor mental health. One big issue is deciding on how much these programs improve the productivity of maternal time investments. Since we don't have actual estimates any number will be arbitrary. I argue that a reasonable assumption is to assume these programs lead to a 20% reduction in the gap in productivity of maternal time investments across mothers.[38]

[38] As mentioned, this number comes with many caveats. First, the actual improvement in the productivity of maternal time investments can be very different across programs. Second,

Another issue with estimating the value of improving the productivity of maternal time investments is that in doing so I am changing structural parameters in the model. Since I did not estimate the mother's preference parameters (Equation 1.6), I cannot estimate how changes in the productivity of maternal time investments affect the mother's time allocation decisions. As a result, the results in this section rely on the assumption that the reduced form time allocation policy functions remain the same after the introduction of the program.

With that in mind, I find that, on average, a 20% reduction in the gap in productivity of maternal time investments results in a 6.31% increase in children's cognitive skills. This can be seen in Table 1.6. This effect was computed by re-simulating the model and calculating children's cognitive skills under the 20% reduction in the gap in the value maternal time investments. Again, these can be compared to an increase in family income by the median TANF benefit, which only increases children's scores by 1.2%. Moreover, I find large heterogeneity in the returns of this program across mothers, as can be seen in Figure 1.6. As expected, the benefit of this policy is negligible for children of mothers with low levels of psychological distress. However, this policy improves skills by as much as 11 percentage points for children of mothers with serious psychological distress.[39]

We can use a simple back-of-the-envelope calculation to compare the value

---

the improvement is most likely heterogeneous across mothers. Lastly, as far as I am aware nobody has measure the improvements in the productivity of maternal time investments from home visitation programs.

[39]There is a technical explanation for the difference in heterogeneous effects across the two programs. Mental health treatment decreases the level of maternal distress, which enters the child's technology of skill production in 'log-form'. This is not true for the improvement in parenting policy, which directly alters the α parameters in that function. As a result, a 30% in effect of maternal distress will have a much larger effect for children than a 30% decrease in the level of maternal distress.

of home visitation programs, insofar as they improve maternal parenting, to equivalent income transfers. Burwick et al. (2014) estimated the average cost of home visitation programs to be \$6,583 per family for a 44 week enrollment period. One important questions is whether these programs improve maternal parenting skills in the short or long run. That is, do we need to re-enroll families every year or enrolling them once is enough to improve maternal parenting until the child reaches the end of the developmental period. In the former case we should compare these programs to a \$6,583 permanent annual income increase, which improves children's cognitive scores by 1.74 percentage points. In the latter case we should compare these programs to a one-time \$6,583 increase in annual income, which at age 5 improves children's cognitive scores at age 16 by mere 0.09 percentage points, on average. With all the caveats already mentioned and assuming home visitation programs lead to a 20% reduction in the gap in productivity of maternal time investments, these results suggest that investments in home visitation programs are between 3-70 times more valuable for improving children's cognitive skills than comparable income transfers.

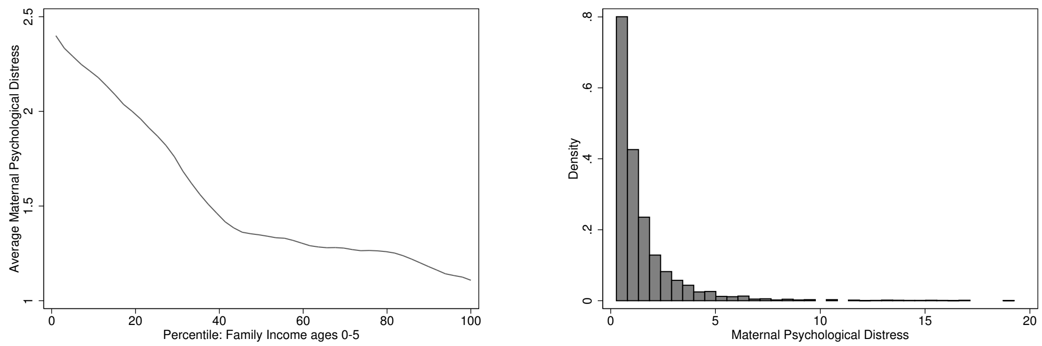
1.7 Conclusion

This paper builds on many strands of the child development literature in order to evaluate the different mechanisms that relate maternal mental health to children's cognitive development. In the analysis, I allow maternal mental health to determine the quantity as well as the quality of maternal investments. My model estimates, which were derived using data from the Panel of Study of Income Dynamics, shed light on the importance of mother's mental health for children's development.

Using the parameter estimates from the model, I first show that on average a 1% decrease in maternal psychological distress increases children's cognitive scores by 0.17 percentage points. Moreover, I show that the effect of maternal distress on the productivity (quality) of maternal time investments explains 70% of the effect on children. Next, I investigate policy interventions that can mitigate these effects. I argue that treating mothers for psychological distress could improve children's cognitive scores by 5.09 percentage points, similar to a $350 weekly increase in family income. Similarly, I argue that home visitation programs, which improve maternal parenting behavior, also have large benefits for children and are a viable option when treatment does not work. Moreover, back-of-the envelope calculations show that both policies are much more cost effective at improving child outcomes than income transfers.

These results open up many avenues for future research. In particular, future research should explore the different ways that government programs interact with maternal mental health in producing child outcomes. For example, we now know that increases in the earned income tax credit (EITC) are related to improvements in mothers' mental health (Evans and Garthwaite, 2014). We also know that increases in the EITC are related to improvements in children's outcomes (Dahl and Lochner, 2012). As a result, we can ask whether maternal mental health mediates the effect of EITC on children. Moreover, we now know that maternal mental health changes the productivity of maternal time investments, and as a result, we can ask whether differences in maternal mental health can partially explain the puzzling heterogeneity in the returns to childcare programs (van Huizen and Plantenga, 2015).

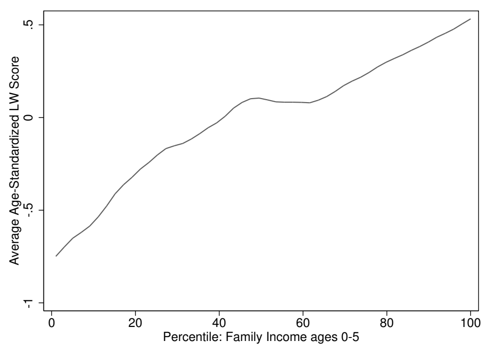
1.8 Figures



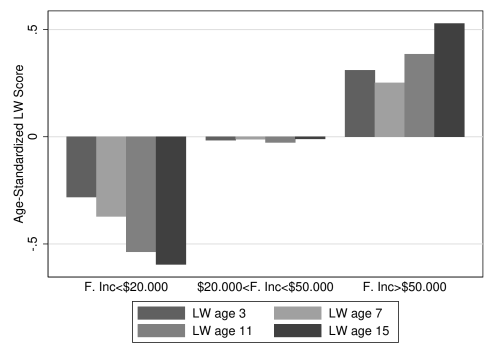
(a) Psych. Distress Scale by Family Income (b) Psychological Distress Scale - Density

Figure 1.1: PSYCHOLOGICAL DISTRESS SCALE

Figure 1.1(a) plots the average distress scale by family income percentiles measured between the ages 0 to 5. Individuals in the lower end of the income distribution are at a much higher risk of developing mental health problems than individuals at the higher end of the distribution. There a sharp drop until the median family income, which is equivalent to \$46,000 (measured in 2000 dollars). Figure 1.1(b) plots the density of the distress scale constructed using Item Response Theory. The responses are concentrated in the left part of the distribution with a long tail.



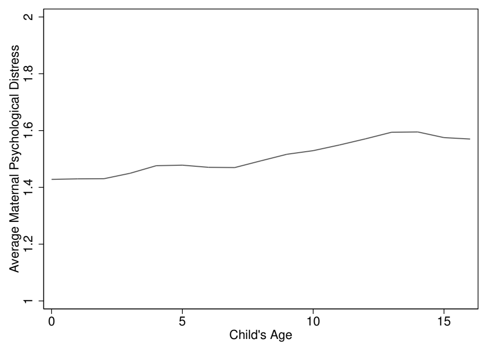
(a) LW Score by Family Income



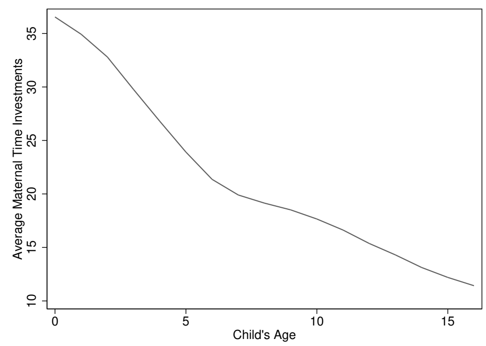
(b) LW Score by Family Income and Age Groups

Figure 1.2: Chidren's Letter-Word Score

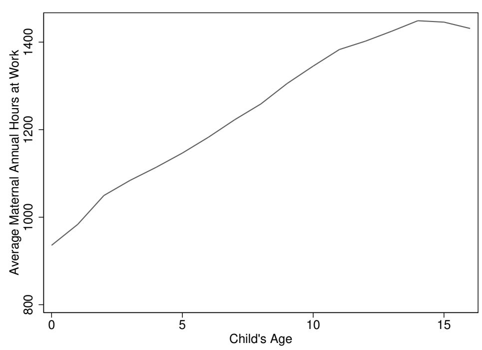
Figure 1.2(a) plots average age-standardized Letter-Word scores by family income percentile. LW score was measured from ages 3 to 16 and family income was measured between the ages 0 to 5. Socioeconomic disparities in children's cognition are large, the cognition gap can be as large as one standard deviation. Figure 1.2(b) does a similar analysis for different age groups. These socioeconomic disparities are present as early as by age 3 and tend to grow over time.



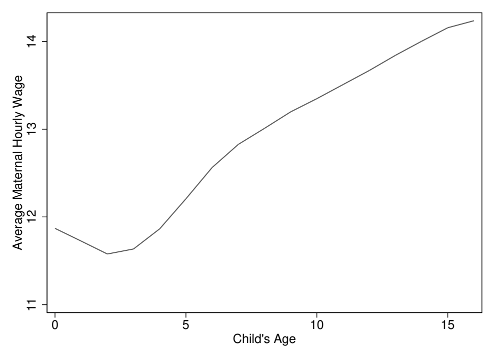
(a) Maternal Distress by the Child's Age



(b) Maternal Time Investments by the Child's Age



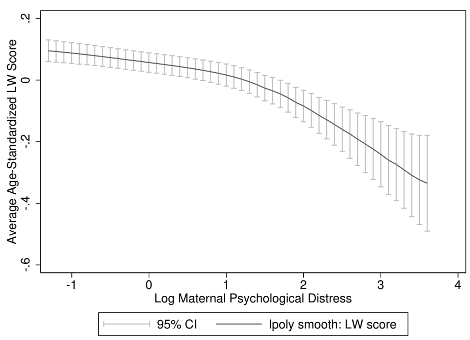
(c) Hours Worked by the Child's Age



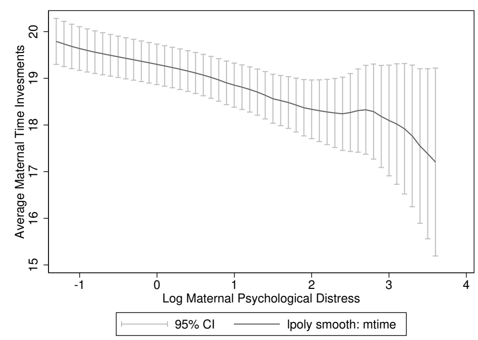
(d) Hourly Wage by the Child's Age

Figure 1.3: Investments by the Child's Age

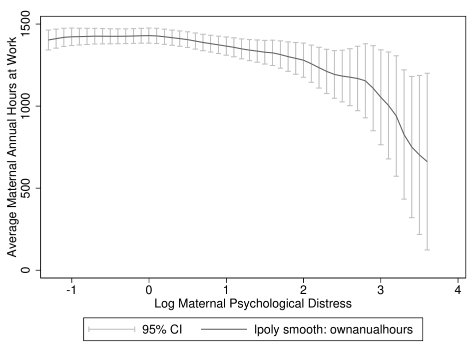
These figures plot changes in maternal investments as the child ages. Figure 1.3(a) plots changes in maternal distress, figure 1.3(b) plots changes in maternal time investments, figure 1.3(c) plots changes in the mother's labor force participation and figure 1.3(d) plots changes in the mother's wage rate.



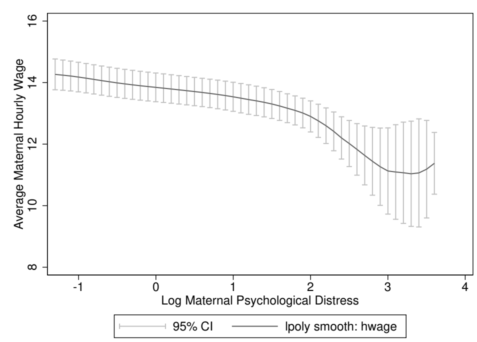
(a) Letter Word Score



(b) Weekly Maternal Time



(c) Annual Hours Worked



(d) Hourly Wage

Figure 1.4: Outcomes by Maternal Distress

Figures 1.4(a)-1.4(d) plot the raw relationship between the maternal distress and children's cognitive skills, maternal time investments, maternal labor force participation, and hourly wage received if employed. 95% confidence intervals are also plotted.

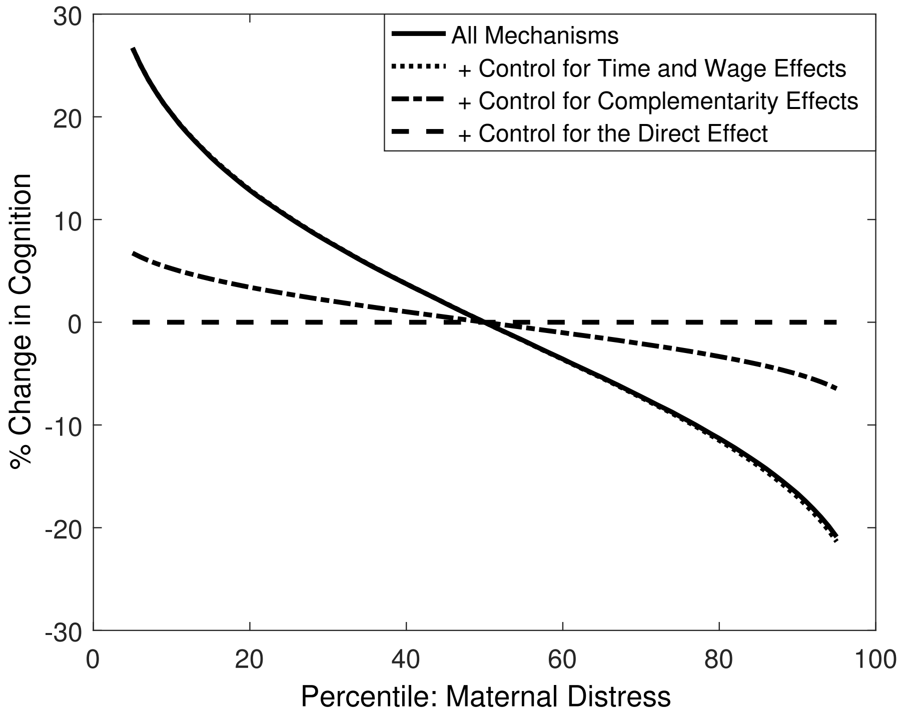


Figure 1.5: Decomposition of the Effect of Maternal Psychological Distress

This figure plots the simulated change in child cognitive score at age 16 for the median child for different levels of maternal distress. I re-simulate child cognitive skills under different specifications. That is, I control for different channels through which maternal mental health can affect child outcomes. First, I control for the effect of maternal distress on her labor force participation, time investments and wages. These are very small, so the second curve overlaps with the first. Then, I control for the effect of maternal mental health on the productivity of maternal time investments. Controlling for this mechanism significantly reduces the effect of maternal distress, highlighting its importance. Lastly, I control for the remaining effects. Once I do so, the line becomes flat, as expected.

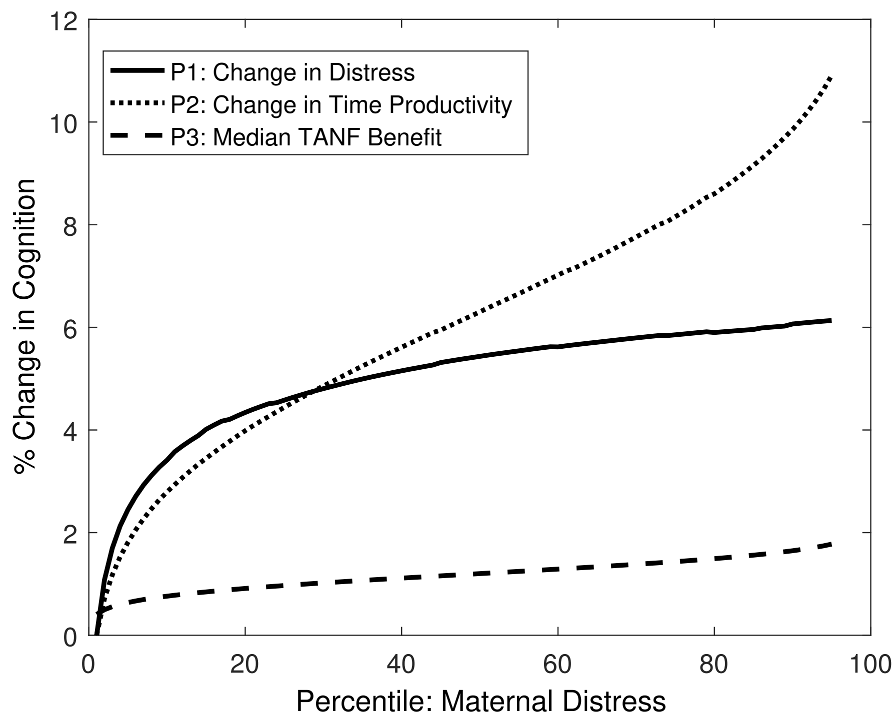


Figure 1.6: Policies

This figure plots the effect of different policies on children's cognitive development at age 16. These effects are calculated for the median child for different levels of maternal distress. In the first policy, I decrease psychological distress by 30% in the whole population. The second policy, decreases the gap in the returns of maternal time investments across individuals by 20%. Then, I compare these policies to an increase in family income by the median TANF benefit, which was was $379 per month in 2,000.

1.9 Tables

Table 1.1: Summary Statistics

Variable	Mean	SD	Min	Max	# obs
Child is Female	0.495	0.500	0	1	41803
Child is Black	0.401	0.490	0	1	41803
Mother's Years of Edu.	12.98	1.978	10	17	41803
Mother's Age at Birth	25.68	5.727	13	41	41803
Number of Children	2.791	1.189	1	11	41803
Mother is Single	0.317	0.2165	0	1	41803
Letter-Word Score	1.417	1.093	0.001	12.79	4582
Maternal Distress	1.533	2.066	0.271	36.64	4534
Maternal Time Investments	19.36	14.13	0	77.33	5006
Annual Hours Worked	1209	913.6	0	3640	25795
Hourly Wage	12.76	84.08	1.970	65.80	19864
Mother's Depression B.17	0.043	0.202	0	1	41803
Child-Age Observations					41803

Notes: Summary statistics for the analytic sample of 2,459 children. Children and their mothers were observed over 17 years for a total of 41803 child-age observations. Entries for the child's race and gender and the mother's cohabiting status and depression before age 17 are in the form of percentages divided by 100. Maternal time investments is measured in weekly hours, hours worked in annual hours and hourly wages are in 2000 dollars. The child's letter-word score has been log-age-standardized to have mean 0 and standard deviation equal to 1 at all ages. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 1.2: Preliminary: Log Letter-Word Score

Variable	[1]	[2]	[3]	[4]
Log Maternal Distress	-.129***	-.069**	-.068**	-.063**
Log Maternal Time	.	.	.035	.038
Log Family Income	.	.	.	.093**
Controls	(N)	(Y)	(Y)	(Y)

Notes: This table contains parameter estimates from OLS regressions used link maternal investments to child cognitive scores. I regress log age-standardized letter word scores at time $t + 5$ on maternal investments at time t. Controls include the mother's years of education, age at the child's birth, cohabiting status, number of children and the child's race and gender. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 1.3: Preliminary: Maternal Investments

	Maternal Time		Annual Hours Worked		Hourly Wages	
	[1]	[2]	[3]	[4]	[4]	[6]
Log M.Dis.	-.825***	-.547**	-57.662***	-41.810***	-.085***	-.048***
Controls	(N)	(Y)	(N)	(Y)	(N)	(Y)

Notes: This table contains parameter estimates from OLS regressions used link log maternal psychological distress to other maternal investments. I regress weekly maternal time with the child, annual hours worked and hourly wages if employed on log maternal distress. Controls for the time decisions include the mother's years of education, age at the child's birth, cohabiting status, number of children and the child's race and gender. Controls for the wage offer include the mother's age and age squared, her education and state level labor market conditions indicators.

Table 1.4: State Level Variables: Description and Sources

Instruments	Description	Source
	Labor Market Conditions	
$servwage_{st}$	Median hourly wage rate for workers in the Service Sector in State s	MORG
$empservice_{st}$	Share of the employed population working in the Service sector in State s	MORG
	State Variation in TANF Rules	
$DiversionProgram_{st}$	Dummy for whether State s has a Diversion Program in place at Period t	UI-WRD
$JobSearchRequired_{st}$	Dummy for whether State s requires applicants to search for a job before application	UI-WRD
$TPElegible_{st}$	Dummy for whether Two-Parent Families are eligible for benefits in State s	UI-WRD
$TPLimitonHours_{st}$	Dummy for whether State s has a limit on the number of hours a month the principal wage earner can work	UI-WRD
$TPWorkHistory_{st}$	Dummy for whether State s performs a work history test in order to determine eligibility of TP Families	UI-WRD
$TPWaitingPeriod_{st}$	Dummy for whether State s implements waiting periods for TP Families	UI-WRD
$NetIncomeTest_{st}$	Dummy for whether State s performs a Net Income Test before determining eligibility	UI-WRD
$EIDpercent_{st}$	Percent amount of Income disregarded in determining net income for the income eligibility tests n State s	UI-WRD
$EIDflat_{st}$	Flat amount of Income disregarded in determining net income for the income eligibility tests in State s	UI-WRD
$MaxIncEligF3_{st}$	Maximum monthly income for initial eligibility for a family of three in State s	UI-WRD
$MaxBenF3_{st}$	Maximum monthly benefit awarded for a family of three with no income in State s	UI-WRD
$WRHowLong_{st}$	Number of assistance months after which work is required in place in State s "	UI-WRD
$WRYoungCExemption_{st}$	Dummy for whether Stats s exempts the work requirement for a parent caring for a young child	UI-WRD
$WRChildExempt_{st}$	How old (in months) a child can be for the caregiver to be exempt from the work requirement in State s	UI-WRD
$WRNumWExemptions_{st}$	Number of different work exceptions allowed by State s	UI-WRD
$LifetimeTimeLimit_{st}$	Dummy for whether State s has Time Limits in place	UI-WRD
$TLLength_{st}$	Maximum number of assistance months before benefits are terminated in place in State s	UI-WRD
	Other Policy Variables	
$unemp_{st}$	Unemployment Rate in State s	LAUS
$SEITC_{st}$	State earned income tax credit as a percentage from the federal EITC	TAXSIM
$SEITCrefundable_{st}$	Dummy for whether the state EITC is refundable	TAXSIM
$MHparity_{st}$	Dummy for whether the state has passed by time t a comprehensive or mandate offering law.	NAMI & NCLS

Note: LAUS refers to the Local Area Unemployment Statistics data provided by the Bureau of Labor Statistics, MORG refers to the CPS Merged Outgoing Rotation Groups data provided by the National Bureau of Economic Research using data from the Current Population Survey (CPS), UI-WRD refers to the Urban Institute's Welfare Rules Database, TAXSIM refers to the National Bureau of Economic Research's TAXSIM program data on State Earned Income Tax Credits, NAMI refers to the National Alliance on Mental Illness parity laws table, and NCLS the National Conference of State Legislatures mental health benefits state mandates table.

Table 1.5: MECHANISMS

Mechanism:	%Δ in Cognition	% Contribution
Total Effect	5.09	100.00
Δ in L.F. Participation	-0.06	-1.18
Δ in Time Investments	0.03	0.59
Δ in L.M. Productivity	0.02	0.39
Complementarity of Mental Health	3.72	73.08
Direct Effect	1.38	27.12

Notes: This table describes and decomposes the average effect of a 30% decrease in psychological distress in the overall population on children's cognitive scores at age 16. Column [1] depicts the average percentage change in children's cognitive scores due to each mechanism. Column [2] depicts the percentage contribution of each mechanism for the overall effect. The first mechanism captures the effect of maternal mental health on maternal annual hours worked, the second captures its effect on maternal time investments and the third its effect on maternal wages. The forth mechanism captures the effect of maternal mental health on the return of maternal time investments. The fifth mechanism captures the remaining effect of maternal mental health for children's cognitive development.

Table 1.6: POLICY SIMULATIONS

Policy:	%Δ in Cognition
30% $\downarrow$ in Distress	5.09
20% $\downarrow$ in Time Prod. Gap	6.31
$\uparrow$ in Income by the Median TANF Benef.	1.20

Notes: This table describes the average effect of three different policies on children's cognitive scores at age 16. That is, I compute the average percentage change in children's cognitive scores as a result of each policy. In the first policy, I decrease psychological distress by 30% in the whole population. The second policy, decreases the gap in the returns of maternal time investments across individuals by 20%. Then, I compare these policies to an increase in family income by the median TANF benefit, which was was $379 per month in 2,000. On average, this program increases children's cognitive scores by 1.2 percentage points.

Table 1.7: Psychological Distress

Constant	2.012	(0.401)
Years of Education	-0.069	(0.017)
Age	-0.077	(0.003)
Age sqrd.	0.001	(0.000)
Single	0.283	(0.078)
# of Children	0.057	(0.025)
Depressed at 17	0.561	(0.709)
White Dummy	-0.070	(0.049)
Mental Health Parity	-0.031	(0.013)
Unobserved Type 2	-0.002	(0.000)
Unobserved Type 3	-0.119	(0.017)

Notes: This table contains parameter estimates for the mental health function (Equation 1.14). It relates log maternal psychological distress to maternal observable variables, the state level mental health parity law, and the time-invariant unobservable types. Bootstrap standard errors are reported in parentheses.

Table 1.8: Cognition Production Function

	Parameter	First Dev. Stage		Second Dev. Stage	
Total Factor Productivity	K	0.595	(0.017)	0.938	(0.035)
Self Productivity	α_1	0.140	(0.158)	0.484	(0.128)
Log-Family Income	α_2	0.092	(0.001)	0.012	(0.006)
Log Maternal Time Investment	α_3	0.085	(0.005)	0.009	(0.010)
Log Psychological Distress	α_4	-0.037	(0.013)	-0.025	(0.020)
Log-A_{t-1} × Log-F.Income	α_5	0.001	(0.001)	0.000	(0.001)
Log-A_{t-1} × Log-M.Time	α_6	0.002	(0.003)	0.004	(0.006)
Log-A_{t-1} × Log-Distress	α_7	-0.005	(0.008)	-0.002	(0.013)
Log-F.Income × Log-M.Time	α_8	-0.001	(0.000)	0.000	(0.000)
Log-F.Income × Log-Distress	α_9	-0.012	(0.010)	0.015	(0.031)
Log-M.Time × Log-Distress	α_{10}	-0.052	(0.007)	-0.033	(0.014)
Unobserved Type 2	κ_2	0.003	(0.000)	0.003	(0.000)
Unobserved Type 3	κ_3	-0.093	(0.002)	-0.093	(0.002)

Notes: This table contains parameter estimates for the child human capital production function (Equation 1.4). The translog function relates child cognitive scores to parental investments in the form of maternal time investments,family goods investments measured by family income and the mother's mental health measured by a psychological distress scale. Time-invariant unobservable types control for unobserved investments. Bootstrap standard errors are reported in parentheses.

Table 1.9: Initial Child Ability

Constant	-1.279	(0.113)
Mother's Years of Education	0.044	(0.005)
Mother's Age at Child's Birth	0.012	(0.002)
Single	-0.001	(0.000)
# of Siblings	-0.037	(0.031)
White Dummy	0.059	(0.042)
Female	0.146	(0.094)
Unobserved Type 2	0.014	(0.000)
Unobserved Type 3	-0.116	(0.008)

Notes: This table contains parameter estimates for the initial child human capital function (Equation 1.5). Children's initial human capital is assumed to depend on child and mother's observable characteristics as well as time-invariant unobservable types. Bootstrap standard errors are reported in parentheses.

Table 1.10: Labor Market Participation

Constant	337.679	(59.698)
Years of Education	66.325	(4.465)
Age at Child's Birth	-0.041	(0.025)
Single	-18.249	(10.617)
# of Children	-156.983	(25.552)
White Dummy	335.792	(45.755)
Child is Female	-33.106	(16.214)
Child's Age	53.541	(7.212)
Child's Age sqrd.	-2.033	(0.513)
Log Non-Labor Income	-0.023	(0.015)
Median State Service Wage Rate	-35.855	(6.664)
State % Employed in Serv. Sector	243.108	(72.271)
State Variation in Welfare Rules 1	114.095	(26.980)
State Variation in Welfare Rules 2	-12.459	(8.469)
Child Younger Than 4	-12.587	(7.853)
Psychological Distress	16.951	(9.948)
Not Working Last Period	-1.020	(1.219)
Experience	2.338	(1.850)
Hours Working Last Period	0.469	(0.040)
Hours With the Child Last Period	-4.394	(1.935)
Unobserved Type 2	216.167	(42.910)
Unobserved Type 3	347.630	(35.791)

Notes: This table contains parameter estimates for the approximated decision rule for labor market participation (Equation 1.12). It relates annual hours of work to all the state variables in the Conceptual model as well as the time-invariant unobservable types. Bootstrap standard errors are reported in parentheses.

Table 1.11: Weekly Time Investments

Constant	26.799	(6.219)
Years of Education	0.512	(0.385)
Age at Child's Birth	-0.016	(0.009)
Single	-3.780	(1.478)
# of Children	0.011	(0.006)
White Dummy	-6.026	(1.317)
Child is Female	2.194	(1.231)
Child's Age	-0.556	(0.255)
Child's Age sqrd.	0.091	(0.014)
Log Non-Labor Income	-0.589	(0.506)
Median State Service Wage Rate	-2.080	(1.100)
State % Employed in Serv. Sector	18.437	(5.629)
State Variation in Welfare Rules 1	0.019	(0.007)
State Variation in Welfare Rules 2	69.014	(11.989)
Child Younger Than 4	8.887	(2.276)
Psychological Distress	-0.067	(0.031)
Not Working Last Period	0.145	(0.144)
Experience	-0.062	(0.087)
Hours Working Last Period	0.001	(0.000)
Hours With the Child Last Period	-0.061	(0.024)
Unobserved Type 2	2.715	(2.168)
Unobserved Type 3	-3.279	(2.400)

Notes: This table contains parameter estimates for the approximated decision rule for mothers' time investments in their children (Equation 1.12). It relates weekly maternal active time with children to all the state variables in the Conceptual model as well as the time-invariant unobservable types. Bootstrap standard errors are reported in parentheses.

Table 1.12: Hourly Wages

Constant	-1.847	(0.285)
Years of Education	0.194	(0.011)
Age	0.018	(0.002)
Age sqrd.	-0.000	(0.000)
Median State Service Wage Rate	0.010	(0.004)
State % Employed in Serv. Sector	1.522	(0.318)
Log Psychological Distress	-0.002	(0.001)
Not Working Last Period	-0.151	(0.188)
Experience	0.001	(0.000)
Unobserved Type 2	0.082	(0.012)
Unobserved Type 3	-0.119	(0.016)

Notes: This table contains parameter estimates for the wage process (Equation 1.9). It relates log hourly wages to maternal observable variables, state level labor market conditions, and the time-invariant unobservable types. Bootstrap standard errors are reported in parentheses.

Table 1.13: Type Probabilities

	Unobserved Type 1		Unobserved Type 2	
Constant	-0.004	(0.000)	-0.012	(0.000)
Years of Education	-0.021	(0.000)	-0.002	(0.000)
Depression before 17	-0.010	(0.000)	-0.022	(0.000)

Notes: This table contains parameter estimates for the time probability equation described in Equation 1.16.

Chapter 2

The Economic Value of *Breaking Bad*: Misbehavior, Schooling and the Labor Market

This chapter is joint work with Nicholas W. Papageorge and Yu Zheng.[1]

2.1 Introduction

Economists generally recognize that human capital consists of multiple skills that drive educational and labor market outcomes. An early contribution is Willis and Rosen (1979), who distinguish between academic and manual skill. More recently, a burgeoning literature in economics has extended the concept of human capital to incorporate non-cognitive skills such as perseverance and grit

[1]We gratefully acknowledge helpful comments from: Robert Barbera, Pedro Carneiro, Sarah Cattan, Flavio Cunha, Seth Gershenson, Donna Gilleskie, Barton Hamilton, Hans von Kippersluis, Patrick McAlvanah, Robert Moffitt, Albert Park, Richard Spady and Sergio Urzua along with seminar participants at the City University of Hong Kong, Tinbergen Institute, the Brookings Institution, Georgetown University, Universtiy of Western Ontario, European University Institute, the Econometric Society World Congress and SOLE 2016. The usual caveats apply.

(Heckman and Rubinstein, 2001).[2] It is not controversial that returns to skills can differ across sectors and that some skills are more productive in schooling than in work or in one occupation than in another. For example, to explain career choices, Willis and Rosen (1979) emphasize variation in the returns across occupations to manual versus academic skill.[3]

Despite potential differences in returns, however, the skills that constitute human capital are typically seen as all enhancing productivity—both in school and on the labor market. This is likely true for cognition and for many non-cognitive skills such as grit, which captures goal setting (Duckworth et al., 2007). However, this view overlooks how some components of human capital could be productive in some economic contexts but could actually be counter-productive in others. If so, then policies designed to promote human capital accumulation could have mixed effects or even negative economic consequences. This is especially the case for policies that target non-cognitive skill formation aimed at children or adolescents, for whom non-cognitive skills have been shown to be relatively malleable (Heckman and Kautz, 2013).

In this paper, we demonstrate that some components of childhood misbehavior predict higher earnings even though they are associated with lower schooling attainment. In particular, we examine a widely-studied pair of non-cognitive

[2]Excellent summaries of this research are found in Borghans et al. (2008) and Almlund et al. (2011).

[3]This point has its origins in Roy's model of selection into occupations (Roy, 1951).

skills known as *externalizing behavior* and *internalizing behavior*.[4] Externalizing behavior is linked to aggression and hyperactivity while internalizing behavior captures anxiety, depression, shyness, unassertiveness and fearfulness (Ghodsian, 1977; Duncan and Magnuson, 2011; Duncan and Dunifon, 2012). Using a longitudinal data set from Britain, the National Child Development Survey (NCDS), we estimate an econometric model relating childhood misbehavior to educational attainment and labor market outcomes. The two types of non-cognitive skills are identified from teachers' reports of misbehavior or maladjustment among schoolchildren. We approximate schooling, hours of work and wages using linear-in-parameters equations and we model correlation across equations as unobserved heterogeneity in the form of three latent factors: externalizing and internalizing behavior, capturing non-cognitive skills, and cognition. We also estimate the model separately for males and females. The key empirical fact we establish is that, for both genders, one of the factors underlying observed classroom misbehavior, externalizing behavior, lowers educational attainment, but is also associated with higher earnings.[5] In other words, we demonstrate that a penchant for *breaking bad* can be good.[6]

Our results have implications for our understanding of the skills comprising human capital. Whereas previous work has recognized variation in skill

[4] Regarding the nomenclature: "externalizing behavior" and "internalizing behavior" describe the latent factors interpreted as two "non-cognitive skills" identified using childhood misbehavior.

[5] Levine and Rubinstein (2013) demonstrate that some teenagers who engage in risky or illicit behaviors go on to do well in entrepreneurship.

[6] According to www.urbandictionary.com the definition of the term *breaking bad* is to "challenge conventions" or to "defy authority". *Breaking Bad* is also the title of an American television show in which the protagonist is an unsuccessful chemist who reveals a striking talent for producing illicit drugs. The show offers an extreme example of how certain skills or behaviors may lead to low productivity in one sector and high productivity in another.

prices across economic sectors, our findings on externalizing behavior go further, demonstrating that a single skill can be productive in some economic contexts and counter-productive in others. Identifying a skill that raises wages, but lowers educational attainment is a particularly striking illustration since it runs counter to the typical view of ability bias in estimates of the returns to education (Becker, 1967). Generally, the presumption is that the unobserved skills leading to success in education would also promote earnings.[7] In line with this assumption, among individuals in our sample, we demonstrate that schooling predicts higher earnings; that internalizing behavior predicts lower education attainment and lower earnings; and that cognition predicts higher degrees and higher earnings. In contrast, externalizing behavior has mixed effects. Despite its negative impact on schooling, it is also associated with higher wages for males and females and with more hours in the labor market for females.[8]

An important question is whether externalizing behavior is a direct determinant of earnings or whether it merely influences some third variable or variables which then affect earnings. One obvious example is selection into employment. Suppose unproductive high-externalizing individuals select out of the labor market. Then estimates of positive labor market returns to externalizing behavior could be an artifact of differential sorting, which would undermine the idea that a valuable component of human capital is counterproductive in some economic

[7]There are a number of exceptions. For example Card (2012) shows that IV estimates could lead to larger coefficients on education in wage equations. The argument is based on heterogeneity in treatment effects coupled with the particular group for whom the IV affects attendance.

[8]Several studies have examined the relationship between these two behaviors to better known measures like the "Big 5" personality traits. Evidence suggests that externalizing behavior is related to conscientiousness, agreeableness, and openness to new experience, while internalizing behavior is mostly related to neuroticism (Ehrler, Evans, and McGhee, 1999; Almlund et al., 2011).

contexts. In a series of sensitivity analyses, we therefore assess how externalizing behavior predicts labor supply, occupation choice, work experience, fertility and partnership. While we show evidence that externalizing behavior is strongly related to many of these economic outcomes, we also demonstrate that these relationships do not drive our main finding that externalizing behavior, despite being unproductive at school, is productive in the labor market.

Another important question is whether the returns to non-cognitive skills vary across socioeconomic groups. This question is partly motivated by findings in Heckman, Pinto, and Savelyev (2013), who show that an early childhood intervention (the Perry Preschool Program) raised earnings and that about 20% of this rise is attributable to a reduction in externalizing behavior. In contrast, we find that, for a 1958 British cohort, externalizing behavior raises earnings. To explore this difference, we consider a sub-sample of the NCDS British cohort that is selected to mimic the financially disadvantaged group studied in Heckman, Pinto, and Savelyev (2013).[9] We show that, among individuals who grew up in poverty, externalizing behavior carries no significant earnings premium. This finding is in line with Lundberg (2013), who demonstrates that the payoff to non-cognitive skills is context-dependent and may vary by socioeconomic status. One possible reason is selection into criminality (Aizer, 2009; Heckman, Pinto, and Savelyev, 2013). However, for our sample, we find that differential sorting into police involvement does not appear to drive differences in returns to externalizing behavior across socioeconomic groups. Therefore, we cannot rule out the possibility that some skills are simply priced differently in the labor

[9] To study black-white differentials in labor market outcomes in the U.S., Urzua (2008) allows the distribution and impact of underlying skills to vary by race.

market depending on an individual's background. This possibility is troubling if it means that individuals who are already disadvantaged are excluded from realizing the full returns to their skills.

Turning to policy, mixed effects of externalizing behavior suggest that a productive labor market skill may be easily overlooked or difficult to detect or foster since it is not productive in school. Relatedly, our findings also point to a mismatch between the types of skills promoted in school and the skills that are valuable on the labor market. This point echoes findings in Heckman and Rubinstein (2001), who show that the GED is a "mixed" signal of productivity since it is taken by high school dropouts with low non-cognitive skill. As a result, educational attainment or certification is a potentially flawed signal of a future worker's productivity. An informative signal should be designed to accurately reflect all skills that are productive in the labor market. Similarly, in our context, if externalizing behavior carries an earnings premium, then at the very least it should not carry an education penalty. More generally, our findings illustrate that broadening our understanding of what constitutes human capital, which the literature on non-cognitive skills has done, also opens up the possibility that some human capital investments can have negative economic returns in some sectors. Despite the positive returns to educational attainment, investments designed to curb or eliminate childhood misbehavior may be ill-conceived or short-sighted since a subset of children who misbehave may be expressing non-cognitive skills that are valuable in the labor market. This is not a hypothetical concern since school districts are increasingly poised to begin using high-stakes tests to evaluate students, teachers and schools based on character or non-cognitive skills (West et al., 2015). Finally, our findings on group differences imply further difficulties in evaluating human capital investments involving children's non-cognitive skills since the returns to skills can differ not only by the economic context in question, but also by socioeconomic status.

The paper is organized as follows. In Section 2.2, we introduce the data set, discuss measurements of misbehavior that identify externalizing and internalizing behavior and conduct a preliminary data analysis. In Section 2.3, we describe the econometric framework and estimation. In Section 2.4, we present results. In Section 2.5, we discuss what our findings mean for policy. Section 2.6 explores policy implications and Section 2.7 concludes.

2.2 Data and Preliminary Analysis

First, we provide details on the NCDS and on how we construct the analytic sample. Second, we discuss how classroom behavior is used to identify two latent factors: externalizing and internalizing behavior. Third, we report summary statistics on education, labor market outcomes and childhood misbehavior in the classroom. Fourth, we provide estimates from a preliminary econometric model relating childhood misbehavior with schooling and earnings. In particular, we demonstrate that once we treat externalizing and internalizing behaviors separately, externalizing behavior is associated with higher earnings even though it also predicts lower schooling attainment.

2.2.1 The National Child Development Study

The NCDS is an ongoing longitudinal survey that follows the universe of individuals born in the same week in 1958 in Great Britain. The data set contains information on physical and educational development, wages, employment, family life, well-being, social participation and attitudes. The NCDS is particularly well-suited for our study since it documents teachers' reports of classroom misbehavior for a large sample of children and then follows these children through adulthood. Therefore, the data set allows us to relate misbehavior in elementary school to educational attainment along with labor market outcomes. To date, there have been eight surveys to trace all the members of the cohort still living in Great Britain. Surveys occurred when subjects were born and when they were aged 7 (1965), 11, 16, 23, 33, 42 and 50 (2008).

We focus on information gathered at birth and in the first five sweeps, covering ages 7 to 33. The NCDS initially contained 18,555 births. At the second wave, 15,356 of the original sample remained as respondents and by the fifth survey, at age 33, 11,407 individuals remained. In constructing our analytic sample, we keep respondents with valid information on test scores and classroom misbehavior at age 11 and educational attainment at age 33, which leaves us with 9,511 individuals. We drop individuals with missing information on relationship status, fertility and employment status at age 33. We also drop individuals with missing information on their employment history or who are reported as employed but have missing information on earnings at age 33.[10] The resulting analytic sample has complete information on 7,324 individuals, of whom 3,612 are males and 3,712 are females.[11]

[10]We drop individuals when there is missing information on one of the key outcome variables used in our analysis. However, we impute data for missing control variables, including parents' education and occupation.

[11]To assess whether sample attrition drives our main results, we compare our final sample to the sample of individuals observed at age 11, to which we refer as the "full sample". We report the summary statistics of the full sample in Tables B.1 and B.2 in B.1. Compared to the full sample, our analytic sample is slightly less educated, more likely to be self-employed

2.2.2 Classroom Misbehavior and Non-Cognitive Skills

While cognitive skills are identified, as usual, by math and reading test scores, non-cognitive skills and classroom misbehavior are identified with inventories completed by teachers describing student behavior in the classroom. When a child in the sample was 11 years old, the child's teacher was asked to complete an inventory listing the child's behaviors in the classroom. The teacher was given a list of roughly 250 descriptions of specific behaviors and asked to underline the items which best fit the child. These descriptions include statements such as: "too timid to be naughty"; "brags to other children"; "normally honest with school work"; "adopts extreme youth fashions"; and "has stolen money". Completed inventories were then used to compute scores on a set of ten maladjustment syndromes, known as the Bristol Social Adjustment Guide or *BSAG* maladjustment syndromes. The syndromes were first defined in Stott, Sykes, and Marston (1974). They are: hostility towards adults, hostility towards children, anxiety for acceptance by adults, anxiety for acceptance by children, restlessness, inconsequential behavior, writing off adults and adults standards, depression, withdrawal and unforthcomingness. The syndromes have been used since their introduction in Stott, Sykes, and Marston (1974) to assess the psychological development of children.[12] They have also been externally validated in the sense that the inventories used to measure the ten syndromes are positively correlated with a range of other measurements of social maladjustment from teachers, professional observers, parents and peers (Achenbach, McConaughy,

and more likely to have a partner at age 33. However, none of these differences are statistically significant.

[12]Unfortunately, the NCDS does not provide access to the original completed inventories. We only have access to the computed maladjustment syndrome scores.

and Howell, 1987).

Using principle components factor analysis, Ghodsian (1977) showed that the BSAG maladjustment syndrome variables could be described by two distinct latent factors.[13] Ghodsian (1977) also proposed a mapping between the measurements and the two factors, which gives a meaningful interpretation to each one. The mapping assigns each observed maladjustment syndrome to one of the two factors. According to the mapping, the first factor corresponds to anxious, aggressive, outwardly-expressed or *externalizing behavior* and includes maladjustment syndromes such as "hostility towards adults" and "restlessness". The second factor corresponds to withdrawn, inhibited or *internalizing behavior* and includes maladjustment syndromes such as "depression".[14] In Table 2.1, we list each factor along with the maladjustment syndromes used to identify it.[15] The two factors have been studied extensively by psychologists researching child development and, of late, by some economists (Blanden, Gregg, and Macmillan, 2007; Aizer, 2009; Agan, 2011; Heckman, Pinto, and Savelyev, 2013).[16]

Though there is some debate on the key assumptions underlying the structure of the mapping between maladjustment syndromes and underlying factors,

[13] In B.1, we use our sample to confirm that two independent factors adequately describe data on the ten BSAG maladjustment syndromes.

[14] In B.1, we confirm that a standard rotation method reveals that variables such as "hostility towards adults and children" and "inconsequential behavior" represent outwardly-expressed behaviors and are strongly related to the first factor in the factor analysis. This factor represents externalizing behavior. Observed maladjustment syndromes such as "depression," "unforthcomingness" and "withdrawal" represent inwardly expressed behaviors and are strongly related to the second factor in the factor analysis.

[15] Syndromes we do not use are "miscellaneous nervous symptoms, "miscellaneous symptoms", "appearance", "attendance" and "health factors". In results available upon request, we repeat our analysis using "miscellaneous nervous symptoms" and "miscellaneous symptoms" and find no significant differences in results.

[16] Both Aizer (2009) and Agan (2011) study how externalizing behavior is linked to antisocial and criminal activity. For general surveys on the use of externalizing and internalizing behaviors, see Duncan and Magnuson (2011) and Duncan and Dunifon (2012).

in our analysis, we generally follow Ghodsian (1977).[17] We use 10 BSAG maladjustment syndrome variables to identify two latent factors, externalizing and internalizing behavior. We also assume dedicated measurements, which means that each measurement is related to only one unobserved factor, even though this assumption is not required (Rao and Sinharay, 2006).[18] Following Ghodsian (1977) has the advantage of making our results comparable to other work studying childhood misbehavior and economic outcomes. This includes work using the NCDS data set studying externalizing and internalizing behaviors (Farmer, 1993, 1995; Jackson, 2006). It also includes research using different samples since the division of misbehavior into these two factors now extends to other data sets, including the CNLSY and the PSID (Yeung, Linver, and Brooks-Gunn, 2002; Agan, 2011).

2.2.3 Summary Statistics

Our three key variables are those for educational attainment, labor market outcomes and childhood misbehavior. For the first of these, we note that, in the UK, schooling is compulsory until age 16. Thereafter, students can leave school without any qualifications (no certificate), study for an exam to obtain a Certificate of Secondary Education (CSE) or study towards obtaining the Ordinary

[17]For example, while traditional factor analytic methods determine the number of unobserved factors in one step and the mapping in a second step, newer Bayesian methods estimate the number of factors and their mapping to the measurement system simultaneously (Conti et al., 2014). In B.1, we discuss issues surrounding factor analytic methods for childhood misbehavior in greater detail.

[18]The exception is "writing off adults and adult standards", which could represent an outwardly or inwardly expressed behavior and is statistically related to both factors. In this case, we again follow previous work and allow the variable to be related to both factors (Ghodsian, 1977; Shepherd, 2013). We also perform robustness checks where we assign the ambiguous variables to either factor. Results remain largely unchanged.

Levels (O-Levels), where the latter are more academically demanding.[19] If students decide to stay in school at age 16, another set of examinations is available, the Advanced Levels (A-Levels). Students who are successful in their A-Levels are able to continue to higher education and obtain either a higher-education diploma (after two years of study) or a bachelor's degree (after three years of study). At the postgraduate level, students can obtain a higher degree: Master of Philosophy (MPhil) or Doctor of Philosophy (PhD). In summary, individuals in our sample can sort into six mutually exclusive schooling levels: no certificate, CSE, O-Levels, A-Levels, higher education (including diploma and bachelors) or higher degree (including MPhil and PhD).

Summary statistics on education, labor market and other adult outcomes are found in Table 2.2. Perhaps most striking are large gender differences, which reflects the fact that the analytic sample is a 1958 cohort. According to the table, females in our sample are less educated than the males. Roughly half of the males obtain O-Level qualifications or less, whereas roughly two-thirds of the females do. We also find large gender differences in employment and, conditional on employment, hourly wages and hours worked. Conditional on working, hourly wages average about 7.64 pounds for males and 5.46 pounds for females and weekly earnings average 329 pounds for males and 162 pounds for females, all measured in 1991 pounds. Differences in educational attainment only offer a partial explanation for labor market disparities. In Figure 2.1, we show that, at each education level, males have higher wages, work longer hours and earn more. Despite these differences, the relative returns to schooling are

[19] CSEs and O-Levels were replaced by the General Certificates of Secondary Education (GCSE) in 1986 after individuals in our sample had finished their schooling.

higher for females than for males (Panel 2.1(d) in the same figure). Females with a higher degree earn 3 times as much as females with no formal education, while for males this ratio is 1.75. This may reflect gender differences in how individuals sort into schooling based on their cognitive and non-cognitive skills or differences in skill prices across genders, both of which our econometric analysis will account for.

Another factor explaining differences in labor market outcomes is fertility. According to Table 2.2, females and males in our sample are equally likely to have a partner, though females are roughly 20% more likely to have children. A stark gender difference emerges if we compare the earnings of males and females with and without children. Females with children work many fewer hours than those without children. For males, having children in the household predicts no drop in labor supply. These patterns can be observed in Figure 2.2. Accordingly, our econometric analysis will consider the role of partnership and fertility in mitigating the gender-specific relationship between non-cognitive skills and earnings. In general, large gender differences in schooling and labor market outcomes suggest that we should allow the parameters of our econometric model to vary for males and females.

In Table 2.3, we present averages for each BSAG variable separately by gender. Values of the BSAG variables ranges from 0 to 15, with a higher value indicating a higher prevalence of a particular maladjustment syndrome. These scores were constructed using the teacher responses to particular statements about the student's behavior. The means are usually low due to a clustering around zero and fairly low values in general. Nonetheless, there are significant differences across gender. In general, females appear to misbehave less frequently than males. Specifically, males score higher for each of the BSAG variables except for "anxiety for acceptance by adults". For example, for "inconsequential behavior" and "anxiety for acceptance by children", the average for males is roughly double that of females. Gender differences in misbehavior are consistent with earlier findings documented in Great Britain (Duncan and Magnuson, 2011; Duncan and Dunifon, 2012) and in the U.S. (Bertrand and Pan, 2013).

2.2.4 Misbehavior, Schooling and Earnings

In our main econometric analysis, we jointly estimate the mapping from unobserved factors to observed BSAG maladjustment syndrome variables with the impact of these factors on outcomes. However, for our preliminary analysis conducted here, we construct measures for externalizing and internalizing behaviors by simply summing the BSAG variables associated with each factor according to Table 2.1 and then including the sums as additional regressors in models where outcomes are schooling categories and earnings. We refer to these as "crude" models since summing up scores likely inflates measurement error and ignores correlation across factors. We provide estimates from the crude model to compare our findings with previous work and to demonstrate that main results, in particular mixed effects of externalizing, are not driven by the factor analytic methods used to estimate the measurement system in our main econometric analysis.

We start by estimating an ordered probit model to explain educational attainment. The outcome variable is one of the six possible schooling levels.[20]

[20]Our results are robust to the specification of the educational model. The main message

Formally, defining s_i^* as a latent variable determining schooling, we estimate regressions of the following form:

$$s_i^* = E_i\psi^{\mathbf{E}} + I_i\psi^{\mathbf{I}} + C_i\psi^{\mathbf{C}} + X_i'\beta_s + e_i^S \tag{2.1}$$

where observed schooling $s_i = s$ if $\mu_L^s \leqslant s_i^* < \mu_H^s$ and μ_L^s and μ_H^s are the particular bounds for schooling level s. E_i and I_i are the measures of externalizing and internalizing behaviors based on a simple summation of the BSAG scores. Similarly, C_i is based on the sum of the reading and math test scores listed in Table 2.1. Here, and elsewhere, we normalize our measures of cognitive and non-cognitive skills with mean equal to 0 and variance equal to 1. Finally, X_i is a vector of covariates and e_i^S is a normally distributed disturbance.

Estimates of equation (2.1) are presented in Table 2.4 for varying sets of covariates X_i. Column [1] contains an indicator for being female and a single measure of misbehavior, obtained by summing E_i and I_i for each individual. Aggregating misbehavior into a single variable allows us to compare our results to earlier research that relates childhood misbehavior to economic outcomes, but which ignores how childhood misbehavior is driven by two separate factors, reflecting two distinct non-cognitive skills. We find that misbehavior predicts lower educational attainment. In Column [2], we add cognition C_i, which is associated with higher education. Including cognition decreases the magnitude of the negative coefficient on misbehavior from -0.37 to -0.14, which suggests strong correlation in measurements of cognition and childhood behavior. In Columns [3] and [4], we again address misbehavior and schooling with and without cognition, though here we separate misbehavior into externalizing

remains when we use multinomial probit model instead.

and internalizing behavior. Both non-cognitive skills predict lower educational attainment and the inclusion of cognition decreases the magnitude of the coefficients by over half. In Column [5], we assess the robustness of estimated coefficients on non-cognitive skills to the inclusion of a number of controls including parents' education, father's social class and whether the mother is working. As expected, higher parental education and occupation are positively related to a higher educational attainment. Coefficients on the three skills, however, remain largely unchanged. Finally, we estimate the schooling model separately for males (Column [6]) and females (Column [7]). We show that the negative effects of externalizing are larger for males (-0.12 versus -0.06). For females, internalizing has a larger effect (-0.09 versus -0.04). Cognition has a slightly larger coefficient for females. Importantly, both non-cognitive skills predict less education, while cognition predicts higher educational attainment even when we estimate the crude model separately by gender.

We perform a similar analysis for earnings, regressing log weekly earnings at age 33, conditional on being employed, onto measures of non-cognitive skills.[21] Defining y_i as log earnings at age 33 for individual i, we estimate OLS regressions of the following form:

$$y_i = E_i\phi^{\mathbf{E}} + I_i\phi^{\mathbf{I}} + C_i\phi^{\mathbf{C}} + X_i'\beta + e_i^Y \tag{2.2}$$

where cognitive and non-cognitive skills are defined as in equation (2.1) and e_i^Y is an iid disturbance.[22] The results from OLS regressions for varying sets of

[21]Conditioning on employment raises the possibility that the preliminary results are driven by compositional effects, which we address in Section 2.5.1. In particular, we assess whether the positive association between externalizing behavior and earnings can be explained by high-externalizing and low productive individuals selecting into unemployment.

[22]We include a London indicator to account for possible earnings differences arising from cost-of-living. Omitting it does not affect results.

regressors are found in Table 2.5. Column [1] contains estimates using the single aggregated measure of misbehavior obtained by summing E_i and I_i. We find that aggregate misbehavior is associated with lower earnings, which is in line with previous research (Segal, 2013). Column [1] suggests that a one-standard-deviation rise in aggregated misbehavior is associated with a 10.5% decline in earnings at age 33. In Column [2], we add cognition to the regression. A one-standard-deviation rise in cognition predicts a 20.5% increase in earnings. Further, adding cognition lowers the coefficient on aggregated misbehavior to 2.8%. This sharp decline in the magnitude of the coefficient means that our measures of misbehavior are related to our measures of cognition. Our main econometric analysis explicitly treats observables as correlated measurements of underlying factors and also permits correlation among the latent factors capturing cognition and non-cognitive skills.

Results on misbehavior change dramatically, however, when we view childhood misbehavior as reflecting two distinct factors. In Columns [3]-[6] of Table 2.5, we regress log earnings onto E_i and I_i separately. Beginning with Column [4], where we also condition on cognition, gender and a London indicator, we find that externalizing behavior predicts higher earnings. In other words, externalizing behavior, as a non-cognitive skill, appears to carry an earnings premium. Adding schooling, we find that higher degrees predict higher earnings (Column [5]). Moreover, the positive coefficient on externalizing rises once we control for educational attainment, which makes sense since adding schooling helps to control for how externalizing could lower earnings through its negative impact on schooling. Next, we add fertility, partnership, experience and occupation (Column [6]). All controls are positively related to earnings with the exception of number of children for females. After adding these controls, the coefficient on externalizing rises once again, which suggests that the association between externalizing and earnings might work through its relationship with other lifecycle outcomes, such as fertility. We explore this possibility explicitly when assessing mechanisms and selection in Section 2.5.1. Finally, in all models from Columns [4]-[8], cognition continues to predict higher earnings while internalizing behavior is associated with lower earnings.

It is also worth highlighting that, according to Table 2.5, the coefficient on externalizing is positive whether or not we control for schooling. An alternative possibility would be that externalizing behavior predicts higher earnings only after we have controlled for its negative impact on schooling. Such a finding would still support the idea that externalizing is potentially valuable in the labor market. However, it would also suggest that lowering externalizing behavior could have a positive net effect on labor market outcomes since the negative effect of externalizing through schooling on earnings would overwhelm the direct positive effect on earnings. In contrast, our estimates suggest that policies lowering externalizing would have a negative effect on individual earnings even if we account for how externalizing affects schooling.

We also estimate the earnings regression separately for males and females. We find the externalizing earnings premium is more pronounced for females than for males (see Columns [7] and [8] for males and females, respectively). In terms of the magnitude, once we control for educational attainment, partnership and fertility, the coefficient on externalizing behavior for females is comparable to the coefficient on cognition. Gender differences in coefficients may reflect true heterogeneity in returns, but could also reflect instability of the measurement of the two factors. In our main empirical analysis, we account for the second possibility by estimating the measurement system mapping latent factors to measurements of misbehavior separately by gender.

The crude model results presented in Tables 2.4 and 2.5 provide preliminary evidence of our main result. A non-cognitive skill that is productive on the labor market is not productive in school. The positive association between externalizing behavior during childhood and adult earnings has not been recognized in previous literature on the economic consequences of childhood misbehavior.[23] There are several reasons for this lack of recognition. First, most of the literature on the long run effects of childhood misbehavior takes for granted that externalizing is broadly unproductive, focusing instead on school-related outcomes (Bertrand and Pan, 2013). This may be a result of data limitations since linking childhood misbehavior to labor market outcomes requires a long panel beginning with a sample of children. However, even studies using the NCDS data have not linked externalizing behavior to earnings. For example, Farmer (1993, 1995) shows that males who display high levels of externalizing behavior leave school earlier, obtain fewer qualifications, and begin their careers in lower social class positions. However, neither study considers actual performance in the labor market.[24] Second, many studies use a single, aggregated measure of misbehavior. Segal (2013) shows that misbehavior during the eighth grade can

[23] One exception is Levine and Rubinstein (2013), who recognize an empirical pattern that is similar in spirit. They show that individuals who engage in illicit behaviors as teenagers report high earnings in self employment. One possible extension to our research would be to assess whether the successful entrepreneurs they identify were high-externalizing children.

[24] Nor do these studies control for internalizing behavior, which means they may suffer from omitted variables bias if the two are correlated. Other work from psychology and sociology uses the NCDS data to examine selection into occupations. Jackson (2006) shows that having low levels of internalizing behavior is an important predictor of managerial occupations.

have a negative impact on future earnings even after controlling for schooling attainment and Sciulli (2016) demonstrates that adult employment outcomes are negatively related to childhood maladjustment. Also using the NCDS data set, Fronstin, Greenberg, and Robins (2005) show that a single measure of misbehavior predicts lower earnings in adulthood. As we demonstrate in Columns [1] and [2] of Table 2.5, using the NCDS data set, we can replicate the basic result that aggregated misbehavior predicts lower earnings. However, estimates from our crude model also demonstrate how key implications change dramatically once we recognize that misbehavior reflects two distinct factors with potentially different returns in the labor market. Building on our preliminary analysis, we now turn to the specification of our main econometric framework, which treats observed classroom behavior as mis-measurements of underlying factors.

2.3 Measurement Error Model and Inference

The preliminary analysis just presented has several shortcomings. Simply summing the BSAG maladjustment syndrome variables assigned to each underlying skill does not account for differences in explanatory power of each measurement or correlation across measurements. This can inflate measurement error, increasing attenuation bias. In what follows, we instead factor analyze the data on childhood classroom behavior, which means that we treat each BSAG variable as a mismeasurement of one of the underlying factors. Factor analysis reduces measurement error and maps underlying factors to observed variables according to the explanatory power of each variable. In addition, in the preliminary analysis we estimated equations for schooling and earnings separately. In what follows, the equations describing the relationship between skills, schooling and labor market outcomes are instead estimated jointly with equations describing how underlying skills map into BSAG variables. Joint estimation reduces estimation error.

2.3.1 Description of the Model

There are three latent skills affecting education and labor market outcomes: externalizing behavior, internalizing behavior and cognition. Each skill is measured from a set of variables with measurement error (Table 2.1). We denote the k-th measurement of skill $j \in \{1, 2, 3\}$ for individual i with gender $n \in \{0, 1\}$ as m_{ijkn}, where $n = 1$ denotes male and $n = 0$ denotes female. m_{ijkn} is specified as:

$$m_{ijkn} = \overline{m}_{jk} + \alpha_{jkn} f_{ij} + \varepsilon_{ijkn} \tag{2.3}$$

where $\overline{m}_{jk}$ is the mean for that measurement for the whole sample, which does not vary by gender.[25] f_{ij} is the value of latent skill j for individual i, α_{jkn} is the factor loading of latent skill j on the k-th measurement of that skill, which can vary by gender, and ε_{ijkn} is an error term capturing mis-measurement and it is assumed to be normally distribution for the measurements of cognition. In order to account for the high number of zero responses we assume the measurements of the two behaviors follow a Poisson distribution. The latent factors f_{ij} are drawn from a joint normal distribution with a probability density function f^M, the parameters of which can vary by gender:[26]

[25] This setup allows us to compare the latent skill mean across genders.

[26] The results are robust to allowing for more flexible distributional assumptions on the measurement errors. In particular, we have permitted mixed normal distributions with two components and obtain qualitatively similar results.

$$\begin{pmatrix} f_{i1} \\ f_{i2} \\ f_{i3} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{1,n} \\ \mu_{2,n} \\ \mu_{3,n} \end{pmatrix}, \begin{bmatrix} \sigma_{11,n} & \sigma_{12,n} & \sigma_{13,n} \\ \sigma_{12,n} & \sigma_{22,n} & \sigma_{23,n} \\ \sigma_{13,n} & \sigma_{23,n} & \sigma_{33,n} \end{bmatrix}\right) \tag{2.4}$$

Further, the model assumes that the latent skills are independent of measurement errors, or $cov(f_{ij}, \varepsilon_{ijkn}) = 0$, $\forall k$. The latent skill j' affects the measurement of the latent skill j only through its correlation with the skill j, or $cov(m_{ijkn}, f_{ij'}|f_{ij}) = 0$, for $j \neq j'$ and all k.[27]

We approximate the schooling problem with a linear-in-parameters ordered probit model, so that the probability that agent i chooses education level $s \in \{0, ..., 5\}$ is given by:

$$\mathrm{P}_i(s) = \Phi_s\left(\mu_s + X_{i,S}\beta_S + \sum_{j=1}^{3}\alpha_{j,S}f_{ij}\right) - \Phi_{s-1}\left(\mu_{s-1} + X_{i,S}\beta_S + \sum_{j=1}^{3}\alpha_{j,S}f_{ij}\right) \tag{2.5}$$

where μ_s is the cutoff for each schooling choice and where $\mu_0 = -\inf$ and $\mu_6 = \inf$. $X_{i,S}$ is the vector of observable characteristics that affect the schooling decision and β_S is the vector of returns associated with $X_{i,S}$. $X_{i,S}$ contains a number of variables that are excluded from other equations: whether the mother studied beyond the minimum schooling age, whether the father studied beyond the minimum schooling age, father's occupation and mother's employment status, all observed when the child is age 11. We also include an indicator for financial difficulties during childhood. The variable takes the value one if (i) the interviewer reported that the household appeared to be experiencing poverty in 1965 or (ii) a member of the household self-reported having financial difficulties in the 12 months prior to being observed in either 1969 or 1974, and

[27]The only exception is "writing off adults and adult standards", which depends on both externalizing and internalizing behaviors.

zero otherwise.[28]

We model the hourly wage and weekly hours worked for individuals that are employed at age 33 as follows: log hourly wage for individual i, denoted y_i, is modeled with a linear specification and a normally distributed disturbance:

$$y_i = X_{i,Y}\beta_Y + \sum_{s=0}^{5} \gamma_{s,Y}\mathbf{1}_i[s] + \sum_{j=1}^{3} \alpha_{j,Y} f_{ij} + \varepsilon_{i,Y}. \tag{2.6}$$

Here, $X_{i,Y}$ is a vector of observables that include partnership, fertility, months of experience, occupation and an indicator for financial difficulties during childhood. The log weekly working hours are modeled in a similar fashion as:

$$h_i = X_{i,H}\beta_H + \sum_{s=0}^{5} \gamma_{s,H}\mathbf{1}_i[s] + \sum_{j=1}^{3} \alpha_{j,H} f_{ij} + \varepsilon_{i,H}. \tag{2.7}$$

where β_H captures how partnership, fertility, experience and occupation (included in the vector of observables $X_{i,H}$) affect the number of hours worked in a usual week. In addition, both of the above equations include dummies of schooling levels, $\mathbf{1}_i[s]$, and the latent skills, f_{ij}.

We summarize the parameters to be estimated by a vector denoted Φ:

$$\Phi = (\beta, \gamma, \alpha, \Xi) \tag{2.8}$$

where β denotes the set of coefficients of the vectors of observables absent the schooling level in equations (2.5)-(2.7), γ is the set of coefficients governing the returns to schooling, α is the set of coefficients governing the returns to latent skills and Ξ are coefficients of the measurement system described in equations (2.3) and (2.4).

[28]Including this variable does not affect main results. However, it is included as an additional control in our main analysis since we use it to explore differences in the returns to externalizing behavior by childhood socioeconomic status in Section 2.5.2.

2.3.2 Estimation Procedure

We estimate the model by simulated maximum likelihood. There are three main steps in the estimation procedure. First, at each suggestion for parameter values, indexed by g and denoted $\Phi^{(g)}$, and for each individual i, we simulate a vector of unobserved skills K times and compute, for each draw of the skills, the probability of observing each schooling level, log weekly hours worked and log hourly wage. More specifically, given a parameter suggestion, we draw a block matrix of size $K \times I \times J$ from a standard normal distribution, where J is the number of latent skills, i.e. 3, and I is the number of individuals. Then, for each individual i and draw k, we construct a vector of latent skills $(f_{i1k}^{(g)}, f_{i2k}^{(g)}, f_{i3k}^{(g)})$. We compute the density functions corresponding to each outcome: the probability of individual i reaching a schooling level s ($\mathrm{P}_{ik}^{(g)}(s)$), the probability of observing wage y_i ($f_{ik}^{Y,(g)}(y_i)$) and hours worked h_i ($f_{ik}^{H,(g)}(h_i)$), for individual i, draw k and parameter suggestion (g). We also compute $f_{ik}^{M,(g)}(m_i)$, the probability of observing the classroom misbehavior measurements, for individual i, draw k and parameter suggestion (g).

Second, we compute each individual's average likelihood contribution, where the average is taken over the K draws:

$$\begin{aligned} L_i^{(g)} &= \tfrac{1}{K}\textstyle\sum_{k=1}^{K} f_{ik}^{M,(g)}(m_i) \times \prod_{s=0}^{5} \mathrm{P}_{ik}^{(g)}(s)^{\mathbf{1}[s=s_i]} \\ &\times\ f_{ik}^{H,(g)}(h_i)^{\mathbf{1}(e_i=1)} \times f_{ik}^{Y,(g)}(y_i)^{\mathbf{1}(e_i=1)} \end{aligned} \tag{2.9}$$

where s_i represents the observed schooling choice and e_i the observed employment status (with employed taking the value 1) in the data.

Third, we take the log of the individual likelihood contribution and sum over all individuals to form the simulated log likelihood function:

$$l^{(g)} = \sum_{i=1}^{I} \log \left(L_i^{(g)} \right) \tag{2.10}$$

Using both simplex and gradient methods, we evaluate $l^{(g)}$ at different values in the parameter space, indexing these suggestions by (g), and continue until a maximum is found. We implement this model for males and females separately.

2.4 Empirical Results

We present our key empirical findings of the econometric model just presented in three sections. We first discuss estimates of the measurement system mapping unobserved factors to observed BSAG maladjustment syndromes (Section 2.4.1). Next, we discuss key findings, including the externalizing schooling penalty (Section 2.4.2) and the externalizing earnings premium (Section 2.4.3).

2.4.1 Mapping Unobserved Factors to Observed Misbehaviors

Starting with the joint distribution of latent factors, we find a negative correlation between externalizing behavior and cognition and a positive correlation between externalizing and internalizing behavior for both males and females (Table 2.6). The negative relationship between the two non-cognitive skills and cognition could reflect the distribution of skill endowments at birth. It could also reflect early childhood investments if the same environments that promote externalizing and internalizing behaviors also slow cognitive development (Heckman and Cunha, 2007). An example would be childhood poverty. The positive relationship between externalizing and internalizing behavior is well-documented in the child development literature. Children under stress as a result of poverty or a family disruption, for example, tend to develop both aggressive and depressive symptoms (Wolfson, Fields, and Rose, 1987). Accounting for correlation across factors means that we avoid mis-attributing returns to skills. For example, failing to account for the positive association between externalizing and internalizing behavior could lead us to over-estimate the degree to which each non-cognitive skill negatively affects schooling.

In Table 2.7, we report the estimates of factor loadings mapping latent skills to BSAG maladjustment syndrome scores. Larger loadings signal a stronger relationship between the latent factor and the observed measure. Recall, we estimate the measurement system for males and for females separately. The goal is to address the possibility that estimated gender differences in returns to non-cognitive skills in school or on the labor market reflect instability of the measurement system across genders. According to Table 2.7, instability is not a very important concern since the estimated factor loadings are very similar for males and females. However, we find considerable variation across measurements. For both genders, the main variable identifying externalizing behavior is "hostility towards children" and the main variable identifying internalizing behavior is "unforthcomingness". In contrast, "writing off of adults and adult standards", for example, is relatively unimportant for both non-cognitive skills.

Using estimates of the distributions of underlying factors, we next plot the gender-specific distributions of each latent skill in Figure 2.3. We find little evidence of gender differences in the distribution of internalizing behavior or cognition. For externalizing behavior the mean and variance are higher for males. Our findings are consistent with earlier literature studying gender differences in misbehavior. However, since earlier literature has taken for granted that externalizing is broadly unproductive, the rightward-shifted externalizing distribution for boys has been viewed as a cause for concern (Bertrand and Pan, 2013).[29]

2.4.2 The Externalizing Penalty in School

Estimates of the ordered probit model for educational attainment are reported in Table 2.8. We find a negative association between externalizing behavior and schooling for both males and females and the point estimates are of a similar magnitude compared to findings in our crude model. The effect of family characteristics is also consistent with our initial expectations. Having parents with more education and who work in more lucrative occupational categories is related to higher educational attainment of the child. Moreover, individuals living in poverty during their childhood, a measure of family resources, are less likely to reach higher levels of education.

A difference from the crude model estimates is that the negative relationship between externalizing and schooling for females is smaller and no longer significant at conventional levels. In other words, high-externalizing females are better able to finish school in comparison to high-externalizing males. This finding may reflect how teachers are more likely to punish or refer a male versus a

[29]The difference in the distribution of externalizing behavior between males and females coupled with positive returns to externalizing in the labor market raises the possibility that differences in externalizing behavior could explain the gender earnings gap. In results available upon request, we show that this is not the case. The gender earnings gap closes only slightly if we assign females the same distribution of externalizing behavior as males.

female child for special help for the same level of aggressive behavior (Gregory, 1977). On the other hand, we find that internalizing behavior is negatively associated with educational attainment for females, but not for males, for whom the coefficient is both small and insignificant. This is also on par with research that finds stronger effects of conduct disorders and weaker effects of anxiety and depressive symptoms for the educational attainment of males in comparison to females (Kessler et al., 1995). Finally, it is worth mentioning that even the largest coefficients on non-cognitive skills in the schooling equations are between one-fifth and one-tenth the size of coefficients mapping cognition to educational attainment, which predicts schooling at similar magnitudes across genders.

In general, estimates for the schooling model are broadly consistent with literature that studies the impact of emotional problems in school. For example, McLeod and Kaiser (2004) argue that children with internalizing and externalizing problems withdraw from social relationships in school, including those with teachers, in order to minimize their exposure to negative interactions. Moreover, one of the key pathways relating behavioral problems to low educational attainment is through early educational failures such as repeating a grade or falling behind in class. If externalizing or internalizing behavior make learning more difficult, this would explain the strong negative correlation between the two non-cognitive skills and cognition (which is identified from test scores) reported in Table 2.6.

2.4.3 The Externalizing Premium on the Labor Market

Literature studying the consequences of externalizing behavior has generally limited attention to educational attainment. In contrast, we assess the relationship between childhood misbehavior and labor market outcomes. Estimates of hours and wage equations conditional on employment are in Tables 2.9 and 2.10.[30] For males, a one-standard-deviation rise in externalizing behavior predicts a 2.5% rise in hourly wages, but it is not significantly related to weekly hours worked. For females, a one-standard-deviation rise in externalizing predicts a 2.5% rise in hourly wage. In addition, it is associated with a 6.9% increase in hours worked per week. The positive wage returns demonstrate that externalizing behavior is productive in the labor market even though it is counter-productive in school, which is a novel finding in the literature on the economic consequences of childhood misbehavior.

In contrast, internalizing behavior is negatively related to both productivity in the labor market and hours worked. For males, a one-standard-deviation rise in internalizing predicts a 4% decrease in hourly wage. We also find that cognition increases hourly wages, but does not influence the hours decision. The remaining parameters follow conventional wisdom. For example, higher educational attainment increases worker productivity, but has little effect on the number of hours worked for those already employed. Also, individuals living in or around London and who work in more skilled occupations receive higher hourly wages. Finally, males in higher-skilled occupations do not necessarily work more hours but females do.

[30] Selection into employment is discussed in the following section.

One possible explanation of the externalizing premium is that externalizing behavior is highly negatively correlated with agreeableness (Ehrler, Evans, and McGhee, 1999). Agreeableness is one of the "Big-5" personality traits and it predicts lower earnings (Judge, Livingston, and Hurst, 2012). To explain why, Barry and Friedman (1998) show that individuals with higher levels of agreeableness are worse negotiators as they are susceptible to being anchored by early offers in the negotiation process. Relatedly, Spurk and Abele (2011) show that less agreeable individuals are more competitive in the workplace and place a higher emphasis on career advancement. They also find that agreeableness is negatively related to work hours, which is consistent with the positive relationship between externalizing behavior and hours worked for the females in our sample. In summary, high-externalizing individuals may earn more for some of the same reasons that agreeable people earn less. Our findings on externalizing differ from those on agreeableness, however, since agreeableness is generally measured during adulthood and, in contrast, we measure externalizing behavior among schoolchildren. We can therefore demonstrate that externalizing behavior, though productive on the labor market, is also counterproductive in school.

Our findings demonstrate a more nuanced relationship between childhood misbehavior and labor market outcomes than has been recognized in previous literature. Separating aggregate misbehavior into two separate components leads to a new understanding of how childhood misbehavior affects earnings during adulthood. In the following section, we conduct sensitivity and subgroup analyses to gain further insights into the relationship between externalizing behavior and adult outcomes. Thereafter, we discuss some policy implications of our findings.

2.5 Sensitivity Tests and Subgroup Analyses

Here, we assess whether our findings on the mixed effects of externalizing behavior are the result of selection into employment, occupation or fertility (Section 2.5.1). Although we do find that externalizing behavior affects these outcomes, accounting for these relationships does not undermine our main findings that externalizing has positive returns in the labor market. Next, we present results showing that the benefits to externalizing do not extend to children who experienced poverty during childhood, even when we control for additional variables such as police involvement (Section 2.5.2).

2.5.1 Externalizing and Selection

In this section, we conduct a series of sensitivity analyses to explore whether wage returns to externalizing are explained by selection. We begin with selection into employment. Next, we study how the relationship between externalizing and earnings changes when we control for various lifecycle outcomes, including education, fertility, partnership, experience by age 33 and occupation decisions. In general, we find evidence that externalizing behavior is strongly related to a host of lifecycle outcomes. However, accounting for these relationships does not undermine the idea that externalizing is rewarded in the labor market.

Externalizing and Employment

Recall that wage and hours regressions are estimated on individuals who are employed. One possible concern is that the estimated relationship between externalizing and earnings is driven solely by selection into employment. In order to consider this relationship we first estimate a multinomial logit model of selection into self and paid employment while keeping the factor analysis structure constant.[31] The results can be found in Table 2.11 where unemployed individuals are the base group. We find important gender differences in our results. Females with higher levels of externalizing behavior are less likely to be unemployed and are more likely to be self-employed at age 33.[32] For males, externalizing behavior is weakly negatively related to the employment decision. Moreover, women with high levels of internalizing behavior are significantly more likely to be unemployed, but for men it is not important. Cognition predicts higher employment for males, though the coefficient is only significant at the 10% level. For females, it is not significantly related to employment. The main impact of cognition on employment likely works through schooling, which we do control for and does predict employment for both genders.

The results for externalizing behavior among females are especially concerning since they raise the possibility that high-externalizing women who are relatively productive (or who work more hours when employed) tend to self-select into employment. This could be the case if high-externalizing individuals face a lower disutility of working and are therefore observed in unemployment only if they are particularly unproductive due to other (omitted) factors. To address this concern, we exploit earnings data for individuals who were not employed at

[31] In other words, we keep the measurement system mapping latent skills to observed measurements of misbehavior constant so that changes in the parameters are solely attributable to changes in the control variables and not in the measurement system.

[32] This finding is similar to the one in Levine and Rubinstein (2013) They show that teenagers who engage in risky or illicit behaviors are more likely to self-select into entrepreneurship.

age 33, but reported earnings in a previous employment spell. The idea is that labor market outcomes at other periods would provide some insight into how much unemployed individuals would have earned if they had worked at age 33 (Neal and Johnson, 1996). Using this approach, the proportion of individuals in our sample for whom we obtain a measure of earnings rises from 62% to 92% (90% for males and 93.5% for females).[33] If results are driven by highly productive, high-externalizing individuals entering employment, we would expect the estimated relationship between externalizing and earnings to fall once we include earnings information on unemployed individuals.

We re-estimate the model outlined in Section 2.3 using the larger sample that includes individuals with earnings information from other years. Estimates are presented in Table 2.12. In Column [1] we present the estimated parameters using the original measure of labor market earnings. In Column [2] we use the new measure of earnings that include individuals not working at age 33. We do not find a decrease in the estimated relationship between externalizing behavior and weekly earnings once we include earnings for unemployed males. These results provide evidence against the possibility that selection into employment explains the estimated results for the males in our sample. However, as can be seen in Column [4], we do see a decrease of about 20% in the estimated relationship for females. Therefore, our estimates reflect, in part, how high-externalizing females who are high earners for unobserved reasons select into employment. However, the bottom line is that, even after we account for this decrease, the resulting relationship between externalizing behavior and earnings

[33] This percentage is somewhat lower for males because a higher percentage of males are always classified as self-employed.

remains large and significant.[34]

Externalizing and other Lifecycle Outcomes

Next, we assess how estimated coefficients change when we vary the set of controls used to explain weekly hours and wages.[35] We consider four sets of controls (measured at age 33), which are added to the earnings equations successively. They are (i) dummies for educational attainment; (ii) number of children and a partnership indicator; (iii) months of experience; and (iv) occupation dummies. Estimation occurs for males and females separately and results are presented in Table 2.13 for males and in Table 2.14 for females. In Figure 2.4, we illustrate the changes in the estimated coefficients on externalizing as the additional controls are added. In particular, for each set of controls in the wage and hour equations, we simulate weekly earnings as we vary the externalizing factor from the lowest 5th percentile to the highest 95th percentile, keeping other latent skills and covariates at the population median.

To begin, we consider the relationship between externalizing behavior and earnings before we control for any additional outcomes. Estimates can be found in column [1] of Tables 2.13 and 2.14. Even before we control for any additional outcomes the relationship is positive for both males and females (though it is insignificant for females). This reflects results from the crude model showing

[34] As an additional robustness check, we also experimented with a formal Heckman selection model for hourly wages using partnership and number of children as exclusion restrictions. We do not present these results since they suggest a similar story to the one presented in Table 2.12 and because the exclusion restrictions are difficult to defend.

[35] For this exercise, we keep the measurement system mapping latent skills to observed measurements of misbehavior constant so that changes in the parameters are solely attributable to changes in the control variables and not in the measurement system. We also re-estimated the model allowing the factor structure to change at each different version of the model. Results do not change in any apparent way.

that the externalizing behavior leads to a net benefit in terms of earnings, i.e., that the premium does not emerge only after we have controlled for the negative impact on schooling. In Column [2], we add schooling dummy variables and re-estimate the model. The estimated relationship between externalizing and earnings increases by around 15%. This is not surprising given the externalizing penalty in school. In other words, the relationship between externalizing behavior and earnings is stronger once we control for schooling, which is negatively associated with externalizing.

Next, we control for number of children and whether the individual has a partner (Columns [3] and [4] of Tables 2.13 and 2.14, for males and females, respectively). For males, including these additional controls does not change the estimated coefficient on externalizing. In contrast, for females, controlling for fertility doubles the magnitude of the coefficient. This gender difference is also clear in Figure 2.4. In Panel (b) for females, the slope of the curve, which represents how externalizing is associated with earnings, increases noticeably once we add the number of children by age 33 as a control. To understand the gender difference in how fertility affects the externalizing earnings premium, we estimate a linear regression of the number of children by age 33 on the three factors using the previously estimated measurement system. Estimates are found in Table 2.15. Externalizing males and females are both likely to have a larger number of children by age 33, but based on the earnings equations (Tables 2.13-2.14), having more children is somewhat irrelevant to earnings for males, but is associated with a large drop in earnings for females. Recall from Figure 2.2 that female earnings are much lower for women with children in comparison to women without children. For males, there is no discernible relationship.[36]

Finally, we add months of experience and occupational choice as controls (Columns [5] and [6]). However, adding these to the model does not appreciably alter the estimated relationship between externalizing and earnings. In fact, there is little evidence that externalizing behavior drives individuals into any specific occupation once we have controlled for education. These results are found in Table 2.16 where we estimate a multinomial logit model of occupational sorting with unskilled occupations as the basis group. As can be seen in Table 2.16, externalizing is not strongly related to the occupation decision. High-externalizing males are more likely to self-select into skilled manual occupations but the parameter is only marginally significant.[37]

In summary, though externalizing behavior is related to a host of economic outcomes that also predict earnings, we have demonstrated here that the externalizing premium on the labor market is not driven by differential sorting by externalizing behavior into these outcomes. Return to Figure 2.4, which plots wages for different levels of externalizing using coefficients estimated assuming varying sets of controls. Though the slope does change, especially for females, depending on which controls are included, the externalizing wage premium is robust across specifications.

[36] For individuals from later cohorts, among whom women are more likely to purchase childcare in the market or men are more likely to take time out of the labor market to care for children, our findings on externalizing, fertility and earnings could be different.

[37] In additional analyses that are available upon request, we also show that the returns to externalizing do not differ significantly across occupations.

2.5.2 Childhood Poverty, Misbehavior and Earnings

Studying a sample of disadvantaged black children in the U.S., Heckman, Pinto, and Savelyev (2013) find that an early childhood education program increased earnings in part by reducing externalizing behavior. In contrast, we show that externalizing can be valuable in the labor market. In this section, we explore whether differences in findings are explained by differences in the socioeconomic status of the group being analyzed. One possibility is that children born into poorer families face a higher likelihood of criminality or police involvement for the same level of externalizing behavior.

We estimate a version of our econometric model with two changes. First, we include a measure of police involvement at age 16 as an additional outcome equation and as an additional explanatory variable in the schooling, wage and hours equations. Second, we estimate the model on a sub-sample of our analytic sample, which is selected to resemble the family characteristics of the sample studied in Heckman, Pinto, and Savelyev (2013). In particular, we construct a subsample of our analytic sample consisting of subjects who faced financial difficulties during childhood. Recall, this occurs if the interviewer reported that the household appeared to be experiencing poverty in 1965 or if a member of the household self-reported having financial difficulties in the 12 months prior to being observed in either 1969 or 1974.[38] We estimate the econometric model separately for the low-SES subsample and for all other subjects in our analytic

[38] An alternative would be to use family income. However, perhaps surprisingly, the NCDS does not collect information on family income or parental pay in the first three surveys. In the fourth survey, when children were 16 years old, categorical information was collected on each parent's work pay. However, this information on parental pay is missing for over 20% of our sample. Therefore, we decided to use the available information about financial difficulties instead.

sample, which we call the high-SES subsample.[39]

Summary statistics for the low-SES sub-sample are found in Table 2.17. Looking at the table, the low-SES sample completes less education and earns lower wages, though hours are similar across groups. They are somewhat less likely to be employed or report having a partner, but have more children, on average. To account for potential instability of the measurement system (the mapping from underlying factors to observed variables), we estimate the measurement system for each group separately. Estimates by SES group for schooling, hours, wages and police involvement are found in Tables 2.18-2.21.

Estimating separate models by childhood SES, we find that many patterns are similar to the main model. However, we also find some important differences by childhood SES. First, we estimate a larger penalty for externalizing behavior for educational attainment among individuals that grew up in low-SES households (-0.108 versus -0.061). This finding is broadly consistent with results in Ramey (2014), who shows that externalizing blacks in the U.S. face a higher likelihood of punishment by suspension in comparison to similarly externalizing whites. This could be because schools that serve low-SES children in the UK (or black children in the U.S.) have fewer resources to address externalizing behavior and therefore react to it through suspensions or expulsions.[40]

Perhaps most importantly, we find that the labor market returns to externalizing behavior fail to extend to the low-SES subsample. For this group, the

[39]In a separate analysis, not presented here, we separated our sample into four groups by gender and socioeconomic status. Main patterns remain largely similar. However, the standard errors for the low-SES groups, when divided by gender, were too large for any useful inference to be made.

[40]There are also some differences in the returns to family characteristics, such as the father's occupation.

point estimate of the coefficient on externalizing behavior is zero in the wage equation. In the hours equation, the coefficient is 0.23 and insignificant for the low-SES group (versus 0.43 and significant at the 1-percent level for the high-SES group). Wage returns to the other skills are similar across the two groups, as are the returns to education, experience and occupation. On the other hand, there are some differences in the influence of internalizing behavior and cognition for the hours worked decision. Internalizing behavior decreases hours worked for the high-SES group but not for the low-SES group and cognition increases hours worked for the low-SES group only. Other coefficients are mostly similar. However, one important caveat to the results presented in this section is that we cannot statistically differentiate the returns to externalizing behavior for the two socioeconomic groups because the standard errors in the estimates for the financially difficulties group are too large.[41]

Following the results in Heckman, Pinto, and Savelyev (2013), one explanation for possible differences in results by childhood SES status is that low-SES individuals are at a higher risk of criminal behavior for a given level of externalizing behavior. In line with this possibility, we find a strong relationship between externalizing behavior and police involvement (see Table 2.21). In general, our estimates suggest that low-SES individuals are more likely to have some police involvement (the estimated constant in the police involvement equation is -1.00 for the high-SES group and -0.41 for the low-SES group). However, the relationship between externalizing behavior and police involvement is stronger for

[41]One possibility is that differences in returns are due to instability of the measurement system across groups. However, we estimated the measurement system separately for each group and find that the factor loadings are remarkably similar for the high-SES and low-SES group (Table B.4).

the high-SES group.[42] Interestingly, we do not find much evidence that police involvement is related to worse labor market outcomes for either SES subgroup. Therefore, even though externalizing behavior predicts higher police involvement, police involvement does not appear to derail labor market prospects in the British sample we study. It is possible that the returns to externalizing behavior might be negative in a context where police involvement is highly penalized in the labor market. This is the sort of context studied in Heckman, Pinto, and Savelyev (2013), who examine a sample composed mostly of at-risk black youths in the U.S. However, for our sample, police involvement cannot explain why low-SES individuals in the British sample we study receive little payoff to externalizing behavior.

Therefore, despite our initial results showing that externalizing behavior is associated with better labor market outcomes, this positive association does not extend to individuals who faced poverty during childhood. In other words, the payoffs to non-cognitive skills are context-dependent, as argued in Lundberg (2013). To explain differences in returns to skills across socioeconomic groups, we are therefore left with several distinct, but related possibilities. The first is that there are true differences in the productivity of externalizing behavior across groups. This is possible if, for example, children born into wealthier families are better able to channel aggressive tendencies into productive activities.[43] A second possibility is that high-externalizing individuals from lower classes face different selection rules than their higher-SES counterparts, but which are not observed by the econometrician. For example, managers or co-workers may view

[42] Interestingly, internalizing behavior and cognition are associated with less police involvement, though the coefficients are much larger in magnitude for high-SES individuals.

[43] See, for example, Doyle et al. (2009) on the timing of investments to decrease inequality.

high-externalizing individuals from high-SES families as ambitious leaders and be willing to hire them in high-wage positions or to promote them. In contrast, high-externalizing individuals from lower SES families may find their advancement thwarted if they are viewed as disruptive, aggressive or impolite. If so, high-externalizing individuals from low-SES families are not unproductive *per se*, but instead sort into jobs where they earn less. In both cases, childhood SES and externalizing exhibit complementarities and children from poorer families are unable to unleash the potential of an otherwise lucrative skill.

2.6 Externalizing and Education Policy

In this section, we discuss the economic implications of our findings on externalizing behavior. Connecting childhood misbehavior to earnings connects our findings to a well-developed literature linking childhood characteristics and behaviors to long-term economic outcomes. An implication of this literature is that human capital investments during childhood can have large payoffs in adulthood (Heckman and Masterov, 2007; Doyle et al., 2009; Cunha, Heckman, and Schennach, 2010; Carneiro, Løken, and Salvanes, 2011). For example, Currie (2001, 2009) shows that early childhood health disparities can affect future labor market outcomes through a variety of mechanisms, including performance at school. This suggests that interventions that address health disparities can improve the labor market performance of children born into poverty.

Researchers have also linked childhood misbehavior to labor market outcomes, typically seeing misbehavior as uni-dimensional and unproductive. Our departure from the typical view suggests the need for caution in implementing policies that affect childhood non-cognitive skills. The concern is not a hypothetical one as many school systems are poised to enact policies that evaluate schools on character skills development (West et al., 2015). In response to such proposals, Duckworth and Yeager (2015) emphasize concerns related to measurement, arguing that assessments of non-cognitive skills could be misleading and are subject to strategic manipulation or outright cheating.[44] These concerns are certainly valid, but our findings on mixed effects of externalizing behavior raise additional serious doubts about the utility of uniformly penalizing or rewarding schools for the development of students' non-cognitive skills. The reason is that such policies could stifle skills that are productive in other sectors.

Even if a character skill is shown to be valuable in the labor market, however, this does not imply that it should be promoted in school. In our case, we do not think that the externalizing premium we identify should be used to justify policies encouraging externalizing behavior. Here, we discuss two reasons why. The first reason is tied to our results on differences in returns by socioeconomic group. Leaving aside important questions about the source of such differences in returns (e.g., a higher propensity for police involvement due to biases in criminal justice systems), policies that promote skills could harm some groups. This is especially concerning for skills that are negatively priced among individuals from poor backgrounds as such policies could exacerbate existing inequality. Another reason not to support policies promoting externalizing behavior is the possibility of negative spillover effects in the classroom if externalizing children

[44] See also Ivcevic and Brackett (2014) on issues with the measurement of grit.

are disruptive and limit other students' learning (Henneberger, Coffman, and Gest, 2016). Recall, our results show that externalizing behavior loads heavily onto the maladjustment syndrome "hostility towards children". Given documented negative impacts of bullying on education, policies increasing hostility among schoolchildren are likely to be unproductive (Brown and Taylor, 2008; Carrell and Hoekstra, 2010).

However, our results do suggest that it could be useful to explore policies and interventions that accommodate externalizing behavior rather than penalizing or simply attempting to eliminate it. Such alternatives might increase schooling without stifling valuable labor market skills. In making this distinction between potential policies, we draw on pedagogical research that discusses "control-oriented" teaching methods, which are designed to reduce externalizing behavior versus "relationship-oriented" methods, which are designed to strengthen the learning environment for externalizing children.[45] A simple example illustrates the difference in the two approaches. Young students who often initiate conversations with teachers at inopportune times could be punished for interrupting a lesson. Instead, they could be given a "raincheck" and invited to initiate a discussion at an appropriate time. The effectiveness of such practices is demonstrated by a randomized controlled trial of the My Teaching Partner-Secondary program (MTP-S), in which a web-mediated program on improving teacher-student in-class interaction has produced reliable gains in student achievement (Allen et al. (2011)).

[45] For an overview of pedagogical techniques that foster a caring and positive student-teacher relationship, in particular, in dealing with student misbehavior, see Hamre and Pianta (2006).

2.7 Conclusion

Few would argue against the idea that stronger cognition or better health would improve outcomes on almost any conceivable economic dimension. Some non-cognitive skills, such as grit, also appear to have positive returns in many sectors. In this paper, we illustrate that it is generally not meaningful to think of non-cognitive skill as either good or bad *per se*. We have demonstrated that the same non-cognitive skill can be productive in one context and counterproductive in another. Our findings suggest that investments in human capital should be evaluated in light of this possibility. In particular, mixed effects of externalizing behavior suggest caution in devising policies that target children with apparently undesirable behaviors or so-called negative non cognitive skills. Such policies may pay off in the short-run by improving educational outcomes, but may also be costly in the long-run by stifling a productive labor market skill. We also show important differences across socioeconomic groups in the returns to skills. This further complicates policies centering around non-cognitive skill formation, suggesting that individuals from disadvantaged backgrounds may suffer from an inability to profit from productive skills. Our results are particularly salient given recent efforts to include measures of non-cognitive skills as part of schools' and teachers' performance ratings.

One direction for future research would aim to better understand heterogeneity in the effects of non-cognitive skills across groups. For example, Ramey (2014) studies a cohort of individuals born in the U.S. in the 1980s and 1990s. He shows that high-externalizing blacks are more likely to be suspended from school than equally externalizing whites. This could lead to differences in the returns to externalizing behavior across racial groups since suspensions are associated with low schooling attainment and lower earnings. Extending the findings in Ramey (2014) to consider labor market outcomes could help to clarify whether differences in returns to the same non-cognitive skill help to explain stubbornly persistent inequality across racial groups. Finally, studying externalizing behavior among a relatively young cohort in the U.S. could lead to a better understanding of mixed returns to the skills that constitute human capital.

2.8 Tables and Figures

Table 2.1: LATENT FACTORS AND THEIR MEASUREMENTS

Latent Skill	Measures
Externalizing Behavior	◇ Hostility Towards Adults ◇ Hostility Towards Children ◇ Anxiety for Acceptance by Adults ◇ Anxiety for Acceptance by Children ◇ Restlessness ◇ Inconsequential Behavior ◇ Writing Off of Adults and Adult Standards
Internalizing Behavior	◇ Depression ◇ Withdrawal ◇ Unforthcomingness ◇ Writing Off of Adults and Adult Standards
Cognition	◇ Reading Comprehension Test Score ◇ Mathematics Test Score ◇ Non Verbal Score on General Ability Test ◇ Verbal Score on General Ability Test

Notes: This table lists the three latent factors used in the empirical analysis (externalizing behavior, internalizing behavior and cognition) and the observed variables used to identify them. Measures for externalizing and internalizing behaviors are drawn from the BSAG maladjustment variables, derived from teachers' reports of misbehavior. For cognition, a series of aptitude test scores are used as measures. See B.1 for further details.

Table 2.2: Summary Statistics

	Both	Males	Females	Diff
No Formal Education	0.112	0.102	0.121	*
	(0.315)	(0.303)	(0.326)	
CSE	0.128	0.112	0.142	***
	(0.334)	(0.316)	(0.349)	
O Level	0.347	0.307	0.386	***
	(0.476)	(0.461)	(0.487)	
A Level	0.146	0.190	0.103	***
	(0.353)	(0.393)	(0.305)	
Higher Education	0.146	0.150	0.142	
	(0.353)	(0.357)	(0.349)	
Higher Degree	0.121	0.138	0.105	***
	(0.326)	(0.345)	(0.307)	
Hourly Wage	6.637	7.643	5.457	***
	(3.053)	(2.967)	(2.712)	
Weekly Hours Worked	36.35	43.54	27.93	***
	(12.65)	(7.757)	(12.07)	
Weekly Earnings	252.5	329.2	162.4	***
	(152.5)	(134.5)	(119.5)	
Experience	145.9	164.0	128.2	***
	(50.92)	(45.65)	(49.56)	
In Paid Work	0.804	0.919	0.692	***
	(0.397)	(0.273)	(0.462)	
Self Employed	0.162	0.197	0.115	***
	(0.368)	(0.398)	(0.319)	
Has a Partner	0.873	0.877	0.869	
	(0.333)	(0.328)	(0.338)	
Number of Children	1.475	1.349	1.597	***
	(1.125)	(1.152)	(1.085)	
London	0.300	0.293	0.306	
	(0.458)	(0.455)	(0.461)	
Observations	7324	3612	3712	7324

Notes: Summary statistics for the analytic sample of 7,324 individuals (Column [1]) and then separately for males (Column [2]) and for females (Column [3]). For education categories, employment and partnership, entries are in the form of percentages divided by 100. Experience is measured in months and wages and weekly earnings are in 1992 British pounds. Self Employed means the percentage of individuals in paid work who are also self-employed. In Column [4], *, ** and *** mean that differences between males and females are significant at the 10, 5 and 1 percent levels, respectively.

Table 2.3: Summary Statistics - BSAG Variables

	Both	Males	Females	Diff
Hostility Towards Adults	0.766	0.896	0.639	***
	(1.754)	(1.866)	(1.628)	
Hostility Towards Children	0.240	0.266	0.215	**
	(0.719)	(0.777)	(0.656)	
Anxiety for Acceptance by Adults	0.515	0.481	0.548	*
	(1.154)	(1.094)	(1.210)	
Anxiety for Acceptance by Children	0.298	0.403	0.197	***
	(0.761)	(0.899)	(0.579)	
Restlessness	0.195	0.242	0.149	***
	(0.522)	(0.575)	(0.459)	
Inconsequential Behavior	1.263	1.676	0.862	***
	(1.868)	(2.153)	(1.432)	
Depression	0.933	1.086	0.784	***
	(1.452)	(1.534)	(1.350)	
Withdrawal	0.307	0.374	0.242	***
	(0.770)	(0.876)	(0.645)	
Unforthcomingness	1.479	1.538	1.421	*
	(2.036)	(2.009)	(2.060)	
Writing Off of Adults and Adult Standards	0.910	1.128	0.698	***
	(1.588)	(1.788)	(1.333)	
Observations	7324	3612	3712	7324

Notes: Summary statistics for maladjustment syndrome scores for our sample of 7324 individuals. Measures constructed using teachers' reports of misbehavior or misconduct in school. Statistics are reported separately for all individuals (Column [1]), for males (Column [2]) and for females (Column [3]). For each maladjustment syndrome, a child receives a score, which is an integer between 0 and 15, with 15 indicating a persistent display of behavior described by the maladjustment syndrome. In the table, entries are averages for each syndrome for the analytic sample. In Column [4], *, ** and *** mean that differences between males and females are significant at the 10, 5 and 1 percent levels, respectively.

Table 2.4: Crude Model: Educational Attainment

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Misbehavior	-.364***	-.138***	.	.	.	.	.
Externalizing	.	.	-.231***	-.092***	-.097***	-.121***	-.062**
Internalizing	.	.	-.183***	-.066***	-.060***	-.036*	-.090***
Cognition	.	.806***	.	.807***	.718***	.695***	.753***
Father Edu	.	.	.	.	.254***	.189***	.316***
Mother Edu.	.	.	.	.	.268***	.224***	.311***
No Info on Father Figure	.	.	.	.	.173**	.130	.220**
Father in Skilled Occupation	.	.	.	.	.167***	.203***	.133***
Father in Managerial Occupation	.	.	.	.	.414***	.462***	.369***
Working Mother	.	.	.	.	.019	.001	.037
Female	-.303***	-.333***	-.304***	-.334***	-.335***	.	.
Obs.	7324	7324	7324	7324	7324	3612	3712

Notes: This table contains parameter estimates from ordered probit used to link non-cognitive skills to educational attainment. We estimate the ordered probability of choosing 1 of 6 schooling levels on a set of observable variables along with proxies for unobserved skills. To construct proxies for unobserved skills, we sum up all variables used to measure that skill in subsequent analysis and then normalize each unobserved skill. Models [1]-[5] include all individuals and a gender dummy, Model [6] includes only males and Model [7] only females. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.5: Crude Model: Log Weekly Earnings

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Misbehavior	-.105***	-.028***	.	.	.	.	.	.
Externalizing	.	.	-.019*	.025**	.034***	.046***	.018**	.090***
Internalizing	.	.	-.101***	-.059***	-.050***	-.045***	-.045***	-.041**
Cognition	.	.205***	.	.204***	.084***	.053***	.052***	.057***
CSE	.	.	.	.	.078**	-.015	.013	-.021
O Level	.	.	.	.	.207***	.033	.070**	.0003
A Level	.	.	.	.	.343***	.116***	.096***	.132**
Higher Education	.	.	.	.	.515***	.178***	.156***	.154***
Higher Degree	.	.	.	.	.644***	.386***	.291***	.368***
Has a Partner	.	.	.	.	.	.085***	.121***	.031
Number of Children	.	.	.	.	.	-.106***	.015**	-.258***
Experience	.	.	.	.	.	.003***	.001***	.002***
Skilled Manual Occu.	.	.	.	.	.	.259***	.091***	.310***
Skilled Non-manual Occu.	.	.	.	.	.	.241***	.172***	.301***
Managerial Occupation	.	.	.	.	.	.514***	.266***	.695***
Female	-.933***	-.916***	-.932***	-.915***	-.867***	-.739***	.	.
London	.249***	.219***	.248***	.219***	.205***	.161***	.193***	.124***
Const.	5.639***	5.615***	5.639***	5.616***	5.319***	4.808***	5.026***	4.403***
Obs.	4936	4936	4936	4936	4936	4936	2664	2272

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to earnings. We regress log earnings of workers on a set of observable variables along with proxies for unobserved skills. To construct proxies for unobserved skills, we sum up all variables used to measure that skill in subsequent analysis and then normalize each unobserved skill. Models [1]-[6] include all individuals and a gender dummy, Model [7] includes only males and Model [8] only females. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.6: Measurement Error Model: Latent Factor Correlation Matrix

	Males		
	Externalizing	Internalizing	Cognition
Externalizing	1.000	0.575	-0.380
Internalizing	0.575	1.000	-0.358
Cognition	-0.380	-0.358	1.000
	Females		
	Externalizing	Internalizing	Cognition
Externalizing	1.000	0.593	-0.400
Internalizing	0.593	1.000	-0.398
Cognition	-0.400	-0.398	1.000

Notes: This table lists the correlation matrix of the three latent skills by gender.

Table 2.7: Measurement Error Model: Factor Loadings

Latent Skill	Measures	Males	Females
Externalizing Behavior	Inconsequential Behavior	1.000	1.000
	Hostility Towards Adults	1.680	1.312
	Hostility Towards Children	2.387	1.632
	Anxiety for Acceptance by Adults	1.204	0.763
	Anxiety for Acceptance by Children	1.699	1.522
	Restlessness	1.784	1.572
	Writing Off of Adults and Adult Standards	0.397	0.299
Internalizing Behavior	Withdrawal	1.000	1.000
	Depression	0.932	1.137
	Unforthcomingness	1.711	1.878
	Writing Off of Adults and Adult Standards	0.605	0.847
Cognition	Verbal Score on General Ability Test	1.000	1.000
	Reading Comprehension Test Score	0.596	0.579
	Mathematics Test Score	1.086	1.065
	Non Verbal Score on General Ability Test	0.733	0.766

Notes: This table lists the factor loadings that express the relationship between each observed measure and the underlying factor it identifies.

Table 2.8: Measurement Error Model: Ordered Probit for Educational Attainment

	[M]	[F]
Externalizing Factor	-0.119***	-0.046
Internalizing Factor	-0.019	-0.064**
Cognition	0.702***	0.725***
Mother Education	0.189***	0.327***
Father Education	0.250***	0.329***
No Father Info.	0.200*	0.271**
Father in Skilled Occupation	0.174***	0.113**
Father in Managerial Occupation	0.442***	0.331***
Working Mother	0.019	0.039
In Financial Difficulties	-0.311***	-0.303***

Notes: This table contains parameter estimates from an Ordered Probit model used to link non-cognitive skills to educational attainment. We estimate educational attainment on a set of observable variables along with the unobserved factors. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.9: Measurement Error Model: Log Hourly Wages

	[M]	[F]
Externalizing Factor	0.025**	0.025**
Internalizing Factor	-0.040***	-0.021*
Cognition	0.053***	0.048***
CSE	0.003	-0.001
O-Level	0.083***	0.035
A-Level	0.118***	0.122***
Higher Education	0.184***	0.257***
Higher Degree	0.333***	0.409***
Partner Dummy	0.109***	0.064***
Number of Children	0.011*	-0.067***
Experience	0.001***	0.001***
Skilled Manual Occu.	0.070***	0.070**
Skilled Non-manual Occu.	0.199***	0.173***
Managerial Occu.	0.255***	0.374***
London Dummy	0.180***	0.123***
In Financial Difficulties	-0.026	-0.014
Constant	1.334***	1.179***

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to hourly wages. We regress log hourly wages on a set of observable variables along with the unobserved factors. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.10: Measurement Error Model: Log Weekly Hours Worked

	[M]	[F]
Externalizing Factor	0.009	0.069***
Internalizing Factor	-0.016**	-0.037**
Cognition	0.000	0.018
CSE	0.007	-0.022
O-Level	-0.021	-0.040
A-Level	-0.034*	0.003
Higher Education	-0.031	-0.110***
Higher Degree	-0.051**	-0.047
Partner Dummy	0.012	-0.033
Number of Children	0.005	-0.190***
Experience	0.000	0.001***
Skilled Manual Occu.	0.023*	0.235***
Skilled Non-manual Occu.	-0.027*	0.127***
Managerial Occu.	0.011	0.317***
London Dummy	0.013	-0.000
In Financial Difficulties	-0.008	0.043*
Constant	3.748***	3.426***

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to hours worked. We regress log weekly hours worked on a set of observable variables along with the unobserved factors. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.11: MEASUREMENT ERROR MODEL: EMPLOYMENT DECISION

	Males		Females	
	Self-Employed	Employee	Self-Employed	Employee
Externalizing Factor	-0.055	-0.211*	0.377***	0.144**
Internalizing Factor	-0.198	-0.074	-0.307**	-0.208***
Cognition	0.154	0.246*	0.063	-0.006
CSE	0.726***	0.740***	0.422	0.182
O-Level	0.672***	0.434**	0.355	0.240*
A-Level	1.093***	1.064***	0.431	0.022
Higher Education	0.448	0.891***	0.356	0.499***
Higher Degree	0.210	0.639*	0.271	0.281
Partner Dummy	1.545***	1.566***	0.280	0.274**
Number of Children	-0.168**	-0.255***	-0.279***	-0.549***
Father in Skilled Occupation	-0.316	-0.107	-0.195	0.277***
Father in Managerial Occupation	-0.362	0.035	-0.414*	0.201*
Working Mother	-0.091	0.145	-0.108	0.238***
In Financial Difficulties	-0.366*	-0.329**	0.005	0.303***
Constant	-0.639*	0.463	-1.012***	0.713***

Notes: This table contains parameter estimates from a multinomial logit model used to link non-cognitive skills to the employment decision. We model the the employment decision as a linear function of a set of observable variables along with the unobserved skills. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. The base category is not-employed at age 33. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.12: Measurement Error Model: Log Weekly Earnings, Imputing Missing Earnings

	[Males]		[Females]	
	[1]	[2]	[3]	[4]
Externalizing Factor	0.039***	0.045**	0.084***	0.066***
Internalizing Factor	-0.061***	-0.054***	-0.049**	-0.041*
Cognition	0.052***	0.088***	0.062***	0.029
CSE	0.017	-0.019	-0.020	-0.065
O-Level	0.073***	0.002	0.002	0.012
A-Level	0.094***	0.062	0.135**	0.178***
Higher Education	0.158***	0.103**	0.158***	0.180***
Higher Degree	0.295***	0.277***	0.376***	0.453***
Partner Dummy	0.124***	0.151***	0.032	0.045
Number of Children	0.015**	0.001	-0.257***	-0.229***
Experience	0.001***	0.002***	0.002***	0.002***
Skilled Manual Occu.	0.089***	0.080**	0.306***	0.358***
Skilled Non-manual Occu.	0.174***	0.195***	0.301***	0.405***
Managerial Occu.	0.266***	0.339***	0.693***	0.774***
London Dummy	0.192***	0.201***	0.123***	0.166***
In Financial Difficulties	-0.033*	-0.029	0.029	-0.011
Constant	5.064***	5.019***	4.613***	4.453***
Obs	2264	3257	2272	3470

Notes: This table contains parameter estimates from a linear regression used to link non-cognitive skills to weekly earnings under alternative specifications. We regress log weekly earnings of workers on a set of observable variables along with the three factors. In Model [1], the dependent variable is reported gross weekly earnings for *males* that were working at age 33. In Model [2], we impute weekly earnings for *males* that were not working at age 33 using self-reported weekly earnings from previous jobs and include those observations in the regression. In Model [3], the dependent variable is reported gross weekly earnings for *females* that were working at age 33. In Model [4], we impute weekly earnings for *females* that were not working at age 33 using self-reported weekly earnings from previous jobs and include those observations in the regression. With the imputation, we manage to compute the earnings for 92% of the individuals in our sample. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.13: Measurement Error Model: Log Weekly Earnings (Males), Varying Controls

	[1]	[2]	[3]	[4]	[5]	[6]
Externalizing Factor	0.036***	0.041***	0.039***	0.037***	0.038***	0.039***
Internalizing Factor	-0.079***	-0.077***	-0.073***	-0.069***	-0.068***	-0.061***
Cognition	0.138***	0.072***	0.072***	0.070***	0.073***	0.052***
CSE	.	0.051*	0.050*	0.053*	0.037	0.017
O-Level	.	0.137***	0.139***	0.129***	0.111***	0.073***
A-Level	.	0.172***	0.176***	0.169***	0.157***	0.094***
Higher Education	.	0.286***	0.289***	0.278***	0.267***	0.158***
Higher Degree	.	0.374***	0.383***	0.368***	0.415***	0.295***
Number of Children	.	.	0.029***	0.012*	0.011	0.015**
Partner Dummy	.	.		0.158***	0.147***	0.124***
Experience	.	.	.		0.001***	0.001***
Skilled Manual Occu.	.	.	.	.	.	0.089***
Skilled Non-manual Occu.	.	.	.	.	.	0.174***
Managerial Occu.	.	.	.	.	.	0.266***
London Dummy	0.215***	.0.212***	0.214***	0.215***	0.216***	0.192***
In Financial Difficulties	-0.070***	-0.043**	-0.044**	-0.043**	-0.040**	-0.033*
Constant	5.666***	5.487***	5.447***	5.336***	5.148***	5.064***

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to weekly earnings with different sets of controls. We regress log weekly earnings of male workers on a set of observable variables along with the three factors. The goal is to undertand how the relationship between non-cognitive skills to earnings change as we change the set of additional regressors. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.14: Measurement Error Model: Log Weekly Earnings (Females), Varying Controls

	[1]	[2]	[3]	[4]	[5]	[6]
Externalizing Factor	0.036	0.043*	0.079***	0.080***	0.086***	0.084***
Internalizing Factor	-0.046	-0.029	-0.062***	-0.060***	-0.065***	-0.049**
Cognition	0.279***	0.109***	0.089***	0.087***	0.080***	0.062***
CSE	.	0.089	0.067	0.068	0.041	-0.020
O-Level	.	0.251***	0.154***	0.154***	0.119**	0.002
A-Level	.	0.509***	0.339***	0.339***	0.329***	0.135**
Higher Education	.	0.732***	0.579***	0.579***	0.541***	0.158***
Higher Degree	.	0.961***	0.727***	0.726***	0.799***	0.376***
Number of Children	.	.	-0.323***	-0.327***	-0.285***	-0.257***
Partner Dummy	.	.	.	0.064	0.048	0.032
Experience	.	.	.	.	0.002***	0.002***
Skilled Manual Occu.	.	.	.	.	.	0.306***
Skilled Non-manual Occu.	.	.	.	.	.	0.301***
Managerial Occu.	.	.	.	.	.	0.693***
London Dummy	0.218***	0.188***	0.136***	0.135***	0.135***	0.123***
In Financial Difficulties	-0.077*	-0.026	0.024	0.026	0.035	0.029
Constant	4.996***	4.634***	5.153***	5.106***	4.744***	4.613***

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to weekly earnings with different sets of controls. We regress log weekly earnings of female workers on a set of observable variables along with the three factors. The goal is to undertand how the relationship between non-cognitive skills to earnings change as we change the set of additional regressors. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.15: Measurement Error Model: Number of Children

	[Males]	[Females]
Externalizing Factor	0.070***	0.051***
Internalizing Factor	-0.089***	-0.026
Cognition	-0.014	-0.014
CSE	-0.059	-0.005
O-Level	-0.028	-0.104***
A-Level	-0.088	-0.206***
Higher Education	-0.076	-0.231***
Higher Degree	-0.229***	-0.370***
Children in HH at 11	0.037***	0.031***
In Financial Difficulties	0.039	0.008
Constant	0.222***	0.542***

Notes: This table contains parameter estimates from a regression model used to link non-cognitive skills to the number of children. We model the number of children as a linear function of a set of observable variables along with the unobserved skills. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.16: Measurement Error Model: Occupation Decision

	Males		
	Skilled Manual	Skilled Non-Manual	Managerial
Externalizing Factor	0.239*	-0.021	0.032
Internalizing Factor	-0.324**	-0.191**	-0.263***
Cognition	0.029	0.279***	0.208*
CSE	0.077	0.689***	0.805***
O-Level	0.536**	1.374***	1.526***
A-Level	1.070***	1.537***	2.470***
Higher Education	1.137***	0.840***	3.679***
Higher Degree	-0.112	1.209***	4.397***
Partner Dummy	0.129	0.487***	0.376**
Number of Children	-0.446***	-0.382***	-0.640***
Father in Skilled Occupation	-0.331	-0.220	-0.354**
Father in Managerial Occupation	-0.277	-0.413***	-0.556***
Working Mother	0.110	-0.105	0.045
In Financial Difficulties	0.243	-0.185	0.122
Constant	-0.839**	-0.169	-1.087***
	Females		
	Skilled Manual	Skilled Non-Manual	Managerial
Externalizing Factor	0.127	-0.099	-0.034
Internalizing Factor	-0.116	-0.163	-0.245**
Cognition	0.041	0.689***	0.594***
CSE	0.564***	0.992***	0.534**
O-Level	1.089***	1.630***	1.209***
A-Level	1.691***	2.350***	2.089***
Higher Education	1.228***	2.484***	2.990***
Higher Degree	0.746	2.914***	4.077***
Partner Dummy	0.459***	0.412*	0.821***
Number of Children	-0.013	-0.155**	-0.142**
Father in Skilled Occupation	0.221	-0.205	-0.583***
Father in Managerial Occupation	0.089	-0.537**	-0.979***
Working Mother	0.126	0.155	-0.058
In Financial Difficulties	-0.390***	-0.482**	-0.323**
Constant	-0.533**	-2.000***	-0.850***

Notes: This table contains parameter estimates from a multinomial logit model used to link non-cognitive skills to the occupation decision. We model the occupation decision as a function of a set of observable variables along with the unobserved skills. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. The base category are unskilled occupations. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.17: Summary Statistics, Subsamples by SES

	Both	High-SES	low-SES	Diff
No Formal Education	0.112	0.0839	0.257	***
	(0.315)	(0.277)	(0.437)	
CSE	0.128	0.116	0.191	***
	(0.334)	(0.320)	(0.393)	
O Level	0.346	0.351	0.324	
	(0.476)	(0.477)	(0.468)	
A Level	0.146	0.158	0.0871	***
	(0.354)	(0.365)	(0.282)	
Higher Education	0.146	0.156	0.0982	***
	(0.354)	(0.362)	(0.298)	
Higher Degree	0.122	0.137	0.0427	***
	(0.327)	(0.344)	(0.202)	
Hourly Wage	6.635	6.832	5.595	***
	(3.052)	(3.073)	(2.718)	
Weekly Hours Worked	36.35	36.57	35.18	**
	(12.65)	(12.51)	(13.32)	
Weekly Earnings	252.3	260.6	208.8	***
	(152.4)	(153.6)	(137.8)	
Experience	145.8	146.8	140.5	***
	(50.96)	(49.82)	(56.28)	
In Paid Work	0.804	0.808	0.783	*
	(0.397)	(0.394)	(0.412)	
Self Employed	0.161	0.164	0.146	
	(0.367)	(0.370)	(0.353)	
Has a Partner	0.873	0.879	0.839	***
	(0.333)	(0.326)	(0.367)	
Number of Children	1.474	1.444	1.635	***
	(1.125)	(1.121)	(1.130)	
London	0.299	0.309	0.247	***
	(0.458)	(0.462)	(0.431)	
Observations	7296	6125	1171	7296

Notes: Summary statistics for the analytic sample of 7,296 individuals. Statistics are reported separately for all individuals (Column [1]), for individual that did not experience financial difficulties growing up (Column [2]) and for those that did (Column [3]). For education categories, employment and partnership, entries are in the form of percentages divided by 100. Experience is measured in months and wages and weekly earnings are in 1992 British pounds. The Self Employed row reports the percentage of individuals in paid work that are self-employed. In Column [4], *, ** and *** mean that differences between males and females are significant at the 10, 5 and 1 percent levels, respectively.

Table 2.18: Measurement Error Model: Ordered Probit for Educational Attainment, by SES

	[High SES]	[Low SES]
Externalizing Factor	-0.061**	-0.108*
Internalizing Factor	-0.053**	-0.032
Cognition	0.698***	0.629***
Mother Education	0.246***	0.357***
Father Education	0.297***	0.185*
No Father Info.	0.259***	0.201
Father in Skilled Occupation	0.162***	0.076
Father in Managerial Occupation	0.390***	0.282*
Working Mother	-0.002	0.069
Police Involvement	-0.416***	-0.559***
No Police Inv. Info	-0.378***	-0.406***

Notes: This table contains parameter estimates from the Ordered Probit model used to link non-cognitive skills to educational attainment with the additional control "police involvement at age 16". We estimate educational attainment on a set of observable variables along with the unobserved factors. The estimation is done separately for individuals having low-SES family backgrounds and those having high-SES family backgrounds. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.19: Measurement Error Model: Log Hourly Wages, by SES

	[High SES]	[Low SES]
Externalizing Factor	0.034***	0.009
Internalizing Factor	-0.039***	-0.038*
Cognition	0.047***	0.039**
CSE	0.012	-0.009
O-Level	0.074***	0.061
A-Level	0.153***	0.039
Higher Education	0.235***	0.258***
Higher Degree	0.397***	0.428***
Partner Dummy	0.086***	0.101**
Number of Children	-0.022***	-0.025*
Experience	0.002***	0.001***
Skilled Manual Occu.	0.108***	0.086**
Skilled Non-manual Occu.	0.182***	0.115***
Managerial Occu.	0.345***	0.215***
London Dummy	0.147***	0.189***
Police Involvement	0.019	0.005
No Police Inv. Info	-0.017	0.010
Constant	1.276***	1.340***

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to hourly wages with the additional control "police involvement at age 16". We regress log hourly wages on a set of observable variables along with the unobserved factors. The estimation is done separately for individuals having low-SES family backgrounds and those having high-SES family backgrounds. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 2.20: Measurement Error Model: Log Weekly Hours Worked, by SES

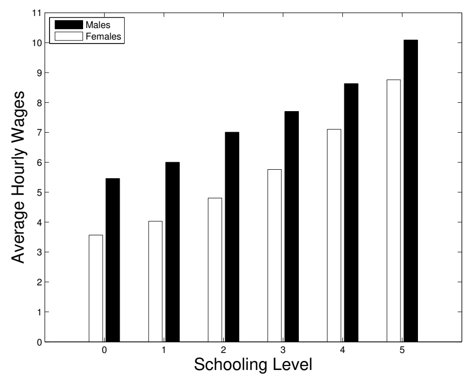
	[High SES]	[Low SES]
Externalizing Factor	0.043***	0.023
Internalizing Factor	-0.036***	-0.001
Cognition	0.002	0.036*
CSE	-0.021	-0.004
O-Level	-0.039*	-0.010
A-Level	-0.030	0.008
Higher Education	-0.069***	0.047
Higher Degree	-0.018	0.089
Partner Dummy	-0.003	-0.008
Number of Children	-0.081***	-0.087***
Experience	0.001***	0.001***
Skilled Manual Occu.	0.168***	0.109**
Skilled Non-manual Occu.	0.097***	-0.034
Managerial Occu.	0.216***	0.090*
London Dummy	0.004	0.032
Police Involvement	0.068**	0.031
No Police Inv. Info	-0.057	-0.008
Constant	3.603***	3.686***

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to hours worked with the additional control "police involvement at age 16". We regress log weekly hours worked on a set of observable variables along with the unobserved factors. The estimation is done separately for individuals having low-SES family backgrounds and those having high-SES family backgrounds. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

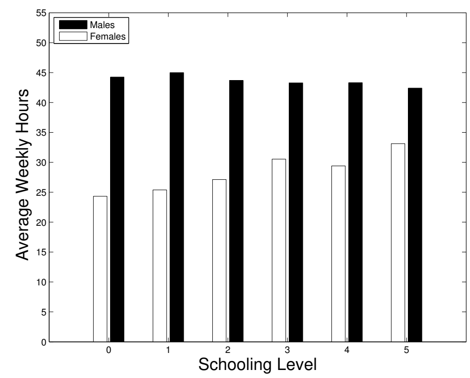
Table 2.21: Measurement Error Model: Linear Probability Model - Police Involvement at 16

	[High SES]	[Low SES]
Externalizing Factor	0.351***	0.244**
Internalizing Factor	-0.174***	-0.059
Cognition	-0.242***	-0.123
Mother Education	0.011	-0.521**
Father Education	-0.042	-0.199
No Father Info.	0.431*	0.349
Father in Skilled Occupation	-0.210***	-0.222*
Father in Managerial Occupation	-0.358***	-1.059*
Working Mother	0.041	0.045
Constant	-0.994***	-0.414**

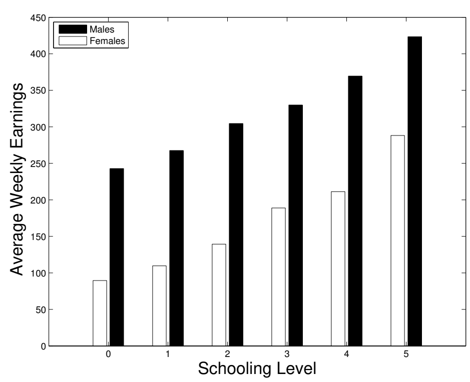
Notes: This table contains parameter estimates from a linear probability model used to link non-cognitive skills to "police involvement at age 16". We regress police involvement on a set of observable variables along with the unobserved factors. The coefficients on the three factors have been standardized to represent a 1 standard deviation effect. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.



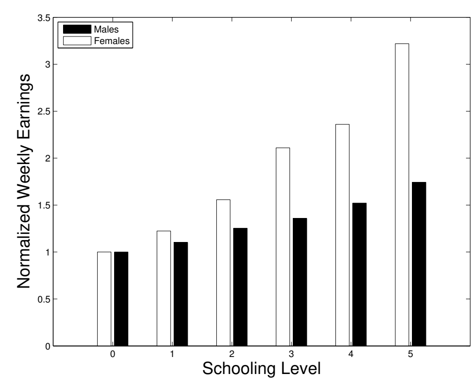
(a) Wages by schooling



(b) Hours by schooling



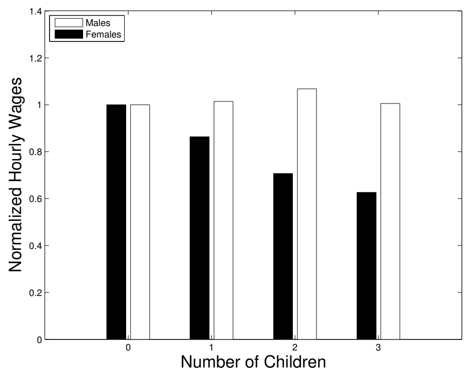
(c) Earnings by schooling



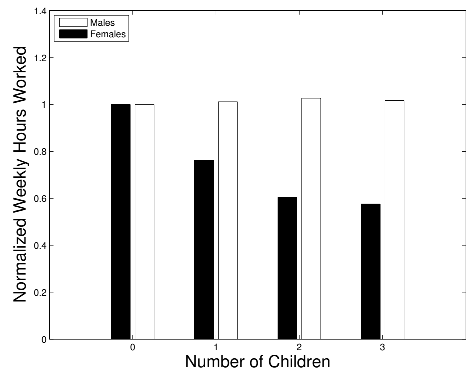
(d) Normalized earnings by schooling

Figure 2.1: Gender Differences in Labor Market Outcomes by Schooling

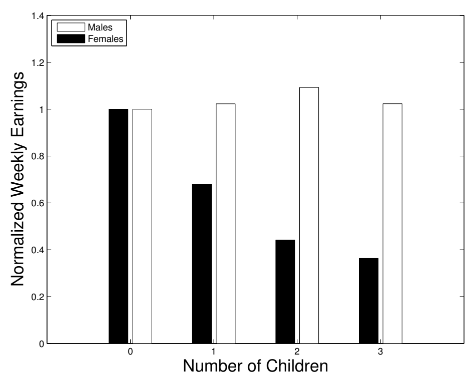
Figure 2.1(a) compares hourly wages by schooling level and gender, Figure 2.1(b) compares weekly hours worked by schooling level and gender, and Figures 2.1(c) and 2.1(d) compare weekly earnings and normalized weekly earnings by schooling level and gender.



(a) Normalized wages by fertility



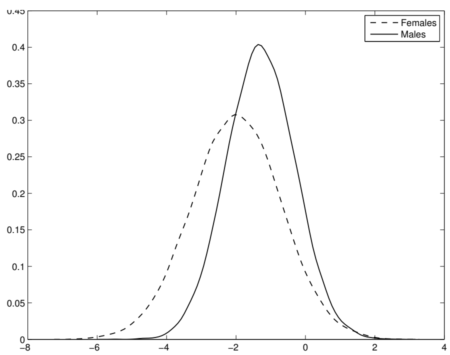
(b) Normalized hours by fertility



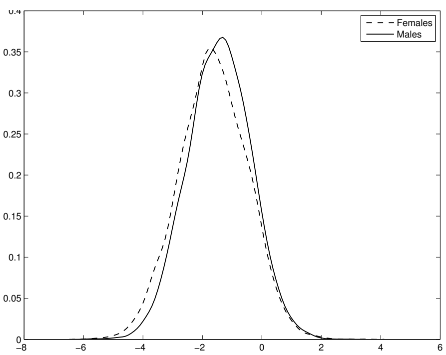
(c) Normalized earnings by fertility

Figure 2.2: Gender Differences in Labor Market Outcomes by Fertility

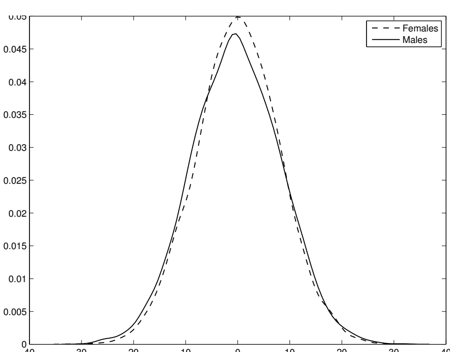
Figure 2.2(a) compares hourly wages by number of children and gender, Figure 2.2(b) compares weekly hours worked by number of children and gender, and Figure 2.2(c) compares normalized weekly earnings by number of children and gender.



(a) Estimated Distribution Externalizing



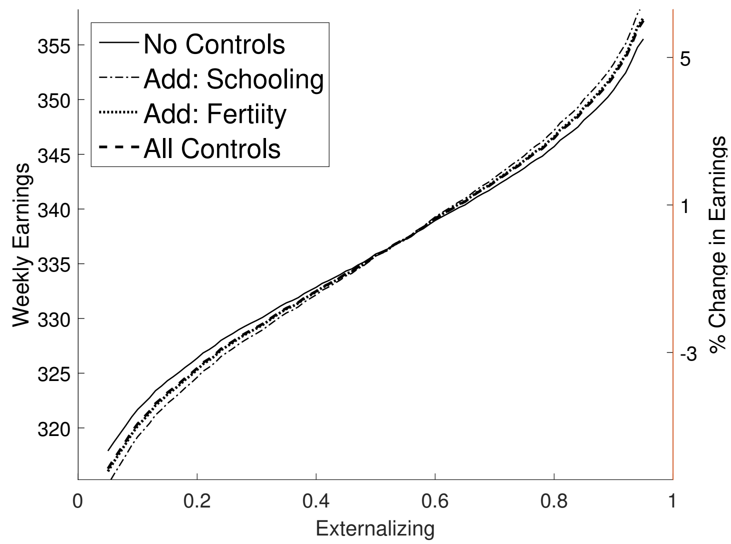
(b) Estimated Distribution for Internalizing



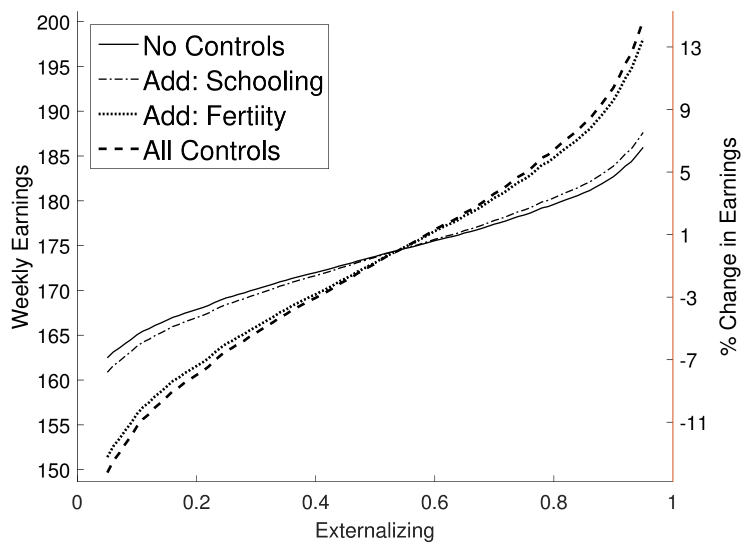
(c) Estimated Distribution Cognition

Figure 2.3: GENDER DIFFERENCES IN LATENT FACTORS

Figure 2.3(a) shows the estimated distribution of externalizing behavior by gender. Figure 2.3(b) shows the estimated distribution of internalizing behavior by gender. Figure 2.3(c) shows the estimated distribution of cognition by gender.



(a) Males



(b) Females

Figure 2.4: Decomposition of Effects of Externalizing on Weekly Earnings

Figure 2.4 visualizes the results from regressing weekly earnings on a varying set of controls presented in Tables 2.13 and 2.14. It illustrates how the predicted weekly earnings in regression models with different sets of controls vary, when we increase the externalizing from the lowest 5th percentile to the highest 95th percentile, keeping other latent skills and covariates at the population median.

Chapter 3

When Mothers and Teachers Disagree: Observer Reports and Children's Noncognitive Skills

3.1 Introduction

A large literature in Economics has recognized the importance of noncognitive skills developed in childhood in explaining economic outcomes. Studies have shown that noncognitive skills, which are defined by Borghans et al. (2008) as personality traits or the "patterns of thought, feelings, and behavior", predict a variety of adult outcomes, from educational attainment to labor force productivity and health outcomes (Heckman, Stixrud, and Urzua, 2006; Currie and Almond, 2011; Durlak et al., 2011; Moffitt et al., 2011). Measuring noncognitive skills in children, however, present unique methodological challenges. Since these skills are not directly observed, researchers have to rely on observer reports, usually the mother but sometimes the child's teacher, on the child's behaviors.[1] [2]

Observer reports, however, provide an imperfect picture of the child's emotional development. Mothers and teachers only observe children in specific settings – at home or at the school. As a result, these reports do not paint a complete picture of the child's behavior (Achenbach, McConaughy, and Howell, 1987; De Los Reyes and Kazdin, 2005; Youngstrom, Loeber, and Stouthamer-Loeber, 2000). More problematic is that observer reports can be influenced by socioeconomic contexts. For example, we now know that parental beliefs regarding children's abilities are influenced by the school composition and that teacher beliefs are influenced by children's and teachers' race (Kinsler, Pavan et al., 2016; Gershenson, Holt, and Papageorge, 2016; Papageorge, Gershenson, and Kang, 2016). Likewise, random mean-zero measurement error is likely a problem since we now know that mothers are often misinformed about different aspects of child development (Cunha, 2013; Boneva and Rauh, 2016).

In this paper, I exploit information from both maternal and teacher reports on the *same* behaviors to better understand how these reports are formed and the implications for our understanding of child development. First, I document the fact that teachers and mothers rarely agree on children's behavioral problems when asked the same questions about children's behaviors - the correlation between maternal and teacher reports is 0.37.[3] In order to better understand

[1]This is the case, for example, of the behavior problem index (BPI) collected as part of the National Longitudinal Survey of Youth - Children and Young Adults. The BPI scale is constructed from a series of maternal report about different child behaviors (Parcel and Menaghan, 1988). This scale is widely used in the literature, including, for example, the seminar paper by Cunha, Heckman, and Schennach (2010).

[2]Self-reports can also be used for older children. However, these also come with many different limitations. For example, Dunning, Heath, and Suls (2004) shows that most individuals tend to overrate their skills.

[3]This is consistent with the literature in psychiatry and psychology. In a seminal paper

this disagreement, I develop a measurement error model, where I model noncognitive skills and the disagreement between teachers and mothers as latent factors and estimate these jointly with adult outcomes influenced by the latent factors. The model allows me to separate the information in the maternal and teacher reports into three components, a common factor that is measured by both maternal and teacher reports, a factor that is uniquely reported by mothers, referred to as maternal disagreement factor, and a third that is uniquely reported by teachers, referred to as teacher disagreement factor. I find that maternal and teacher reports are measuring very different aspects of child development. I find that both observer reports measure the common factor, which captures general child misbehavior and is related to educational attainment, labor market involvement and adult risky behaviors. I also find that the teacher disagreement factor explains a small proportion of teacher reports and seems to capture random mean-zero measurement error. On the other hand, the maternal disagreement factor explains a large proportion of the variation in maternal reports and is a strong predictor of adult mental health.

These results suggest that differences between maternal and teacher reports are not solely explained by random measurement error, but seem to be driven by teachers and mothers reporting on different aspect of child development. An important open question, however, is why are mothers and teachers are responding differently when asked the same questions. One explanation is systematic bias in reporting by mothers and teachers. There is some evidence in the literature that mothers and teachers can be influenced by the socioeconomic context

Achenbach, McConaughy, and Howell (1987) found that the mean correlation between reports of similar respondents (e.g., pairs of parents) is 0.6, while between different types of respondents (e.g., a parent and a teacher) is 0.28.

when describing children's academic abilities. For example, parental beliefs have been shown to be influenced by the school composition and teacher beliefs by the child's race (Kinsler, Pavan et al., 2016; Gershenson, Holt, and Papageorge, 2016). In this paper, I find suggestive evidence that reporting bias might also be at play when describing children's noncognitive skills. For example, I find that maternal mental health is the main predictor of maternal disagreement but is not related to the common factor. This finding has been argued in the literature as evidence of bias in maternal reports (Boyle and Pickles, 1997; Najman et al., 2000).

A second possible explanation is that mothers and teachers observe children in different settings - at home and at the school - and as a result are reporting on different dimensions of noncognitive skills (Achenbach, McConaughy, and Howell, 1987; De Los Reyes and Kazdin, 2005). I also find suggestive evidence of this explanation as different factors are related to different adult outcomes. For example, the maternal disagreement is a strong predictor of adult mental health, while the common factor is a strong predictor of adult risky behaviors. Further work is needed to disentangle these two effects.[4]

My results highlight the empirical challenges in measuring noncognitive skills, as even when using the same scale different observers report different aspects of children's noncognitive development. As a result, research using the same scale but measured from different observers could lead to very different conclusions on the returns of investing in children and the importance of

[4]One important caveat is that I don't have exogenous variation for investments in children and for children noncognitive skills. As a result, further works is needed to establish whether the results shown here are causal. The reader should take this into consideration when interpreting my results.

noncognitive skills for adult outcomes. This accentuates the need for a better understanding of what aspects of children's noncognitive development are measured by different scales and observers, and how to better elicit noncognitive skills from observer reports.

The remainder of the paper is organized as follows. In Section 3.2, I describe the conceptual framework that explains mother-teacher disagreement. In Section 3.3, I introduce the dataset, describe my measure of noncognitive skills, and conduct a preliminary data analysis. In Section 3.4, I describe the econometric framework and estimation. In Section 3.5, I present my findings and discuss the implications for our understanding of children's noncognitive development. Section 3.6 concludes.

3.2 Conceptual Framework

In this section, I describe the main conceptual problems with observer reports. I start by describing conceptually how observer reports are formed and why maternal and teacher reports might be different. I argue that there are three main conceptual reasons for maternal-teacher disagreement: random and mean-zero measurement error, systematic biases in maternal and teacher reports, and behaviors that are only expressed in a specific setting and, as a result, only observed by the mother or by the teacher. Then, I discuss the conceptual implications of these three sources of disagreement for our understanding of noncognitive skill formation and for our understanding of the returns of noncognitive skills for adult outcomes.

3.2.1 Sources of Disagreement

Conceptually observer reports of children's behavioral problems measure a combination of three different objects. First, they measure noncognitive skills, here described as θ. If this were not so, the reports would be of little interest to us. Importantly, noncognitive skills are multidimensional, so θ is a vector of different skills (Currie and Almond, 2011). Second, since noncognitive skills are imperfectly observed by the observers, the reports of child behaviors are always measured with some error. Here, I describe this random and mean-zero measurement error as ϵ. Third, observers can be influenced by the socioeconomic context. For example, we know that parental beliefs about children's cognitive ability can be influenced by the composition of children in the child's school (Kinsler, Pavan et al., 2016). Similarly, we know that teacher beliefs about children's future educational attainment can be influenced by children's and teachers' race (Gershenson, Holt, and Papageorge, 2016; Papageorge, Gershenson, and Kang, 2016). As a result, reports can be systematically biased and influenced by observable socioeconomic variables, here described by the vector X.

Formally, we can write the production of maternal (MR_i) and teacher (TR_i) reports as:

$$MR_i = f_m(\theta_i, X_i^m, \epsilon_i^m) \tag{3.1}$$

$$TR_i = f_t(\theta_i, X_i^t, \epsilon_i^t) \tag{3.2}$$

where f_m and f_t are functions that describe how the mother and the teacher incorporate the three sources of information into the reports of different child behaviors.

Similarly, mother-teacher disagreement (Dis_i) can be formally described as the difference between the maternal and teacher reports:

$$Dis_i = MR_i - TR_i = f_m(\theta_i, X_i^m, \epsilon_i^m) - f_t(\theta_i, X_i^t, \epsilon_i^t) \tag{3.3}$$

This simple framework suggests three main conceptual reasons for why maternal and teacher reports can differ. First, mothers and teachers imperfectly observe children's noncognitive development. As a result, both maternal and teacher reports suffer from random measurement error – ϵ_i^m and ϵ_i^t in Equation 3.3. When taking the difference between the two reports, the random measurement error problem is amplified and can explain differences in the two reports. Random, mean-zero measurement error is not a significant problem, and can be corrected with multiple sources of information on noncognitive skills, as it has been suggested in the econometric and psychometric literature (De Los Reyes and Kazdin, 2005; Hu and Schennach, 2008).

A second reason for disagreement between different observers is systematic bias in reporting, as captured by X_i^m and X_i^t in Equation 3.3. This is a possibility if observers with different characteristics or in different social contexts disagree about the underlying noncognitive skill when they observe the same child behaviors (Achenbach, McConaughy, and Howell, 1987; De Los Reyes and Kazdin, 2005).[5] Systematically bias in reporting can be explained by both biases in beliefs, where observers mistakenly believe the child's behavior to be higher (lower) than it actually is, and differences in how individuals report the same belief about the child behavior. For example, observers in different social

[5]One such example is Gupta, Lausten, and Pozzoli (2016) that shows that mothers in distress over report child behavioral problems in comparison to fathers.

contexts might agree on the underlying child behavior but disagree on what 'often' means when asked whether "the child *often* has changes in mood or feelings".

Thirdly, disagreement might arise because noncognitive skills are multidimensional and different observes interact with children in different settings and, as a result, observe and report on a different subset of the noncognitive skills. In Equation 3.3, this is captured by f_m and f_t, which describe how different dimensions of noncognitive skills (θ_i) are transformed into maternal and teacher reports. One explanation is that mothers mainly interact and observe children at home, while teachers mainly observe children in the school, and children express different behaviors in these different settings (Achenbach, McConaughy, and Howell, 1987; Smith, 2007).[6]

3.2.2 Conceptual Implications

Understanding the sources of disagreement is important because they can clarify conceptual problems in using observer reports to study the determinants and returns of noncognitive skills. The sources of disagreements can explain why the returns of child behavioral problems for adult outcomes differ when using maternal and teacher reports, which I demonstrate Section 3.3.3. Similarly, it provides conceptual explanations for why estimates for the formation of noncognitive skills in children can differ when using maternal and teacher reports. I discuss these issues in more detail in the next two sections.

[6] Similarly, children might express a different set of behaviors as they interact with different individuals. De Los Reyes et al. (2009) showed this phenomenon in a laboratory setting. The researchers showed that maternal reports of child disruptive behavior are more predictive of laboratory observations of child disruptive behavior in the presence of the mother, whereas teacher reports are more predictive of child behavior in the presence of an unfamiliar observer.

Returns for Adult Outcomes

Suppose we are interested in estimating the returns of noncognitive skills on some relevant adult outcome (Y_i), say years of education or earnings. A typical exercise would be to linearly regress the adult outcome on the noncognitive skill and on relevant controls (X_i^y), as described in Equation 3.4 below. However, the true noncognitive skill is not observed and instead we would use maternal report and teacher report as measures of the noncognitive skill, as described in Equations 3.5 and 3.6.

$$Y_i = X_i^y \beta_y + \alpha_y \theta_i + \epsilon_y \tag{3.4}$$

$$Y_i = X_i^y \widetilde{\beta}_y + \widetilde{\alpha_y} MR_i + \widetilde{\epsilon}_y \tag{3.5}$$

$$Y_i = X_i^y \widehat{\beta}_y + \widehat{\alpha_y} TR_i + \widehat{\epsilon}_y \tag{3.6}$$

The three sources of disagreement discussed in the previous section suggest that conceptually the estimated returns of noncognitive skills can be miss-estimated in Equations 3.5 and 3.6 ($\alpha_y \neq \widetilde{\alpha_y} \neq \widehat{\alpha_y}$). First, it is likely that both maternal and teacher reports to be measured with error. This would lead to attenuation bias, or an underestimate of the returns of noncognitive skills. Second, reports might suffer from reporting bias, which can lead to bias estimates on the returns to noncognitive skills if the source of the bias also affects the adult outcome. Thirdly, it possible for mothers and teachers report on a different set of child behaviors. In this case, the problem is more nuanced since estimates for the returns of noncognitive skills in Equations 3.5 and 3.6 are measuring returns to different sets of skills and, as a result, should not be expected to be the same. Moreover, it would imply that estimates of noncognitive skills measured with teacher reports and maternal reports are not necessarily comparable and highlight the need to treat noncognitive skills as multidimensional objects.[7] As a result, understanding the sources of disagreement between mothers and teachers can highlight important problems in estimating the returns of noncognitive skills.

Noncognitive Skill Formation

Similarly, suppose we are interested in estimating the returns of different investments in children (X_i^θ). A typical exercise would be to linearly regress the noncognitive skill on the relevant investments as described in Equation 3.7 below. Since the underlying skill is unobserved we would use maternal report and teacher report as measures of the noncognitive skill as described in Equations 3.8 and 3.9.

$$\theta_i = X_i^\theta \beta_\theta + \epsilon_\theta \tag{3.7}$$

$$MR_i = X_i^\theta \widetilde{\beta}_\theta + \widetilde{\epsilon}_\theta \tag{3.8}$$

$$TR_i = X_i^\theta \widehat{\beta}_\theta + \widehat{\epsilon}_\theta \tag{3.9}$$

The sources of disagreement suggest that the estimated returns of different investments could be different in the three equations (i.e. $\beta_\theta \neq \widetilde{\beta}_\theta \neq \widehat{\beta}_\theta$). In

[7] The literature on noncognitive skills has started to realize the importance of treating noncognitive skills as multidimensional objects. One example is Papageorge, Ronda, and Zheng (2017). They show that misbehavior in school has two dimensions, one that is valued in the labor market (externalizing behaviors) and one that is not (internalizing behaviors). This difference in effects, however, would not be uncovered if misbehavior was treated as an uni-dimensional object, as it is commonly done in the literature.

this case, random measurement error is less of a problem since the variable measured with error is the outcome of interest. However, as before, bias in reporting would lead to miss-estimation of the returns of different investment on children's noncognitive skills if the source of the bias in reporting is correlated with the investments. That is, we would attribute a higher effect to an investment if it also influences the observer reports. A second problem arises if different sets of behaviors are measured by mothers and teachers. In this case, as before, the estimated returns of investments in children are expected to be different when using maternal and teacher reports, since different skills are affected differently by the investments. As before, understanding the sources of disagreement between mothers and teachers can highlight important problems when estimating the returns of investments on children's noncognitive development.

3.3 Data and Preliminary Results

In this section, I introduce the data set used in this paper and discuss the different ways we can measure noncognitive skills in children. Then, I show that mothers and teachers display a high level of disagreement about children's noncognitive skills when asked the same questions about different child behaviors. Moreover, I provide preliminary evidence that mother-teacher disagreement is not random and is correlated with family characteristics such as maternal age, maternal mental health and family income. Finally, I provide preliminary evidence that maternal and teacher reports are correlated with different adult outcomes, which suggests that they might be capturing different dimensions of children's noncognitive skills. These results motivate the empirical model developed in Section 3.4.

3.3.1 The Child Development Supplement to the PSID

In this paper, I use information from the *Panel Study of Income Dynamics* (PSID) and its *Child Development Supplement* (CDS). The PSID is an ongoing dynastic longitudinal survey. It started as a nationally representative sample of 18,000 individuals living in 5,000 families in 1968 in the United States. The CDS collected information on 3,563 children living in 2,394 of the PSID families. Information was collected in three waves: 1997, 2002 and 2007. Eligible children were between the ages of 0 and 12 in 1997, at the time of the first survey. The CDS includes a broad array of developmental measures as well as information on the home and schooling environment of the child.

In this paper, I focus on children attending elementary and middle school in 1997, during the first wave. For these children, the survey collected information on their schooling environment and interviewed their elementary and middle school teachers. This allows me to compare maternal and teacher reports about child behavioral problems. Information on teacher characteristics and experience was also collected as part of the schooling supplement. Schooling information was collected for 1,024 children out of the 3,563 children in the original CDS sample. My final sample includes 834 of the 1,024 children in elementary school in 1997. I exclude children missing information on key family characteristics, such as maternal years of education, children with missing information on key teacher characteristics and missing information on either maternal or teacher reports on their behavior.

Table 3.1 provides summary statistics for the 834 children included in the study. In 1997, these children were between 4 and 13 years old, with an average age of 9.26. Half of my sample is female and about 37% of my sample is reported by the mother as black. The high percentage of black children is a direct result of the over-sample of low income families in the main PSID. For the same reason, about 28% of mothers were single in 1997. In addition to the common family characteristics, such as family income, maternal years of education and age at birth, the PSID also provides information about mothers' mental health, measured by the Kessler 6 (K6) distress scale. The K6 is a widely used measure of general mental health developed by Kessler et al. (2002). It measures psychological distress and captures a symptoms of depression and anxiety (Drapeau, Marchand, and Beaulieu-Prévost, 2011).

Moreover, it also provides information on maternal investment measured by the emotional support scale and the cognitive stimulation scale from the Home Observation Measurement of the Environment-Short Form (HOME-SF). The HOME-SF scales use both maternal reports and interviewer observations to measure the quality of the cognitive stimulation and emotional support provided by the child's family. The emotional support scale includes interviewer observations, such as whether the"mother answered child's questions or requests verbally?", and maternal reported measures, such as "about how many times, if any, have you had to spank child in the past week?". Similarly, the cognitive stimulation scale includes interviewer observations, such as the "building has no potentially dangerous structural or health hazards within a school-aged child's range.", and maternal reported measures, such as "How many books does the child have?".

Furthermore, the PSID continued to survey these children throughout their lives. Starting in 2005, individuals 18 and older were interviewed as part of the 'transition into adulthood' (TA) supplement. As part of this study, information was collected on many adult outcomes. As a result, the TA supplement allows me to link the maternal and teacher measures of children's behavioral problems to adult outcomes. I included a varied range of adult outcomes in this study in order to better understand what the maternal and teacher reports are measuring. Information on the adult outcomes can be seen in Table 3.1. First, I include information on individuals' years of completed education. On average the individuals in my sample completed 13.26 years of education at age 23. Second, I include information on their employment, on average 82% of my sample was working or in school at age 23. Third, I include a measure of young adult risky behavior. This measure captures how often the individual engages in five different behaviors. Individuals were asked 'about how often in those 6 months did you' 'do something dangerous', 'damage public or private property', 'get into a physical fight', 'drive when drunk or high' and 'ride with a driver who had too much to drink'. Each behavior was coded as follows: never as (1), once as (2), 2-3 times as (3), 4-6 times as (4), and more than 6 times as (5). The risky behavior scale is an average of the self-reported score on each of these five questions, ranging from 1 to 5. On average individuals score 1.41 in the risky behavior scale. Finally, I also include a measure of adult mental health. This measure is the same scale available for their mothers in 1997, which measures symptoms of anxiety and depression. These individuals report slightly higher distress than the mothers , 4.88 versus 3.60 on average.

3.3.2 Measuring Noncognitive Skills - The BPI

As mentioned throughout the paper, one issue with studying children's noncognitive development is that noncognitive skills are not directly observed. As a result, researchers are dependent on information about different child behaviors that are related to the child's noncognitive skills. Information on these different behaviors is usually elicited from an adult that is close to the child, usually the child's mother but sometimes the child's teacher (Borghans et al., 2008).[8] [9] Alternatively, some researchers have relied on child outcomes that are highly correlated with noncognitive skills as measures of these skills. For example, Jackson (2012) and Gershenson (2016) rely on student absences and suspensions as measures of students' noncognitive skills.

In this paper, I use both maternal and teacher reports of child behaviors from the Behavior Problem Index (BPI). The BPI was developed to measure the incidence and severity of child behavior problems in a survey setting (Peterson and Zill, 1986). The BPI in the PSID-CDS includes 26 questions[10] administered to children ages 3 and older. These questions ask respondents about specific behaviors that children may have exhibited in the previous three months. For

[8]For adults and older children, self-reports about these behaviors can also be used to elicit individual's noncognitive skills.

[9]For example, Cunha, Heckman, and Schennach (2010) uses maternal responses about different child behaviors in the Behavior Problem Index to measure children's noncognitive skills. This information is collected as part of the widely used National Longitudinal Survey of Youth - Children and Young Adults. In comparison,Papageorge, Ronda, and Zheng (2017) uses teacher reports about different child behaviors in the Bristol Social Adjustment Guides to measure similar children's noncognitive skills. These are part of the British National Child Development Study.

[10]I use 26 of the original set of 30 questions administered in the PSID. I exclude questions that were administered differently for teachers and primary caregivers and questions that do not belong to the six sub-scales described in Table 3.2.

example, respondents were asked whether the child "is stubborn, sullen or irritable" and whether the child "has trouble getting along with other children" (see Table 3.2 for the full list of questions). Respondents describe the child's behavior using three categories: (1) "often true", (2) "sometimes true", and (3) "not true". In the PSID-CDS the BPI was administered to various respondents, including the primary caregiver, usually the mother, a secondary caregiver, usually the father, and the child's teacher in preschool, elementary and middle school. These questions vary slightly by child age and respondent but the underlying construct remains the same.

The 26 questions can be separated into six different subscales measuring different dimensions of child behaviors. Table 3.2 lists the 26 questions and the corresponding sub-scale classification. For example, mothers were asked whether their children "has sudden changes in mood or feeling", capturing anxiety and depression, and also whether their children "cheats of tells lies", capturing antisocial behavior. The sub-scales are constructed by summing the scores on the individual questions. I rely on the six sub-scales when estimating the model described in Section 3.4.

In addition to the BPI, the PSID-CDS also includes information about different child outcomes that are related to children's noncognitive development. For example, it includes information about whether the child has ever been suspended or expelled from school, ever saw a behavioral specialist, and how often she has been late or absent from school. The specific questions are described in Table 3.3. These observable outcomes can also be used to measure children's noncognitive skills and will be used to identify the model described in Section 3.4.

3.3.3 Preliminiary Results

In the main econometric analysis, I formally model noncognitive skills and the disagreement between teachers and mothers as latent variables and estimate these jointly with investments in children and adult outcomes. However, for the preliminary analysis conducted here, I construct measures of noncognitive skills by simply summing the scores in the 26 BPI questions from maternal and teacher reports. Similarly, mother-teacher disagreement is constructed by taking the difference between the maternal noncognitive score and the teacher noncognitive score. This 'crude' model allows me to demonstrate the key patterns in the data and to demonstrate that the main results are not driven by the factor analytic methods in the main econometric analysis.

Table 3.4 summarizes the BPI total score and subscales from teachers and mothers separately. These are crude measures, constructed simply by summing the score in the corresponding BPI questions. In general, mothers tend to report higher behavioral problems than teachers, the difference however is not statistically significant. Despite the non-significant difference in average reports, mothers and teachers rarely agree on children's behavioral development. The correlation between maternal and teacher reports is only 0.36. This result is in line with other results in the literature that found the average correlation between maternal and teacher reports to be 0.27 across several studies (see Achenbach, McConaughy, and Howell (1987)). The low correlation is also present in the sub-scales, with correlations ranging from 0.22 for the 'Dependent' sub-scale to 0.42 for the 'Hyperactive' sub-scale.

In order to understand what explain this high level of disagreement, I estimate the relationship between different family characteristics, maternal investments and teacher characteristics with the disagreement score (Dis_i) and the child behavior measured from maternal and teacher reports (MR_i and TR_i). Formally, I estimate linear regressions of the following form:

$$MR_i = X_i^D \beta^{mr} + \epsilon_i^{mr} \tag{3.10}$$

$$TR_i = X_i^D \beta^{tr} + \epsilon_i^{tr} \tag{3.11}$$

$$Dis_i = MR_i - TR_i = X_i^D \beta^D + \epsilon_i^D \tag{3.12}$$

where X_i^D is a vector of different family socioeconomic characteristics and maternal investments and ϵs are normally distributed disturbances.

Estimates of Equations 3.10-3.12 are presented in Table 3.5. I find some important differences in the association between the different observable characteristics and the maternal and teacher reports. First, teacher reports are influenced by the child's age and gender, while maternal reports are not. Second, while I find that some family investments are associated with both maternal and teacher reports, others are only associated with one of the observer reports. For example, I find that maternal education, maternal mental health and family income are associated with maternal reports but not with teacher reports. Similarly, cognitive stimulation is strongly associated with teacher reports but only weakly related to the maternal reports. Of these relationships the most striking is the difference in the association between the reports and maternal mental health. As a result, maternal mental health is the main predictor of mother-teacher disagreement. This finding is well documented in the literature

(Fergusson, Lynskey, and Horwood, 1993; Najman et al., 2000). Finally, teacher and school characteristics do not seem to influence either maternal or teacher reports.

Since mothers report higher behavioral problems on some sub-scales and teachers report higher behavioral problems on others (see Table 3.2), it is possible that these differences are driven by differences in the weights given to different sub-scales in the maternal and teacher reports. In order to understand whether this is a problem, I re-estimate Equation 3.12 on each of the six sub-scales separately. These results can be seen in Table 3.6. I don't find any discernible differences in the sub-scales that could be driving the results in Table 3.5.

These results suggest that mother-teacher disagreement cannot be solely explained by random measurement error. However, differences in association between observable characteristics and maternal and teacher reports can be explained by either bias in reporting by one of the observers or by the idea that different observers report on different dimensions of child behaviors. The results in Table 3.5 do not allow me to differentiate between the two sources of disagreement.

In order to try to distinguish between these two sources of disagreement, I explore the relationship between the maternal and teacher reports with different dimensions of adult outcomes at age 23. These outcomes capture educational attainment, involvement with the labor market, risky behaviors and mental health. The idea being that different dimensions of noncognitive skills have different returns for different adult outcomes, and if maternal and teacher reports are measuring different dimensions we should also see different returns across the outcomes. Formally, let Y_i be adult outcome of interest for individual i, I estimate linear regressions of the following form:

$$Y_i = X_i^y \beta_y + \alpha_1 \theta_i + \alpha_2 Dis_i + \epsilon_y \tag{3.13}$$

where θ_i is a measure of child i's noncognitive skills using either maternal or teacher reports, X_i^y are observable socioeconomic characteristics of the child and ϵ_y is a normally distributed disturbance.

Estimates of Equation 3.13 are presented in Table 3.7. First, note that both maternal and teacher measures of children's behavioral problem are negatively associated with educational attainment and involvement with the labor and schooling sectors at age 23 (columns [1] and [2]). The maternal report, however, is not significantly related to the risky behaviors scale at age 23, while the teacher report is. In contrast, teacher report is not significantly related with the mental health scale at age 23, while the maternal report is. Second, for the first three outcomes, the disagreement factor is significant when I control for the maternal report but not when I control for the teacher report (columns [4] and [5]). This suggests that the unique information in the teacher report is an important predictor of educational attainment, labor market and schooling participation and adult risky behaviors. In contrast, for the mental health scale, the disagreement is only significant when I control for the teacher report. This suggests that the unique information in the maternal report is an important predictor adult mental health.

The results from the crude model presented in Tables 3.5 and 3.7, provide evidence that the mother-teacher disagreement cannot be solely explained by random measurement error. Moreover, it provides suggestive evidence that the maternal and teacher reports are picking up different dimensions of child behavioral problems. These results suggest that the teacher report is capturing a general aspect of child misbehavior that is associated with educational attainment, involvement with the labor and schooling sectors, and with adult risky behaviors. The maternal report, on the other hand, seems to be capturing information about the child's mental health since it is associated with the maternal mental health scale and maternal investments in the form of emotional support, and, importantly, with adult psychological distress.

3.4 Model

This section describes the main empirical framework used in the paper. I start by describing the structural latent factor model and its advantages in comparison to the crude model estimated in the previous section. In short, it allows me to control for measurement error in the reports and to separately identify the unique information in each observer report. Then, I quickly describe the estimation procedure. Finally, I explain the main identification assumptions and the variation that identify the key parameters in the model.

3.4.1 Description of the Model

In the model, noncognitive skills are measured by three sources of information: maternal reports of child behaviors, teacher reports of child behaviors and child outcomes, such as school suspension and absences. Importantly, I treat each sub-scale of the BPI, for each observer, as an independent measure of the child's noncognitive skills. Treating each sub-scale as a mismeasurement allows me to control for measurement error in the observer reports. More importantly, I allow for three latent factors to explain the observer reports and child outcomes in the model. The first factor, the 'common factor', is allowed to influence all noncognitive skills measurements. As a result, it captures the common variation across the maternal and teacher reports and the child outcomes. On the other hand, the second factor, the 'maternal disagreement factor', only influences maternal reports and captures the information that is unique to these reports. Similarly, the third factor, the 'teacher disagreement factor', captures the information that is unique to the teacher reports.

As explained in Section 3.2, the unique information in each observer report can be explained by three different components. First, the disagreement factors could be capturing a dimension of noncognitive skills that is only observed or reported by one of the observers. Second, the disagreement factors could be capturing observer bias in reporting. Third, it can capture measurement error that is correlated across sub-scales, and as a result not corrected by the model. Unfortunately, I cannot separately identify these three components with the available data. Nonetheless, estimated results provide suggestive evidence of which component seems to be the most relevant in explaining the unique information in each observer report.

Formally, I denote the k-th measurement of child i's noncognitive skills from observer $j \in \{m, t\}$ as m^j_{ik}, where m denotes maternal reports and t teacher reports. m^j_{ik} is specified as:

$$m^j_{ik} = \alpha^j_{0k} + \alpha^j_{1k}\theta^c_i + \alpha^j_{2k}\theta^j_i + \epsilon^j_{ik} \qquad j \in \{m, t\} \tag{3.14}$$

where θ_i^c is the 'common factor' that influences both maternal and teacher reports, θ_i^m is the 'maternal disagreement factor' that influences maternal reports only, and θ_i^t is the 'teacher disagreement factor' that influences teacher reports only. In addition, α_{0k}^j is the mean for measurement k of observer j, α_{1k}^j and α_{2k}^j are the factor loading of the two factors on the k-th measurement of observer j and ϵ_{ik}^j is an error term capturing measurement error that is assumed to be normally distributed and independent across measurements.

Moreover, I use child observed outcomes related to misbehavior, such as suspension in school and absences, to separately identify θ_i^c from θ_i^m and θ_i^t. In order to separately identify the three factors, I need a source of information that only depends on θ_i^c but not on either θ_i^m or θ_i^t. I explain the identification issue in more detail in Section 3.4.3. Formally, I denote the k-th measurement of child i's outcome as m_{ik}^o. m_{ik}^o is specified as:

$$m_{ik}^o = \alpha_{0k}^o + \alpha_{1k}^o \theta_i^c + X_i^o \zeta_k^o + \epsilon_{ik}^o \tag{3.15}$$

where, X_i^o are observables that affect the child outcomes in addition to the common factor and ζ_k^o are the estimated returns for these observables for the k-th outcome. In addition, as before, α_{0k}^o is the mean for outcome k, α_{1k}^o the factor loading of the common factor on the k-th outcome and ϵ_{ik}^o is an error term capturing measurement error that is assumed to be normally distributed and independent across outcomes.

The measurement system is estimated jointly with the distribution for the three latent factors, which is allowed to depend on family, teacher and child characteristics. That is, the latent factors are determined as follows:

$$\theta_i^c = X_i^c \beta^c + \epsilon_i^c \tag{3.16}$$

$$\theta_i^m = X_i^m \beta^m + \epsilon_i^m \tag{3.17}$$

$$\theta_i^t = X_i^t \beta^t + \epsilon_i^t \tag{3.18}$$

where the X_is are observable variables that determine the three latent factors and the ϵ_is are exogenous shocks that are normally distributed and independent across factors.[11] [12] The β coeficients estimate the association between the observable family, teacher and child characteristics and the three sources of information.

Finally, I jointly estimate this system with relevant adult outcomes that are influenced by the noncognitive skills. I approximate these outcomes with linear-in-parameter models. Formally, let y_{in} denote the n-th adult outcome for individual i, then:

$$y_{in} = \gamma_{1n}\theta_i^c + \gamma_{2n}\theta_i^m + \gamma_{3n}\theta_i^t + X_i^y \zeta_n^y + \epsilon_{in}^y \tag{3.19}$$

where γ_{1n}, γ_{2n} and γ_{3n} estimate the association between the three sources of information and the relevant adult outcomes, X_i^y include control variables and ϵ_{in}^y are exogenous shocks that are normally distributed and independent across outcomes.

I summarize the parameters to be estimated by a vector denoted Φ:

$$\Phi = (\alpha, \beta, \gamma, \zeta, \Sigma)$$

[11] This is necessary condition for identification.

[12] Also necessary for identification is for X_i^c to include variation not in X_i^o. Otherwise, I cannot separately identify ζ_k^o from β^c.

where α denotes the set of factor loadings in equations 3.14 and 3.15, β is the set coefficients governing the relationship between observables and the three latent factors in equations 3.16-3.18, γ is the set of coefficients relating the three factors to adult outcomes in equation 3.19, ζ is the set of coefficients on control variables in equations 3.15 and 3.19, and Σ are the estimated variance of the exogenous shocks in equations 3.14-3.19.

3.4.2 Estimation Procedure

I estimate the model by simulated maximum likelihood. The estimation is done in steps. First, at each parameter suggestion, indexed by g and denoted $\Phi^{(g)}$, and at for each individual i, I simulate a vector of unobserved latent factors k^3 times, using the Gauss-Hermite quadrature with k points for the exogenous shocks described in equations 3.14 and 3.15. Then, for each draw, I compute the probability of observing the maternal reports, teacher reports, child outcomes and adult outcomes.

More formally, given a parameter suggestion g, I draw $k^3 \times I$ exogenous shocks, where I is the number of individuals in my sample. Then, for each individual i and draw l, I construct a vector of latent factors ($\theta_{il}^{c,(g)}$,$\theta_{il}^{m,(g)}$,$\theta_{il}^{t,(g)}$), and compute the density functions for each outcome in the model: the probability of observing the maternal reports $f_{il}^{M^m,(g)}(m_i^m)$, the probability of observing the teacher reports $f_{il}^{M^t,(g)}(m_i^t)$, the probability of observing the child outcomes $f_{il}^{M^0,(g)}(m_i^o)$, and the probability of observing the adult outcomes $f_{il}^{Y,(g)}(y_i)$. Then, I compute each weighted individual's likelihood contribution, where the weights (w_l) are the weights in the Gauss-Hermite quadrature with k points:

$$L_i^{(g)} = \sum_{l=1}^{k^3} \left(f_{il}^{M^m,(g)}(m_i^m) \times f_{il}^{M^t,(g)}(m_i^t) \times f_{il}^{M^0,(g)}(m_i^o) \times f_{il}^{Y,(g)}(y_i) \right) w_l \tag{3.20}$$

Finally, I take the log of the individual likelihood contribution and sum over all individuals to form the simulated log-likelihood function:

$$l^{(g)} = \sum_{i=1}^{I} log \left(L_i^{(g)} \right) \tag{3.21}$$

Using both simplex and gradient methods, I evaluate $l^{(g)}$ at different values in the parameter space until a maximum is found.

3.4.3 Parameter Identification

Identification of latent factor models has been extensively discussed in the economics literature. While different models require different identification assumptions, the general requirement is for there to exist sufficient correlation between the measurements and instruments (variables that only enter the model through the latent factors) to identify the latent factors. Carneiro, Hansen, and Heckman (2003) provides a good overview of the identification assumptions necessary to identify a linear latent factor model where the latent factors are independent. They argue that a sufficient condition for identification for the number of measurements to be equal or larger than twice the number of latent factors plus one. More generally, conditions for nonparametric identification are provided in the measurement error literature (see Hu (2008); Hu and Schennach (2008); Cunha, Heckman, and Schennach (2010)).

In order to understand the variation in the data that identify the key parameters of the model, I provide an illustrative example of how the correlation between measurements and investments can be used to identify the parameters in the model. The conditions for identification in the main model are similar than those in the simple model below. That is, I need to assume that the exogenous shocks are independent ($\epsilon_i^c \perp \epsilon_i^m \perp \epsilon_i^t$), that only the common factor influences the observed child outcomes, as in Equation 3.15, and to fix the value of at least one factor loading for each of the three factors.

Simple Example

In this section, I show identification for a simple model with 5 measurements (M_1-M_5), two unobserved factors (θ_1 and θ_2) and three explanatory variables that only influence the factors (X_1,X_1 and X_3). Formally, we can describe the system as:

$$M_1 = \mu_1 + \alpha_{11}\theta_1 + \epsilon_1$$

$$M_2 = \mu_2 + \alpha_{12}\theta_1 + \epsilon_2$$

$$M_3 = \mu_3 + \alpha_{13}\theta_1 + \alpha_{23}\theta_2 + \epsilon_3$$

$$M_4 = \mu_4 + \alpha_{14}\theta_1 + \alpha_{24}\theta_2 + \epsilon_4$$

$$M_5 = \mu_5 + \alpha_{15}\theta_1 + \alpha_{25}\theta_2 + \epsilon_5$$

$$\theta_1 = \beta_{11}X_1 + \beta_{12}X_2 + \beta_{13}X_3 + \epsilon_{\theta_1}$$

$$\theta_2 = \beta_{21}X_1 + \beta_{22}X_2 + \beta_{23}X_3 + \epsilon_{\theta_2}$$

Identification requires at least to measurement to be uniquely determined by one of the factors. M_1 and M_2 satisfy this condition in the simple model and the child outcomes satisfy this condition in the main model (see Equation 3.15). Moreover, in order to scale the factors I need to fix one of the factor loadings (αs) for each of the factor. Here I will assume that $\alpha_{11} = 1$ and $\alpha_{23} = 1$. Finally, I also need to assume the error terms are independent, so that $\epsilon_{\theta_1} \perp \epsilon_{\theta_2} \perp \epsilon_1 \perp \epsilon_2 \perp \epsilon_3 \perp \epsilon_4 \perp \epsilon_5$.

We are interested in estimating three main set of parameters: the factor loadings (αs), the determinants of the latent factors (βs) and the distribution of the shocks for the latent factors (ϵ_{θ_1} and ϵ_{θ_2}). Below I describe step by step how each of these parameters can be identified in the simple model.

I start with the parameters describing the distribution of θ_1. First, β_{11}, β_{12} and β_{13} are easily identified from regressing M_1 on X_1, X_2 and X_3. Since $\alpha_{11} = 1$, the β_{11}, β_{12} and β_{13} are easily identified as can be seen in the equations below:

$$M_1 = \mu_1 + \theta_1 + \epsilon_1$$

$$= \mu_1 + \beta_{11}X_1 + \beta_{12}X_2 + \beta_{13}X_3 + (\epsilon_{\theta_1} + \epsilon_1)$$

Once we know β_{11}, β_{12} and β_{13}, we can identify α_{12} also by regressing M_2 on X_1, X_2 and X_3, as can be seen in the equations below:

$$M_2 = \mu_2 + \alpha_{12}\theta_1 + \epsilon_2$$

$$= \mu_2 + \alpha_{12}\beta_{11}X_1 + \alpha_{12}\beta_{12}X_2 + \alpha_{12}\beta_{13}X_3 + (\alpha_{12}\epsilon_{\theta_1} + \epsilon_2)$$

Finally, ϵ_{θ_1} can be identified from the correlation between the residuals of the two regression above. That is: $var(\epsilon_{\theta_1}) = corr(\epsilon_{\theta_1} + \epsilon_1, \alpha_{12}\epsilon_{\theta_1} + \epsilon_2)/\alpha_{12}$ since $\epsilon_{\theta_1} \perp \epsilon_1 \perp \epsilon_2$.

Identifying the parameters for θ_2 is slightly more complicated since we need to solve a system of equations. We can identify the parameters in two steps. First, we regress M_3, M_4 and M_5 on X_1, X_2 and X_3:

$$M_3 = \widehat{\mu}_3 + \widehat{\gamma_{31}} X_1 + \widehat{\gamma_{32}} X_2 + \widehat{\gamma_{33}} X_3 + \widehat{\epsilon_3}$$

$$M_4 = \widehat{\mu}_4 + \widehat{\gamma_{41}} X_1 + \widehat{\gamma_{42}} X_2 + \widehat{\gamma_{43}} X_3 + \widehat{\epsilon_4}$$

$$M_5 = \widehat{\mu}_5 + \widehat{\gamma_{51}} X_1 + \widehat{\gamma_{52}} X_2 + \widehat{\gamma_{53}} X_3 + \widehat{\epsilon_5}$$

Second, the estimated $\widehat{\gamma}$s are functions of the β and α parameters. The relationship can be described by the following system of equations:

$$\widehat{\gamma_{31}} = \alpha_{13}\beta_{11} + \alpha_{23}\beta_{21} \ , \ \widehat{\gamma_{32}} = \alpha_{13}\beta_{12} + \alpha_{23}\beta_{22} \ , \ \widehat{\gamma_{33}} = \alpha_{13}\beta_{13} + \alpha_{23}\beta_{23}$$

$$\widehat{\gamma_{41}} = \alpha_{14}\beta_{11} + \alpha_{24}\beta_{21} \ , \ \widehat{\gamma_{42}} = \alpha_{14}\beta_{12} + \alpha_{24}\beta_{22} \ , \ \widehat{\gamma_{43}} = \alpha_{14}\beta_{13} + \alpha_{24}\beta_{23}$$

$$\widehat{\gamma_{51}} = \alpha_{15}\beta_{11} + \alpha_{25}\beta_{21} \ , \ \widehat{\gamma_{52}} = \alpha_{15}\beta_{12} + \alpha_{25}\beta_{22} \ , \ \widehat{\gamma_{53}} = \alpha_{15}\beta_{13} + \alpha_{25}\beta_{23}$$

The remaining parameters can be obtained from solving the system of equations above, where we have 9 observables ($\widehat{\gamma_{31}}$-$\widehat{\gamma_{53}}$) and 9 unobservables (α_{13},α_{14},α_{15},α_{23},α_{24},α_{25},β_{21} ,β_{22}, β_{23}). Note that we already estimated β_{11} ,β_{12}, β_{13} using M_1 and M_2. Finally, ϵ_{θ_2} can be identified from the correlation between the residuals of two of the measurements that include θ_2.

3.5 Results

In this section, I present the empirical findings from the econometric model described in the previous section. I first discuss the relationship between the three latent factors and the observer reports and child outcomes (Section 3.5.1). Next, I discuss the estimated relationship between observed child, family and teacher characteristics and the three factors (Section 3.5.2). Then, I discuss the estimated relationship between the three factors and the adult outcomes (Section 3.5.3). Finally, I discuss how these different results fit together and the implications for our understanding of children's noncognitive development (Section 3.5.4).

3.5.1 Mapping Factors to Measurements

Starting with the distribution of the three factors, I find a small positive correlation between the common factor and the maternal disagreement factor (Table 3.8). However, I find a zero correlation between the maternal and teacher disagreement factors. The small correlation between the factors is non-surprising given the low estimated correlation between the maternal and teacher reports described in Table 3.2. It is also explained by the relationship between these factors and the observed characteristics and investments that determine the factors, which I describe in Section 3.5.2.

In Table 3.9, I report the estimates of the factor loadings mapping the three factors to the observer reports and child outcomes. The factor loadings describe the relationship between each factor and its measurements, where a higher factor loading implies a higher association between the factor and the corresponding measurement. In order to identify the three factors, I restrict one of the factor loadings for each factor. I restrict the maternal and teacher disagreement factors to have a loading of 1 for the maternal and teacher reports of headstrong, and the common factor to have a loading of 1 for the number of monthly days absent from school.

The factor loadings for the maternal disagreement factor are similar for the six maternal reports sub-scales. The one exception is peer problems, with a factor loading of 0.32. This suggests that each of the first five sub-scales explain the maternal disagreement factor by a similar rate. On the other hand, the teacher disagreement factor disproportionately load higher on the anxious/depressed scale, which suggests that this factor is mainly explained by this subscale. The common factor loads more heavily in the teacher report scales than in the maternal report scales, suggesting that it is mainly driven by the information on the teacher reports.

The factor loadings, however, paint an imcomplete picture of the relationship between factors and measurements. The meaning of the loadings depends on the variance of each factors. For example, while the loadings for the common factor and teacher disagreement are similar for maternal report of headstrong, the maternal disagreement factor explain a large variation of that sub-scale because it has a larger variance. This can be seen in Table 3.10, where I describe the variation of each measurement explained by the three factors. The maternal disagreement factor explain about 67% of the variation in headstrong, in comparison to 6.7% by the common factor.

What is interesting from Table 3.10 is that while the common factor explains a large percentage of the variation in the teacher reports (47.1%-93.5%), it only explains a small percentage of the variation in the maternal reports (4.1%-16.2%). That is, the maternal disagreement factor explains a larger proportion of the maternal reports (31.6%-69.8%). In comparison, the teacher disagreement factor explains a much smaller share of the teacher reports (2%-46.6%). It is also important to note that the remaining measurement error, the variance not explained by both factors, in the maternal reports (28.2%-63.1%) is higher than the measurement error in the teacher reports (3.7%-50.3%).

These results show that maternal and teacher reports are explained by different factors, which explain the low estimated correlation in Table 3.2. While the teacher reports are mainly explained by the common factor the maternal reports are mainly explained by the maternal disagreement factor. These results will be important when interpreting the implications of my findings.

3.5.2 Explaining the Factors

Moving on, Table 3.11 describes the estimates for the relationship between the estimated factors and child, family and teacher characteristics. It is important to note that the estimated relationships can be explained by the effect of these family and school characteristics on the underlying noncognitive skill, but also by bias in reporting. As a result, differences in the relationship across the factors can be explained by both the idea that these factors are measuring different dimensions of children's noncognitive skills and the idea that these factors are picking up differences in observer bias in reporting. It is also important to note that I do not provide exogenous variation in family and school investments, and as a result, the estimated relationships shown here provide suggestive evidence of what is being captured by the three estimated factors but not causal relationships.

The first column describes the relationship between the observable characteristics and the common factor. I find that child behavioral problems, as measured by the common factor, increase as children age, are higher for black children and lower for females. It is also significantly related to maternal investments in the form of cognitive stimulation and emotional support. Other family characteristics, such as maternal mental health, maternal education and family income, are not significant once I control for the measures of maternal investment. Interestingly, teacher and school characteristics, such as teach experience or number of students in the classroom, are not significantly related to the common factor.

The second column describes the relationship between the observable characteristics and the maternal disagreement factor. Interestingly, the family characteristics associated with this factor and very different than the variables associated with the common factor. While the common factor is higher for black children, the maternal disagreement factor is smaller for black children. Moreover, maternal education and family income are associated with the maternal factor but are not statistically related to the common factor. The most interesting result, however, is the strong association between maternal mental health and the maternal disagreement factor. It is not easily discernible from Table 3.11, but the maternal mental health measure is the main determinant of the maternal disagreement factor. The strong relationship between maternal mental health and maternal disagreement has been previously documented in the literature (see Boyle and Pickles (1997); Briggs-Gowan, Carter, and Schwab-Stone (1996); Fergusson, Lynskey, and Horwood (1993); Najman et al. (2000); Webster-Stratton (1988)). Most of the literature interpret this finding as evidence of bias in the maternal reporting of child behavioral problems (Boyle and Pickles, 1997; Najman et al., 2000). Later in the paper I provide evidence that this can also be explained by the idea that the maternal disagreement factor is measuring a factor related to the child's mental health.

The third column describes the relationship between observable characteristics and the teacher disagreement factor. Perhaps surprisingly, none of the teacher and school characteristics are statistically related to the teacher disagreement factor. The only two significant predictors are the child's race and gender. The teacher disagreement factor captures the information in teacher reports not captured by the common factor. However, as described in Section 3.5.1, most of the variation in the teacher reports is explained by the common factor. The lack of explanatory power of the teacher disagreement factor together with the fact that it is uncorrelated with most teacher and family characteristics suggest that the teacher disagreement factor is mainly capturing random measurement error in the teacher reports. This idea is reinforced when looking at the correlation with adult outcomes in the next Section.

3.5.3 Factors and Adult Outcomes

Literature studying disagreement in teacher and maternal reports has generally limited attention to the determinants of the disagreement. In contrast, I explore the relationship between the different factors that explain teacher and maternal reports and different dimensions of adult outcomes at age 23. The results are shown in Table 3.12, where I access the relationship between the three factors and four distinct outcomes capturing educational attainment, labor market involvement, risky behaviors and mental health. Importantly, I control for confounders, such as the child's cognition, gender and race, maternal education, maternal mental health and family income measured in 1997.

The common factor is the sole main predictor of years of completed education at age 23. Similarly, the common factor is strongly related to whether the individual is working or studying at age 23 and strongly related to a measure of risky behavior also measured at age 23. It is not, however, statistically or economically related to mental health measured by the K6 distress scale. In contrast, the maternal disagreement factor is unrelated to measures of educational attainment, employment or risky behaviors. It is however, strongly related with the K6 distress scale, even after I control for maternal mental health. The teacher disagreement factor is unrelated to all adult outcomes.

These results are supported by the results from the crude model described in Section 3.3.3. There I showed that, once I control for the teacher report, the disagreement factor was statistically associated with the K6 distress scale but not with educational attainment, employment and risky behaviors. That is, in both the latent factor model and the crude model the additional information contained in the maternal reports is associated with the mental health scale but not with the other adult outcomes.

3.5.4 Implications

As shown in Section 3.3.3, mothers and teachers systematically disagree when reporting on children's behavioral problems. As described in Section 3.2, there are three main explanation for the maternal-teacher disagreement. First, it can be explained by random measurement error in both reports. Second, mothers and teachers observe children in different settings and, as a result, could be observing and reporting on different dimensions of children's noncognitive skills. Third, disagreement can be explained by systematic bias in the observer reports if certain type of observer tend to over or under report children's noncognitive skills.

The latent factor model provides suggestive evidence on which of the three sources are at play. First I find that teacher and maternal reports measure different factors. As described in Section 3.5.1 and shown in Table 3.10, maternal reports are mainly explained by the maternal disagreement factor and teacher reports are mainly explained by the common factor. The teacher disagreement factor explain a smaller share of the teacher reports. Second, different socioeconomic characteristics are associated with these factors. While child's age, race and gender and family investments are the main predictors of the common factor, maternal mental health, emotional support and family income are the main predictors of the maternal disagreement factor. The teacher disagreement factor is unrelated to most socioeconomic characteristics. Third, these different factors are associated with different adult outcomes. While the common factor is strongly related to educational attainment, employment and risky behaviors at age 23, the maternal disagreement factor is strongly related to measured mental health at age 23. The teacher disagreement factor is unrelated to any of the four adult outcomes and, as a result, seems to be capturing random measurement error in the teacher reports.

These results suggest that maternal and teacher reports are measuring very different aspects of child development. Teacher reports seem to be measuring child misbehavior associated with risky behaviors in adulthood and negative schooling and labor market outcomes. On the other hand, maternal reports seem to be measuring both child misbehavior and a factor related to the child's mental health. This difference explains the preliminary results in Tables 3.5 and

3.7. For example, it explains why maternal mental health and emotional support are strong predictors of maternal reports and why both emotional support and cognitive stimulation are important determinants of teacher reports (Table 3.7. Similarly, it explains why teacher reports are stronger predictors of educational attainment and employment at age 23 but weaker predictors of adult mental health than maternal reports.

My results highlight the empirical challenges in measuring noncognitive skills, as even when using the same scale different observers report on different aspects of children's noncognitive development. As a result, researchers using the same scale but measured from different observers could end up with very different conclusions about the returns of investing in children. For example, using the maternal reports one would conclude that maternal mental health is an important determinant of children's misbehavior. However, using teacher reports one would conclude that maternal mental health is not relevant at predicting children's misbehavior (Table 3.5). Similarly, researchers using the same scale but measured from different observers could end up with very different conclusions about the importance of noncognitive skills in determining adult outcomes. For example, the estimated returns of noncognitive skills for educational attainment using teacher reports is twice the size of the estimated returns using maternal reports (Table 3.7). This highlights the need for a better understanding of what aspects of children noncognitive development are measured by different scales and observers, and how to better elicit noncognitive skills from observer reports.

3.6 Conclusion

Economists have recognized the importance of noncognitive skills developed in childhood in explaining a variety of adult outcomes. Measuring noncognitive skills in children, however, present unique methodological challenges. In this paper, I show that the same scale when applied to different observers measure very different aspects of children's noncognitive development.

An important open question left for future research is why mothers and teachers react to different pieces of information when asked the same questions about the child's behaviors. Conceptually, this can be explained by either systematic bias in maternal and teacher reports or by the fact that mothers and teachers observe children in different settings – at home and at the school – and, as a result, are reporting on different dimensions of noncognitive skills. Understanding these differences is important when thinking about policy interventions that target children noncognitive development.

One direction for future research would be to explore whether disagreement are present on different dimensions of children's noncognitive skills. For example, the questions in the BPI scale can be separated into questions eliciting children externalizing behaviors, capturing anxious, aggressive and outwardly-expressed behaviors, and internalizing behaviors, capturing withdrawn and inhibited behaviors. Moreover, previous research suggests that parents are better informants of children's internalizing problems than teachers, whereas teachers are better informants of children's externalizing problems than parents, which can explain the findings in this paper (Smith, 2007). Another direction for future research would be to explore whether parents and teachers beliefs on children's noncognitive development can be influenced by the social context. For example, Kinsler, Pavan et al. (2016) shows that parents beliefs on their child's ability are influenced by the school composition. It would be interesting to see if their findings extend to beliefs on children's noncognitive skills, and whether teachers are also influenced by the school composition.

3.7 Tables and Figures

Table 3.1: Summary statistics

	Mean	s.d.	Min	Max
Child's Age	9.26	2.36	4	13
Child is Female	0.51	0.50	0	1
Child's is Black	0.37	0.48	0	1
Birth Order	2.03	1.09	1	9
Single Mother	0.26	0.44	0	1
Mother's Years of Educ.	13.03	2.26	2	17
Mother's K6 Distress Scale	3.60	3.53	0	22
Mother's Age at Birth	27.46	5.29	14	42
Home Cognitive Stimulation	0.26	0.87	-3.34	1.99
Home Emotional Support	0.56	0.89	-2.92	2.11
Family Income	51561.41	50359.92	0	577000
Teacher Experience	14.67	9.18	0.5	30
Teacher has Masters	0.47	0.50	0	1
Teacher is Black	0.13	0.34	0	1
More Than 2 Teachers	0.43	0.49	0	1
Large Classroom (25+)	0.34	0.47	0	1
Years of Education at 23	13.26	1.83	6	17
Working or a Student at 23	0.82	0.39	0	1
Risky Behaviors Scale at 23	1.41	0.69	1	5
K6 Distress Scale at 23	4.88	3.75	0	24
Observations		834		

Table 3.2: Behavior Problems Index: Questions and Subscales

Subscale	Question
Anxious/Depressed	(He/She) has sudden changes in mood or feeling.
Anxious/Depressed	(He/She) feels or complains that no one loves him/her.
Anxious/Depressed	(He/She) is too fearful or anxious.
Anxious/Depressed	(He/She) feels worthless or inferior.
Anxious/Depressed	(He/She) is unhappy, sad or depressed.
Headstrong	(He/She) is rather high strung and nervous.
Headstrong	(He/She) argues too much.
Headstrong	(He/She) is stubborn, sullen, or irritable.
Headstrong	(He/She) has a very strong temper and loses it easily.
Antisocial	(He/She) cheats or tells lies.
Antisocial	(He/She) bullies or is cruel or mean to others.
Antisocial	(He/She) is disobedient.
Antisocial	(He/She) does not seem to feel sorry after (he/she) misbehaves.
Antisocial	(He/She) breaks things on purpose or deliberately destroys things.
Hyperactive	(He/She) has difficulty concentrating, cannot pay attention for long.
Hyperactive	(He/She) is easily confused, seems to be in a fog.
Hyperactive	(He/She) is impulsive, or acts without thinking.
Hyperactive	(He/She) has difficulty getting (his/her) mind off certain thoughts.
Hyperactive	(He/She) is restless or overly active, cannot sit still.
Peer Problems	(He/She) has trouble getting along with other children.
Peer Problems	(He/She) is not liked by other children.
Peer Problems	(He/She) is withdrawn, does not get involved with others.
Dependent	(He/She) clings to adults.
Dependent	(He/She) cries too much.
Dependent	(He/She) demands a lot of attention.
Dependent	(He/She) is too dependant on others.

Notes: For each statement, mothers and teachers were asked whether “the following statement is not true, sometimes true, or often true, of the child’s behavior”.

Table 3.3: Behavioral Problems: Child Outcomes

Subscale	Question
Suspended	Has child ever been suspended or expelled from school?
Saw a Beh. Specialist	Has child ever seen a psychiatrist, psychologist, doctor, or counselor about an emotional, mental, or behavioral problem?
Days Absent	How many days in the past month has the target child been absent?

Notes: The mother was asked the questions about whether the child has ever been suspended form school and whether the child has ever seen a psychiatrist. On the other hand, the teacher answered the question about student absence from school.

Table 3.4: Behavior Problem Index

	Maternal Report	Teacher Report	Disagreement	Correlation
Total Score	9.316	7.366	1.950	0.362***
	(8.020)	(9.063)	(9.688)	
Anxious/Depressed	1.962	1.182	0.741	0.288***
	(1.853)	(1.799)	(2.172)	
Headstrong	2.087	1.265	0.823	0.227***
	(1.981)	(1.909)	(2.421)	
Antisocial	1.445	1.202	0.224	0.300***
	(1.742)	(2.071)	(2.262)	
Hyperactive	1.858	2.005	-0.163	0.421***
	(2.120)	(2.481)	(2.484)	
Peer Problems	0.456	0.827	-0.379	0.262***
	(0.934)	(1.273)	(1.364)	
Dependent	1.520	0.815	0.700	0.221***
	(1.745)	(1.335)	(1.942)	
Observations	834	834	834	

Notes: Results for the first three columns are for mean coefficients with standard deviations in parentheses. For the correlations, * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 3.5: Preliminary: Investments

Variable	Maternal Report	Teacher Report	Disagreement
Child's Age	-.001	.394***	-.371***
Child's Age Sqrd.	.0008	-.020***	.019***
Child's is Black	-.056	.117	-.155*
Child is female	-.084	-.258***	.173***
Birth Order	.009	.038	-.028
Single Mother	-.099	.087	-.162
Mother's Years of Educ.	.035**	-.005	.033*
Mother's Age at Birth	-.024***	-.019**	-.001
Mother's K6 Distress Scale	.070***	-.0006	.058***
Log-Family Income	-.103**	-.029	-.056
Home Emotional Support	-.181***	-.135**	-.020
Home Cognitive Stimulation	-.089*	-.142***	.061
Teacher Experience	-.003	-.002	-.0009
Teacher has Masters	.023	-.057	.072
Teacher is Black	-.00008	-.081	.076
More Than 2 Teachers	.018	.082	-.062
Large Classroom (25+)	.103	-.016	.099
Const.	1.146	-.767	1.655**

Notes: For ease of comparison, the maternal and teacher reports and the disagreement measure have been standardized to have mean 0 and variance 1. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 3.6: Preliminary: Investments (2)

Variable	[1]	[2]	[3]	[4]	[5]	[6]
Child's Age	-.347**	-.225	-.315**	-.187	-.359***	-.345***
Child's Age Sqrd.	.019**	.012	.016**	.010	.019***	.017**
Child's is Black	-.094	-.315***	-.256**	-.074	.013	.113
Child is female	.208***	.078	.149**	.160**	.081	.152**
Birth Order	-.059	.004	-.031	-.029	-.020	.025
Single Mother	-.040	-.225**	-.118	-.124	-.077	-.078
Mother's Years of Educ.	.0007	.014	.044**	.034*	.029	.036**
Mother's Age at Birth	-.004	-.005	.006	.00007	.002	-.0006
Mother's K6 Distress Scale	.058***	.050***	.035***	.041***	.027**	.038***
Log-Family Income	-.046	-.050*	-.062	-.029	-.002	-.071
Home Emotional Support	.057	-.074	-.040	.020	.022	-.055
Home Cognitive Stimulation	.051	.052	.030	.088*	.080*	-.032
Teacher Experience	-.003	-.002	-.002	-.003	.006	.0008
Teacher has Masters	.068	.074	.107	.035	.032	.045
Teacher is Black	-.080	.101	.140	.087	.106	.093
More Than 2 Teachers	-.073	-.057	-.021	.007	-.165**	-.011
Large Classroom (25+)	.068	.113	.086	.057	-.035	.091
Const.	1.923***	1.377**	1.292*	.496	1.013	1.606**

Notes: Columns 1-6 are OLS regressions on mother-teacher disagreement among the six BPI sub-scales. Each column represents a sub-scale in the following order: (1) Anxious, (2) Headstrong, (3) Antisocial, (4) Hyperactive, (5) Peer-Problems and (6) Dependent. For ease of comparison, the disagreement measures have been standardized to have mean 0 and variance 1.* denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 3.7: PRELIMINARY: OUTCOMES

Columns:	[1]	[2]	[3]	[4]	[5]
Outcome:	**Years of Education**				
Maternal Report	-.153**	.	.	-.304***	.
Teacher Report	.	-.298***	.	.	-.349***
Disagreement	.	.	.133**	.294***	-.079
Outcome:	**Working or Studying**				
Maternal Report	-.031*	.	.	-.052**	.
Teacher Report	.	-.043**	.	.	-.059**
Disagreement	.	.	.014	.040**	-.024
Outcome:	**Risky Behaviors Scale**				
Maternal Report	.036	.	.	.082**	.
Teacher Report	.	.086**	.	.	.094**
Disagreement	.	.	-.047	-.088**	.012
Outcome:	**K6 Distress Scale**				
Maternal Report	.382**	.	.	.374*	.
Teacher Report	.	.111	.	.	.429*
Disagreement	.	.	.206	.017	.474**

Notes: Control variables are included in all regressions. These include a measure for the child's cognition in 1997, the child's age and age squared in 1997, child gender and race dummies, and 1997 measures of maternal education, mental health and log-family income. For ease of comparison, the maternal and teacher reports and the disagreement measure have been standardized to have mean 0 and variance 1.* denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 3.8: Model: Latent Factor Correlation and Covariance Matrix

	Correlation Matrix		
	Common Factor	Maternal Disa.	Teacher Disa.
Common F.	1.00	0.07	-0.01
Mat. Dis.	0.07	1.00	0.11
Tea. Dis.	-0.01	0.11	1.00
	Covariance Matrix		
	Common Factor	Maternal Disa.	Teacher Disa.
Common F.	0.17	0.04	-0.00
Mat. Dis.	0.04	1.81	0.04
Tea. Dis.	-0.00	0.04	0.09

Notes: This table lists the estimated correlation matrix of the three latent factors.

Table 3.9: Model: Factor Loadings

	Measurement	Common Factor	Maternal Disa.	Teacher Disa.
Maternal Report	Headstrong	1.01	**1.00**	.
	Antisocial	1.23	0.74	.
	Anxious/Depressed	0.70	0.96	.
	Dependent	0.80	0.70	.
	Hyperactive	1.78	0.91	.
	Peer Problems	0.52	0.32	.
Teacher Report	Headstrong	3.90	.	**1.00**
	Antisocial	4.54	.	0.46
	Anxious/Depressed	3.06	.	3.23
	Dependent	1.84	.	1.36
	Hyperactive	4.48	.	1.39
	Peer Problems	2.14	.	1.55
Child Outcomes	Days Absent	**1.00**	.	.
	Saw a Beh. Specialist	0.16	.	.
	Suspended	0.21	.	.
Estimated Standard Deviation:		0.41	1.34	0.30

Notes: This table lists the factor loadings that express the relationship between each observed measure and the underlying factor it identifies. For the three child outcomes, I control for the child's age and age squared, for the child's race and gender, and for log-family income. For ease of comparison of the factor loadings, I describe the estimated standard deviation of each factor in the last row.

Table 3.10: Model: Percent Explained by Each Factor

	Measurement	Common F.	Mat. Disa.	Tea. Disa.	All Fact.	M. Error
Maternal Rep.	Headstrong	0.067	0.670	.	0.711	0.289
	Antisocial	0.119	0.485	.	0.575	0.425
	Anxious/Dep.	0.041	0.698	.	0.718	0.282
	Dependent	0.051	0.412	.	0.445	0.555
	Hyperactive	0.162	0.497	.	0.623	0.377
	Peer Problems	0.073	0.316	.	0.369	0.631
Teacher Rep.	Headstrong	0.793	.	0.020	0.815	0.185
	Antisocial	0.935	.	0.031	0.963	0.037
	Anxious/Dep.	0.471	.	0.466	0.943	0.057
	Dependent	0.350	.	0.144	0.497	0.503
	Hyperactive	0.600	.	0.026	0.628	0.472
	Peer Problems	0.510	.	0.205	0.726	0.274

Notes: This table describes the percentage of the variation in each measurement explained by each factor and by random measurement error. The unique variation of a measurement explained by a single factor is estimated by the R^2 obtained from an OLS regression of the measurement on the factor alone. The variation explained by both factors is estimated by the R^2 obtained from an OLS regressiong of the measurement on both factors together. Since the maternal (teacher) disagreement factor and the common factor are positively (negatively) correlated, the explanatory power of the two factors jointly is less (more) than the sum of the explanatory power of each individual factor.

Table 3.11: Model: Investments

Variable	Common Factor	Maternal Disa.	Teacher Disa.
Child's Age	0.122*	-0.150	0.111
Child's Age Sqrd.	-0.006*	0.009	-0.006
Child's is Black	0.128**	-0.311**	-0.183**
Child is female	-0.161***	0.072	0.118*
Birth Order	0.022	-0.024	-0.015
Single Mother	0.015	-0.181	0.052
Mother's Years of Educ.	-0.002	0.060**	0.005
Mother's Age at Birth	-0.009**	-0.027**	0.004
Mother's K6 Distress Scale	-0.001	0.122***	0.002
Home Cognitive Stimulation	-0.056**	-0.073	-0.019
Home Emotional Support	-0.055**	-0.231***	-0.019
Log-Family Income	-0.015	-0.154***	0.001
Teacher Experience	0.000	-0.005	-0.002
Teacher has Masters	-0.031	0.088	0.015
Teacher is Black	-0.062	0.063	0.080
More Than 2 Teachers	0.003	0.024	0.070
Large Classroom (25+)	-0.033	0.218*	0.056

Notes: * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Table 3.12: Model: Young Adult Outcomes

	Years of Edu.	Work. or Study?	Risky Beh.	K6 Distress
Common Factor	-0.905***	-0.137***	0.203**	0.148
Maternal Disa. Factor	-0.036	-0.014	0.010	0.308**
Teacher Disa. Factor	0.356	0.093	0.069	0.410

Notes: Control variables are included in all regressions. These include a measure for the child's cognition in 1997, age and age squared in 1997, child gender and race dummies, and 1997 measures of maternal education, mental health and log-family income. * denotes the coefficient is significant at the 10% level, ** denotes the coefficient is significant at the 5% level and *** denotes the coefficient is significant at the 1% level.

Appendix A

Appendix for Chapter 1

A.1 Results Assuming Exogeneity of Inputs

This section provides parameter estimates for the model when I do not control for the unobservable correlation across the different decisions and outcomes. As a result, the maternal psychological distress is assumed to enter exogenously in the model. The same true about family income and the mother's time with her child, which are assumed to be exogenous in the child human capital production function. The estimates presented in this section serve as comparison for the results presented in the main paper.

Table A.1: Psychological Distress

Constant	1.252	(0.146)
Years of Education	-0.041	(0.001)
Age	-0.053	(0.002)
Age sqrd.	0.001	(0.000)
Single	0.175	(0.008)
# of Children	0.051	(0.002)
Depressed at 17	0.449	(0.038)
White Dummy	-0.066	(0.006)
Mental Health Parity	-0.104	(0.008)

Bootstrap standard errors are reported in parentheses.

Table A.2: Cognition Production Function

	Parameter	First Dev. Stage		Second Dev. Stage	
Total Factor Productivity	K	0.579	(0.004)	1.019	(0.010)
Self Productivity	α_1	0.243	(0.026)	0.877	(0.074)
Log-Family Income	α_2	0.108	(0.003)	-0.010	(0.000)
Log Maternal Time Investment	α_3	0.072	(0.001)	0.001	(0.003)
Log Psychological Distress	α_4	-0.009	(0.010)	-0.013	(0.011)
Log-A_{t-1} $\times$ Log-F.Income	α_5	-0.000	(0.000)	-0.030	(0.010)
Log-A_{t-1} $\times$ Log-M.Time	α_6	0.005	(0.002)	0.036	(0.051)
Log-A_{t-1} $\times$ Log-Distress	α_7	0.002	(0.000)	-0.014	(0.021)
Log-F.Income $\times$ Log-M.Time	α_8	-0.000	(0.000)	0.000	(0.000)
Log-F.Income $\times$ Log-Distress	α_9	0.011	(0.005)	-0.009	(0.007)
Log-M.Time $\times$ Log-Distress	α_{10}	-0.076	(0.004)	0.002	(0.007)

Bootstrap standard errors are reported in parentheses.

Table A.3: Initial Child Ability

Constant	-0.842	(0.393)
Mother‘s Years of Education	0.001	(0.009)
Mother‘s Age at Child‘s Birth	0.013	(0.001)
Single	0.743	(0.128)
# of Siblings	0.064	(0.030)
White Dummy	-0.031	(0.041)
Female	0.136	(0.081)

Bootstrap standard errors are reported in parentheses.

Table A.4: Annual Hours Worked

Constant	272.423	(64.679)
Years of Education	76.661	(7.358)
Age at Child‘s Birth	0.039	(0.030)
Single	-84.558	(36.578)
# of Children	-307.750	(36.632)
White Dummy	421.549	(41.582)
Child is Female	-40.016	(38.192)
Child‘s Age	31.242	(7.017)
Child‘s Age sqrd.	-0.543	(0.380)
Log Non-Labor Income	-0.020	(0.015)
Median State Service Wage Rate	-18.494	(7.519)
State % Employed in Serv. Sector	811.018	(58.033)
State Variation in Welfare Rules 1	25.941	(26.690)
State Variation in Welfare Rules 2	-14.711	(19.557)
Child Younger Than 4	-81.770	(26.435)
Psychological Distress	8.963	(15.367)
Not Working Last Period	-0.779	(2.048)
Experience	2.471	(3.106)
Hours Working Last Period	0.394	(0.015)
Hours With the Child Last Period	-4.725	(2.869)

Bootstrap standard errors are reported in parentheses.

Table A.5: Weekly Time Investments

Constant	32.774	(6.967)
Years of Education	-0.340	(0.434)
Age at Child's Birth	-0.067	(0.065)
Single	-5.192	(1.614)
# of Children	0.059	(0.074)
White Dummy	-4.825	(0.745)
Child is Female	2.297	(0.415)
Child's Age	-0.252	(0.287)
Child's Age sqrd.	0.074	(0.005)
Log Non-Labor Income	-0.745	(0.574)
Median State Service Wage Rate	-1.599	(0.630)
State % Employed in Serv. Sector	20.420	(7.774)
State Variation in Welfare Rules 1	-2.590	(1.917)
State Variation in Welfare Rules 2	19.946	(8.863)
Child Younger Than 4	3.298	(1.939)
Psychological Distress	-0.654	(0.637)
Not Working Last Period	1.147	(2.263)
Experience	-0.028	(0.020)
Hours Working Last Period	0.001	(0.000)
Hours With the Child Last Period	0.117	(0.050)

Bootstrap standard errors are reported in parentheses.

Table A.6: Hourly Wages

Constant	-1.745	(0.325)
Years of Education	0.216	(0.005)
Age	0.016	(0.001)
Age sqrd.	-0.000	(0.000)
Median State Service Wage Rate	0.015	(0.007)
State % Employed in Serv. Sector	1.335	(0.645)
Log Psychological Distress	-0.024	(0.011)
Not Working Last Period	-0.148	(0.472)
Experience	-0.001	(0.001)

Bootstrap standard errors are reported in parentheses.

Appendix B

Appendix for Chapter 2

B.1 Additional Descriptive Statistics and Factor Analysis

Descriptive statistics for the full sample of observed 11-year-olds are found in Tables B.1 and B.2. Also, descriptive statistics for the BSAG by SES used in Section 2.5.2 are found in Table B.3 and the corresponding factor loadings are found in Table B.4. The remainder of this Section of the appendix explains how factor analysis is used in this paper (see Figure B.1 and Tables B.5-B.6).

In the National Child Development Survey (NCDS), childhood misbehavior and maladjustment are measured as follows. Teachers read a number of phrases and then report whether each phrase applies to the child in question. These measures are then aggregated into 10 variables capturing childhood maladjustment, known as the BSAG maladjustment variables.[1] We use factor analysis, a

[1] There are actually 12 BSAG variables available. We exclude two of the original variables from this study in order to maintain consistency with recent research using the same data set (see Shepherd (2013)). The two omitted variables are called Miscellaneous Symptoms and Miscellaneous Nervous Symptoms. Our main results do not change if these variables are included in the analysis. Results from this robustness test are available upon request from the authors.

statistical technique used for data reduction to assess whether classroom misbehavior can be represented using fewer than the ten dimensions available from the NCDS. These techniques reduce the number of dimensions by uncovering linear combinations of the original variables that contain most of the information in the data and that also have meaningful interpretations. Specifically, the analysis determines the number of dimensions needed to adequately describe observed variation in classroom behavior and which of the BSAG variables are related to which dimensions of classroom misbehavior. From here on, we refer to the original BSAG variables as *measurements* and each dimension of classroom misbehavior as *factors*.

We begin by writing the measurements of classroom behavior as a linear function of unobserved factors. If there are k underlying factors, we can write our original 10 BSAG variables as:

$$BSAG_{ji} = l_{j1} f_{1i} + ... + l_{jk} f_{ki} + \epsilon_{ji}, \text{ for } j = 1, \cdots, m, \text{ and } i = 1, \cdots, n \quad \text{(B.1)}$$

where $BSAG_{ji}$ is the value of the jth BSAG variable for individual i, f_{qi} is the unobserved value of the qth factor for individual i, l_{jq} represents the coefficient relating factors to measurements (usually referred to as the *factor loading*) and ϵ_{ji} is a residual, capturing measurement error. For each individual, we can rewrite equation (B.1) in matrix form as follows:

$$BSAG = LF + \epsilon \quad \text{(B.2)}$$

where $BSAG$ is the $(m \times n)$ measurement matrix, L is the $(m \times k)$ matrix of factor loadings and F is the $(k \times n)$ matrix of unobserved factors, where m is the number of measurements, n represents the number of observations for each measurement and k is the number of factors.

The first goal of factor analysis is to find a small number of unobserved factors ($k < m$) that sufficiently explain the variation in the measurements. We are interested in the variation within and across the measurements instead of the measurements *per se*. In other words, we are interested in explaining the measurement covariance matrix (V), which has size $m \times m$ and is given by:

$$V = BSAG\ BSAG^T = LL^T + \Omega \tag{B.3}$$

The above expression is valid under the assumptions that (i) F and ϵ are independent and (ii) $\mathbb{E}[F] = 0$, with $FF^T = I$, implying that the factors are uncorrelated and Ω is the diagonal error variance matrix. Then, the factor analysis problem amounts to understanding the symmetric positive semidefinite matrix LL^T. We can decompose this matrix using an eigen-decomposition, so that:

$$V - \Omega = LL^T = CDC^T \tag{B.4}$$

$$\approx L_k L_k^T = C_k D_k C_k^T$$

where C is the $(m \times m)$ matrix whose columns are the eigenvectors of $L^T L$ and D is the $(m \times m)$ diagonal matrix whose entries are the eigenvalues. The key 'trick' of factor analysis is that we can reduce the dimensionality of F (or L) by picking the k eigenvalues and eigenvectors that explain a lot of the measurement variance $U \equiv (V - \Omega)$. To accomplish this, we use C_k, the $(m \times k)$ matrix whose columns are the eigenvectors associated with the k largest eigenvalues, and define as D_k the diagonal matrix of the eigenvalues. The resulting $(m \times m)$ matrix $C_k D_k C_k^T$ is not equal to U, but will converge to U as k gets closer to m and will be closer to U than any other matrix with rank k.

Now that we understand the idea behind the factor analysis, we need to decide on k, the number of factors to be used. There are three widely used criteria in psychometrics. The first method is commonly known as Kaiser's criterion or Kaiser's stopping rule. It stipulates that only the number of latent independent factors with eigenvalues greater than 1 should be considered in the analysis. Recall, for a given factor, the eigenvalue measures the variance in all measures that is accounted for by that factor. A low eigenvalue means that the factor contributes little to explaining variance and may be treated as redundant and therefore ignored. The second method is known as the scree plot method. In this method, the researcher plots the relationship between the relative magnitude of the eigenvalues and the number of factors. The researcher then examines the scree plot and decides where the line stops descending precipitously and levels out. The number of points along the precipitously dropping part of the line, excluding the transition point, gives the number of factors that should be used in the analysis. The third method is known as parallel analysis. In this method, we create a dataset with random numbers and the same number of observations and variables as in the original data set. Then we compute the eigenvalues for each factor as we did for the original data using factor analysis. The researcher should keep only the number of factors where the eigenvalues from the random data are smaller than the eigenvalues from the factor analysis using the original data as the remaining factors are effectively capturing random noise.

We use all three methods on the BSAG maladjustment variables in our data and all suggest we should use exactly two factors in our analysis. Our findings match those of Ghodsian (1977) and Shepherd (2013), who also studied childhood misbehavior using NCDS data. To perform the test using Kaiser's criterion, we compute the eigenvalues of the correlation matrix together with the eigenvectors corresponding to each factor. We plot the eigenvalues of the factors in descending order in Figure B.1. Factors 1 and 2 have eigenvalues of 3.69 and 1.78 respectively, whereas factors 3 to 10 have eigenvalues between 0.9 and 0.3. The Kaiser's stopping rule suggests keeping only the factors with eigenvalues equal or higher than 1, or the first two factors in our analysis. In the same figure, we also notice that the line connecting the eigenvalues stops descending and levels out after the second factor. The scree plot method then suggests that we should keep and use only the first two factors. Lastly, in the same figure, we plot the eigenvalues from the random data created by the parallel analysis (dashed line). The two lines intersect before the third factor, suggesting that factors 3 to 10 capture a level of variation in the data that is generated by random noise. Hence, only the first two factors should be used.

Now that we have shown that two independent random variables or "factors" explain the ten BSAG maladjustment variables, it remains to be determined which set of BSAG variables are related to which factor. Deciding on which measurement is related to which factor is less straightforward than deciding on the number of factors. Since the two eigenvectors can be rotated in an infinite number of ways, we rotate the original eigenvectors in order to maximize the variance accounted for by the first two factors using the *quartimin* method. This produces the rotated factor loadings shown in Table B.5. Clearly, there is an association between the first 6 BSAG variables and the first factor. Coefficients range from 0.53 for anxiety of acceptance by adults to 0.80 for inconsequential behavior, which are fairly high. Moreover, intuitively, all these variables seem to be measures of outwardly expressed behaviors. Similarly, there is a clear association between the 7-9th measurements and the second factor. Again, these variables all represent inwardly expressed behavior. The last measurement is less clear. It seems to be statistically related to both factors and it is less clear intuitively if this behavior is inwardly or outwardly expressed. As a result, we permit the 10th behavior to be related to both latent factors.

It turns out this particular mapping between measurements and factors is the same as the one proposed by Ghodsian (1977) and used in other research (see, e.g., Shepherd (2013)). The first six measurements, taken together with the last measure, capture externalizing behavior and the last four capture internalizing behavior. This leads to the mapping from the BSAG measures to the two factors presented in Table 2.1. Finally, once we have decided on the mapping, it remains to estimate the two unobserved factors for each individual conditional on each individual's observed measurements, to be used in our reduced-form analysis. There are many methods available to compute the unobserved factor matrix F. One widely used method to estimate the unobserved factor is the "regression" or "Thompson" method, which constructs the weighting matrix b that minimizes the mean squared error in Equation (B.2). The resulting formula for F is given by:

$$F = L^T V^{-1} BSAG \tag{B.5}$$

where the weighting matrix, also known as the factor score matrix, is given by $b = L^T V^{-1}$. Table B.6 presents the factor scores, or weights, used to construct each latent factor.

In the econometric model in the main paper we use a similar approach to estimate the unobserved factors. There we take both the number of factors and the mapping between the measurements and factors as given. We also re-estimate the factor loadings using the restrictions imposed by the mapping and estimate the resulting distribution of each factor separately by gender. The key differences is that we estimate the factor loadings together with the other relevant outcomes. In a sense, the econometric model allows us to use the outcome equations as additional measurements for the unobserved factors. Moreover, the measurement error model by-passes the estimation of the unobserved factors directly as we use the estimated distribution to integrate out the unobserved factors in the estimation. Both modifications are beneficial. The first modification allows us to use additional variation to identify the latent factors. The second modification allows us to reduce the measurement error in the construction of the latent factors.

B.2 Additional Reduced-Form Evidence

This appendix contains results from additional reduced-form specifications relating externalizing behavior to earnings. Similar to what we did in Section 2.2.4, in all results presented here, we construct skills by summing up corresponding observable BSAG measurements and test scores.

In Tables B.7 and B.8, we explore this relationship once we have controlled for selection into education. In Figures B.2 and B.3 and Tables B.9 and B.10, we explore possible non-linearities, non-monotonicites and interaction effects in the relationship between the unobserved factors and earnings. The general conclusion is the following: the positive relationship between externalizing and adult earnings holds under all specifications and, moreover, there is little evidence of non-linearities in the relationship. Lastly, In Tables B.11 and B.12 we explore if the positive relationship between externalizing behavior and earnings still holds as individuals age.

In the measurement error model in the main text, we jointly model how latent factors affect sorting into education along with other decisions and outcomes. Here, we ask whether the positive reduced-form relationship between externalizing and earnings discussed in Section 2.2.4 holds once we more formally account for selection into schooling. We employ the widely-used Lee (1983) and Dubin and McFadden (1984) methods for selection bias correction. These methods are akin to a two-stage least squares approach when selection is specified as a multinomial logit model.[2] As in the measurement error model, excluded variables include: class size, average class preparation, number of children in the household, mother's education and father's education.

The first stage estimates can be found in Table B.7 and are not very different from what has been shown in the rest of the paper. Cognition is the most important variable for the schooling decision and both other factors describing classroom behavior are negatively associated with educational attainment. The second stage estimates (in addition to OLS estimates for comparison) can be found in Table B.8. The effect of externalizing is positive among all educational groups albeit less so for individuals in higher education groups. Moreover,

[2]The method proposed in Lee (1983) is a generalization of the two-step selection bias correction introduced by Heckman (1979) that allows for any parameterized error distribution. The method proposed in Dubin and McFadden (1984) is also a generalization of the method proposed by Heckman (1979) with the further advantage in comparison to the method in Lee (1983) that it does not make any assumption on covariances between the error term in the outcome and selection equations. These and other selection methods based on the multinomial logit model have been reviewed by Bourguignon, Fournier, and Gurgand (2007).

controlling for selection into schooling does not alter the results in any important way. If anything, the relationship between externalizing behaviors and earnings becomes stronger after we control for selection into schooling.[3]

We do not allow for non-linearities and interactions between the unobserved factors in the measurement error model estimated in the main text. Here, we run additional regressions allowing for non-linearities and interactions. Figures B.2 and B.3 plot the impact of the externalizing behavior on log-earnings after we have controlled for additional regressors. In order to control for the other variables we regress log-earnings on the other explanatory variables and use the residuals from that regression as the dependent variable in the non-parametric regression. The impact of externalizing on earnings appears linear, with a decline in the effect for the few individuals with externalizing behaviors above 4 standard deviations from the mean. We also explore non-linearities in Table B.9, where we separately regress earnings on those individuals with externalizing behaviors that are one standard deviation above the mean versus individuals below that level. We do not find any difference in the relationship between the two groups. Last, in Table B.10 we allow for a quadratic term and interactions between the unobserved factors when regressing those on log-earnings. Again, we find no evidence that interactions or non-linearities are relevant for the factors capturing classroom behavior.

All labor market outcomes in the main paper were constructed when individuals were 33 years old. In this section of the appendix we explore if the

[3]Note that we use collapse education into three educational groups to increase each group's sample size. The reason is that the selection models we use automatically estimate different earnings equations for each education level. Using six educational groups therefore leads to many more coefficients and larger standard errors.

relationship between childhood behaviors and earnings change as individuals age. In Tables B.11 and B.12 we explore this relationship when individuals are 42 and 50 years old respectively. The main patterns remain as individuals age. That is, externalizing behaviors are positively related to earnings even as individuals age. The relationship between externalizing and earnings seems to peak at age 42 and then decrease when individuals reach age 50. It is possible the relationship falls due to changes in the control variables since we use control variables measured at age 33. Nonetheless, these results show that nothing special seems to be happening at age 33. We could have used labor market outcomes from any survey and still obtain similar results.

Appendix Tables and Figures

Table B.1: Summary Statistics - Full Sample

	Both	Males	Females	
No Formal Education	0.126	0.114	0.138	***
	(0.332)	(0.317)	(0.345)	
CSE	0.124	0.111	0.137	***
	(0.330)	(0.315)	(0.344)	
O Level	0.341	0.306	0.375	***
	(0.474)	(0.461)	(0.484)	
A Level	0.141	0.184	0.0997	***
	(0.348)	(0.387)	(0.300)	
Higher Education	0.142	0.144	0.139	
	(0.349)	(0.351)	(0.346)	
Higher Degree	0.126	0.141	0.111	***
	(0.332)	(0.348)	(0.314)	
Hourly Wage	6.749	7.645	5.666	***
	(3.063)	(2.969)	(2.815)	
Weekly Hours Worked	36.71	43.54	28.71	***
	(12.54)	(7.917)	(12.23)	
Weekly Earnings	259.8	329.2	175.9	***
	(152.3)	(135.0)	(127.8)	
Experience	140.4	158.9	122.6	***
	(55.69)	(50.80)	(54.35)	
In Paid Work	0.792	0.902	0.685	***
	(0.406)	(0.297)	(0.464)	
Self Employed	0.142	0.176	0.0987	***
	(0.349)	(0.380)	(0.298)	
Has a Partner	0.794	0.783	0.804	**
	(0.405)	(0.412)	(0.397)	
Number of Children	1.512	1.347	1.666	***
	(1.147)	(1.148)	(1.125)	
London		0.302	0.305	
		(0.459)	(0.460)	
Observations	15,356	7,899	7,457	15,356

Notes: Summary statistics for the full sample of 15,356 individuals observed at age 11. Statistics are reported separately for both genders (Column [1]), for males (Column [2]) and for females (Column [3]). For education categories, employment and partnership, entries are in the form of percentages divided by 100. Experience is measured in months and wages and weekly earnings are in 1992 Great British pounds.

Table B.2: Summary Statistics - BSAG Variables - Full Sample

	Both	Males	Females	
Hostility Towards Adults	0.904	1.079	0.719	***
	(1.946)	(2.088)	(1.766)	
Hostility Towards Children	0.288	0.336	0.237	***
	(0.805)	(0.892)	(0.699)	
Anxiety for Acceptance by Adults	0.559	0.545	0.573	
	(1.212)	(1.188)	(1.237)	
Anxiety for Acceptance by Children	0.334	0.464	0.197	***
	(0.803)	(0.953)	(0.575)	
Restlessness	0.229	0.286	0.169	***
	(0.568)	(0.633)	(0.484)	
Inconsequential Behavior	1.433	1.887	0.953	***
	(1.999)	(2.278)	(1.513)	
Depression	1.049	1.196	0.893	***
	(1.546)	(1.614)	(1.454)	
Withdrawal	0.347	0.410	0.279	***
	(0.826)	(0.910)	(0.720)	
Unforthcomingness	1.606	1.630	1.582	
	(2.137)	(2.059)	(2.216)	
Writing Off of Adults and Adult Standards	1.019	1.263	0.760	***
	(1.703)	(1.911)	(1.406)	
Observations	15,356	7,899	7,457	15,356

Notes: Summary statistics for maladjustment syndrome scores for the full sample of 14,158 individuals observed at age 11. Measures constructed using teachers' reports of misbehavior or misconduct in school. Statistics are reported separately for both genders (Column [1]), for males (Column [2]) and for females (Column [3]). For each maladjustment syndrome, a child receives a score, which is an integer between 0 and 15, with 15 indicating persistent display of behavior described by the maladjustment syndrome. In the table, entries are averages for each syndrome for the analysis sample.

Table B.3: Summary Statistics - BSAG Variables, Subsamples by SES

	Both	High SES	Low SES	Diff
Hostility Towards Adults	0.765	0.700	1.108	***
	(1.756)	(1.647)	(2.210)	
Hostility Towards Children	0.240	0.217	0.360	***
	(0.719)	(0.676)	(0.901)	
Anxiety for Acceptance by Adults	0.515	0.483	0.686	***
	(1.152)	(1.098)	(1.386)	
Anxiety for Acceptance by Children	0.298	0.285	0.369	***
	(0.762)	(0.749)	(0.820)	
Restlessness	0.194	0.178	0.279	***
	(0.521)	(0.497)	(0.625)	
Inconsequential Behavior	1.263	1.166	1.769	***
	(1.868)	(1.774)	(2.231)	
Depression	0.932	0.857	1.324	***
	(1.452)	(1.380)	(1.728)	
Withdrawal	0.307	0.292	0.387	***
	(0.771)	(0.743)	(0.902)	
Unforthcomingness	1.477	1.414	1.810	***
	(2.035)	(1.992)	(2.221)	
Writing Off of Adults and Adult Standards	0.908	0.855	1.183	***
	(1.586)	(1.523)	(1.859)	
Observations	7296	6125	1171	7296

Notes: Summary statistics for maladjustment syndrome scores for our sample of 7,296 individuals. Measures constructed using teachers' reports of misbehavior or misconduct in school. Statistics are reported separately for all individuals (Column [1]), for individual that did not experience financial difficulties growing up (Column [2]) and for those that did (Column [3]. For each maladjustment syndrome, a child receives a score, which is an integer between 0 and 15, with 15 indicating a persistent display of behavior described by the maladjustment syndrome. In the table, entries are averages for each syndrome for the analysis sample. In Column [4], *, ** and *** mean that differences between males and females are significant at the 10, 5 and 1 percent levels, respectively.

Table B.4: Measurement Error Model: Factor Loadings, by SES

Latent Skill	Measures	[High SES]	[Low SES]
Externalizing Behavior	Inconsequential Behavior	1.000	1.000
	Hostility Towards Adults	1.448	1.807
	Hostility Towards Children	1.989	2.005
	Anxiety for Acceptance by Adults	0.934	0.993
	Anxiety for Acceptance by Children	1.570	1.665
	Restlessness	1.697	1.577
	Writing Off of Adults and Adult Standards	0.357	0.443
Internalizing Behavior	Withdrawal	1.000	1.000
	Depression	1.057	0.917
	Unforthcomingness	1.757	1.877
	Writing Off of Adults and Adult Standards	0.692	0.722
Cognition	Verbal Score on General Ability Test	1.000	1.000
	Reading Comprehension Test Score	0.584	0.612
	Mathematics Test Score	1.079	1.049
	Non Verbal Score on General Ability Test	0.740	0.789

Notes: This table lists the factor loadings that express the relationship between each observed measure and the underlying factor it identifies by SES groups.

Table B.5: FA: Rotated Factor Loadings

	Factor 1	Factor 2	Uniqueness
Hostility Towards Adults	0.72	0.19	0.45
Hostility Towards Children	0.73	0.09	0.45
Anxiety for Acceptance by Adults	0.53	-0.23	0.66
Anxiety for Acceptance by Children	0.75	-0.12	0.42
Restlessness	0.61	0.04	0.62
Inconsequential Behavior	0.80	0.15	0.32
Depression	0.39	0.67	0.39
Withdrawal	0.12	0.79	0.35
Unforthcomingness	-0.06	0.79	0.37
Writing Off of Adults and Adult Standards	0.53	0.54	0.42

Notes: This table contains the factor loadings to the two factors that we retain using the *quartimin* rotation.

Table B.6: Reduced Form: Factor Scoring Coefficients

Cognition		Externalizing		Internalizing	
Measurement	Coefficient	Measurement	Coefficient	Measurement	Coefficient
Verbal Score	0.27648	Hostility To Children	0.22625	Depression	0.34172
Non Verbal Score	0.26356	Hostility To Adults	0.22810	Withdrawal	0.35678
Reading Comp.	0.25578	Anxiety Adults	0.13640	Unforthcomingness	0.32127
Mathematics Score	0.27106	Anxiety Children	0.21961	Writing Off	0.30815
Copying Designs	0.13239	Restlessness	0.19218		
		Inconsequential	0.25119		
		Writing Off	0.19477		

Notes: This table contains the scoring coefficients from the regression scoring method, which are used as weights in the construction of proxies for the three unobserved factors in the reduced-form analysis in Section 2.2.4.

Table B.7: Reduced Form: Educational Attainment - First Stage

	O Level or A Level	Higher Ed. or Higher Deg.
Cognition	1.238***	2.064***
Externalizing	-0.070*	-0.174***
Internalizing	-0.155***	-0.225***

Notes: This table contains parameter estimates for the first stage IV regressions in Table B.8. The coefficients are estimated by multinomial logit, where the base group includes individuals with no formal education or with a CSE degree. Instruments include mother's education, father's education, class size, class preparation and number of children in the household measured at age 11. We also include a gender dummy.

Table B.8: Reduced Form: Externalizing and Log Weekly Earnings

	OLS	IV-DMF	IV-Lee	IV-Lee [M]	IV-Lee [F]
No Formal Ed. or CSE	0.039*	0.041**	0.042**	0.017	0.048
O Level or A Level	0.059***	0.062***	0.057***	0.042***	0.069**
Higher Ed. or Higher Deg.	0.010	0.017	0.019	0.070*	-0.052

Notes: This table contains parameter estimates for the externalizing variable from regressions used to link non-cognitive skills to earnings. We regress log earnings of workers on a set of observable variables along with proxies for unobserved skills. To construct proxies for unobserved skills, we apply principal components factor analysis to all the variables used to measure that skill. Column 1 displays the coefficients obtained by OLS. In Columns 2, we perform the Dubin-McFadden (1984) correction method in order to control for selection on schooling. In columns 3-5, we perform the Lee (1983) correction method for selection in schooling. Column 3 displays the results for the whole sample, while column 4 and 5 displays the results for males and females separately.

Table B.9: Extremes

Variable	[1]	[2]	[3]	[4]
Cognition	.273***	.091***	.253***	.078*
Externalizing	.156***	.082***	.173***	.052*
Internalizing	-.006	-.051***	-.061*	-.071***
CSE	.	.008	.	.123
O Level	.	.092**	.	.134
A Level	.	.204***	.	.235**
Higher Education	.	.421***	.	.230**
Higher Degree	.	.664***	.	.689***
london	.	.174***	.	.334***
Female	.	-.816***	.	-.961***
Has a Partner	.	.145***	.	.224***
Number of Children	.	-.152***	.	-.053*
Experience	.	.003***	.	.002***
Const.	5.295***	4.951***	5.094***	4.909***
Obs.	3552	3552	378	378

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to earnings. This table examines the possibility of non-linearities in the returns to externalizing behaviors. Models [3] and [4] include all individuals with externalizing behavior 1 standard deviation above the mean. Models [1] and [2] include the rest of the sample.

Table B.10: INTERACTIONS

Variable	[1]	[2]	[3]	[4]
Cognition	.090***	.086***	.089***	.086***
Cognition^2	.	.012	.	.012
Externalizing	.058***	.069***	.057***	.069***
Externalizing^2	.	-.005	.	-.006
Internalizing	-.053***	-.063***	-.052***	-.064***
Internalizing^2	.	.004	.	.006
Ext.$\times$ Cog.	.	.	-.008	-.007
Int.$\times$ Cog.	.	.	.002	.009
Ext.$\times$ Int.	.	.	-.003	.001
Const.	4.940***	4.917***	4.937***	4.915***
Obs.	3930	3930	3930	3930

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to earnings. To construct proxies for unobserved skills, we apply principal components factor analysis to all the variables used to measure that skill. All models include both male and female individuals and a gender dummy.

Table B.11: Log Weekly Earnings at 42

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Misbehavior	-.098***	-.028*	.	.	.	.	.
Externalizing	.	.	.036**	.049***	.054***	.039**	.072**
Internalizing	.	.	-.064***	-.051***	-.037***	-.046***	-.026
Cognition	.	.184***	.186***	.054***	.040**	.067***	.016
CSE	.	.	.	.184***	.131**	.152**	.153*
O Level	.	.	.	.281***	.189***	.154***	.238***
A Level	.	.	.	.424***	.297***	.272***	.289***
Higher Education	.	.	.	.623***	.385***	.271***	.503***
Higher Degree	.	.	.	.779***	.606***	.475***	.708***
Has a Partner	.	.	.	.	.018	.202***	-.107*
Number of Children	.	.	.	.	-.007	.013	-.044**
Experience	.	.	.	.	.002***	.001***	.002***
Skilled Manual Occu.	.	.	.	.	.084**	.094**	-.010
Skilled Non-manual Occu.	.	.	.	.	.108***	.180***	.094*
Managerial Occupation	.	.	.	.	.397***	.331***	.431***
Female	-.985***	-.968***	-.964***	-.907***	-.795***	.	.
London	.155***	.133***	.134***	.122***	.090***	.197***	-.020
Const.	6.105***	6.080***	6.080***	5.696***	5.231***	5.220***	4.626***
Obs.	4452	4452	4452	4452	4452	2226	2226

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to earnings at age 42. We regress log earnings of workers at age 42 on a set of observable variables at age 33 along with proxies for unobserved skills. The controls are all constructed for individuals when they were 33 years old. To construct proxies for unobserved skills, we sum up all variables used to measure that skill in subsequent analysis and then normalize each unobserved skill. Models [1]-[5] include all individuals and a gender dummy, Model [6] includes only males and Model [7] only females.

Table B.12: Log Weekly Earnings at 50

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Misbehavior	-.101***	-.033***	.	.	.	.	.
Externalizing	.	.	.019	.029**	.031**	.020	.040*
Internalizing	.	.	-.052***	-.045***	-.032***	-.045***	-.017
Cognition	.	.187***	.188***	.079***	.067***	.056***	.079***
CSE	.	.	.	.045	.012	.050	-.001
O Level	.	.	.	.135***	.079**	.028	.115**
A Level	.	.	.	.302***	.213***	.210***	.195***
Higher Education	.	.	.	.457***	.286***	.252***	.317***
Higher Degree	.	.	.	.582***	.452***	.413***	.472***
Has a Partner	.	.	.	.	.026	.154***	-.067
Number of Children	.	.	.	.	.025**	.017	.031*
Experience	.	.	.	.	.002***	.0008**	.002***
Skilled Manual Occu.	.	.	.	.	.071**	.081**	.064
Skilled Non-manual Occu.	.	.	.	.	.089***	.106**	.079**
Managerial Occupation	.	.	.	.	.329***	.317***	.322***
Female	-.745***	-.735***	-.733***	-.681***	-.609***	.	.
London	.191***	.169***	.168***	.156***	.142***	.220***	.070**
Const.	6.393***	6.361***	6.362***	6.111***	5.705***	5.729***	5.130***
Obs.	3639	3639	3639	3639	3639	1723	1916

Notes: This table contains parameter estimates from OLS regressions used to link non-cognitive skills to earnings at age 50. We regress log earnings of workers at age 42 on a set of observable variables at age 33 along with proxies for unobserved skills. The controls are all constructed for individuals when they were 33 years old. To construct proxies for unobserved skills, we sum up all variables used to measure that skill in subsequent analysis and then normalize each unobserved skill. Models [1]-[5] include all individuals and a gender dummy, Model [6] includes only males and Model [7] only females.

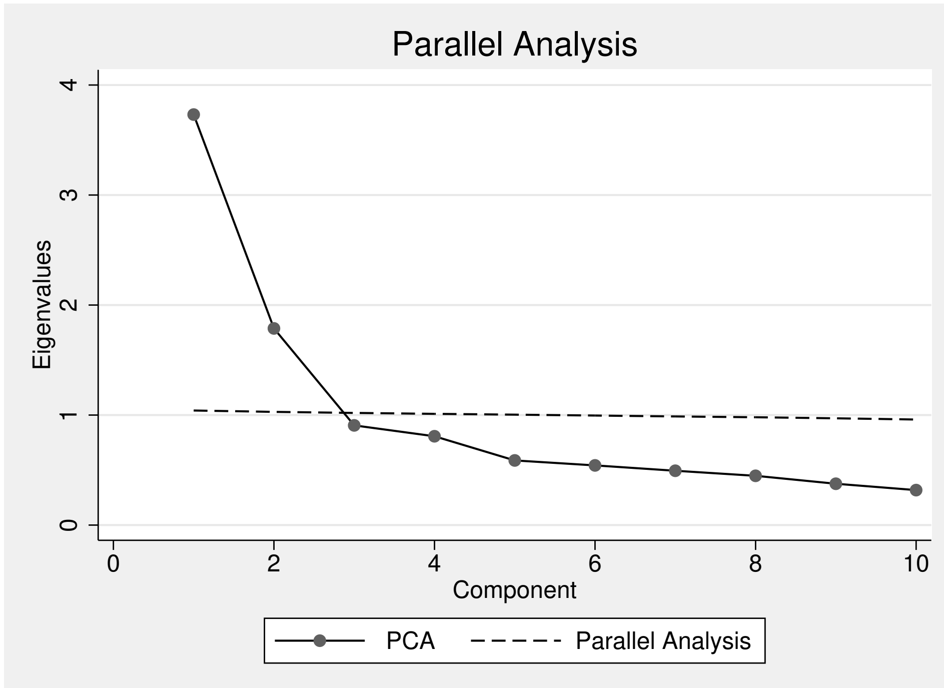


Figure B.1: Factor Analysis

The solid line depicts the eigenvalues associated with each factor in descending order in the principal component analysis. The dashed line depicts the eigenvalues computed from the random data created by the parallel analysis. From the principal component analysis, both the Kaiser's criterion and the scree plot test suggest that we should only keep the first two factors. Moreover, since the two lines intersect before the third factor, the parallel analysis suggests that only the first two factors are informative for our analysis and factors 3 to 10 are mostly random noise.

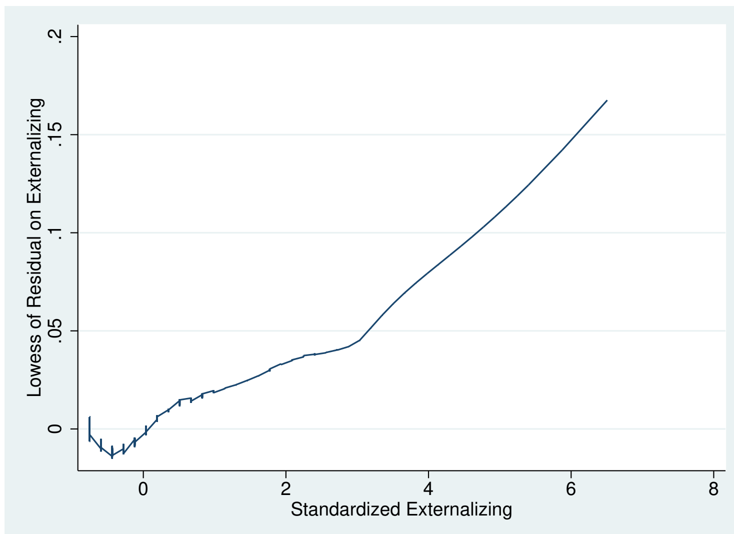


Figure B.2: Non-parametric regression

Here we use a non-parametric regression method (lowess) to plot the impact of the externalizing behavior on log-earnings controlled for cognition, internalizing behavior and a gender dummy. In order to control for the other variables we regress log-earnings on the other explanatory variables and use the residual of that regression as the explanatory variable in the non-parametric regression graphed here.

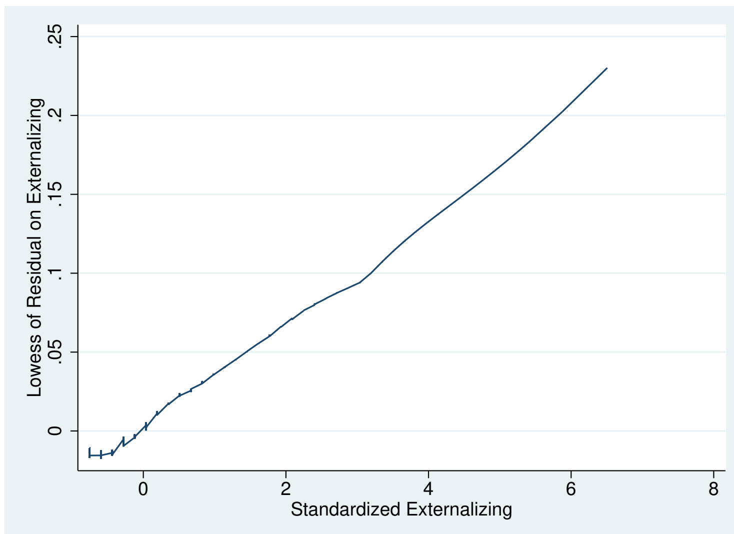


Figure B.3: Non-parametric regression with controls

Here we use a non-parametric regression method (lowess) to plot the impact of the externalizing behavior on log-earnings controlled for cognition, internalizing behavior, a gender dummy, educational choices, partnership status, fertility and experience. In order to control for the other variables we regress log-earnings on the other explanatory variables and use the residual of that regression as the explanatory variable in the non-parametric regression graphed here.

Bibliography

Achenbach, Thomas M., Stephanie H. McConaughy, and Catherine T. Howell. 1987. "Child/Adolescent Behavioral and Emotional Problems: Implications of Cross-Information Correlations for Situational Specificity." *Psychological Bulletin* 101 (2):213–232.

Agan, Amanda Y. 2011. "Non-Cognitive Skills and Crime." University of Chicago.

Agostinelli, Francesco and Matthew Wiswall. 2016. "Estimating the Technology of Children's Skill Formation." *Working Paper* .

Aizer, Anna. 2009. "Neighborhood Violence and Urban Youth." *The Problems of Disadvantaged Youth: An Economic Perspective* :275.

Aizer, Anna and Flavio Cunha. 2012. "The Production of Child Human Capital: Endowments, Investments and Fertility." Mimeo, Brown University.

Alexander, Karl, Doris Entwisle, and Linda Olson. 2014. *The Long Shadow: Family Background, Disadvantaged Urban Youth, and the Transition to Adulthood*. Russell Sage Foundation.

Allen, Joseph P., Robert C. Pianta, Anne Gregory, Amori Yee Mikami, and

Janetta Lun. 2011. "An Interation-Based Approach to Enhancing Secondary School Instruction and Student Achievement." *Science* 333.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." *Handbook of the Economics of Education* 4 (1).

Ammerman, Robert T, Frank W Putnam, Nicole R Bosse, Angelique R Teeters, and Judith B Van Ginkel. 2010. "Maternal Depression in Home Visitation: A Systematic Review." *Aggression and Violent Behavior* 15 (3):191–200.

Barry, Bruce and Raymond A Friedman. 1998. "Bargainer Characteristics in Distributive and Integrative Negotiation." *Journal of Personality and Social Psychology* 74 (2):345.

Barry, Colleen L and Susan H Busch. 2007. "Do State Parity Laws Reduce the Financial Burden on Families of Children with Mental Health Care Needs?" *Health Services Research* 42 (3p1):1061–1084.

———. 2008. "Caring for Children with Mental Disorders: do State Parity Laws Increase Access to Treatment?" *The Journal of Mental Health Policy and Economics* 11 (2):57–66.

Becker, Gary S. 1981. *A Treatise on the Family.*

Becker, Gary S and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* :S1–S39.

Becker, Gary Stanley. 1967. *Human Capital and the Personal Distribution of Income: An Analytical Approach.* Institute of Public Administration.

Bernal, Raquel. 2008. "The Effect of Maternal Employment and Child Care on Children's Cognitive Development." *International Economic Review* :1173–1209.

Bernal, Raquel and Michael P Keane. 2010. "Quasi-structural Estimation of a Model of Childcare Choices and Child Cognitive Ability Production." *Journal of Econometrics* 156 (1):164–189.

———. 2011. "Child Care Choices and Childrens Cognitive Achievement: The Case of Single Mothers." *Journal of Labor Economics* 29 (3):459–512.

Bertrand, Marianne and Jessica Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5 (1):32–64.

Blair, Clancy. 2010. "Stress and the Development of Self-Regulation in Context." *Child Development Perspectives* 4 (3):181–188.

Blair, Clancy, Douglas A Granger, Michael Willoughby, Roger Mills-Koonce, Martha Cox, Mark T Greenberg, Katie T Kivlighan, and Christine K Fortunato. 2011. "Salivary Cortisol Mediates Effects of Poverty and Parenting on Executive Functions in Early Childhood." *Child Development* 82 (6):1970–1984.

Blanden, Jo, Paul Gregg, and Lindsey Macmillan. 2007. "Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education." *The Economic Journal* 117 (519):C43–C60.

Boneva, Teodora and Christopher Rauh. 2016. "Parental Beliefs About Returns to Educational Investments-The Later the Better?" *Working Paper* .

Borghans, Lex, Angela Lee Duckworth, James J Heckman, and Bas Ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43 (4):972–1059.

Bourguignon, François, Martin Fournier, and Marc Gurgand. 2007. "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons." *Journal of Economic Surveys* 21 (1):174–205.

Boyle, Michael H and Andrew R Pickles. 1997. "Influence of Maternal Depressive Symptoms on Ratings of Childhood Behavior." *Journal of Abnormal Child Psychology* 25 (5):399–412.

Briggs-Gowan, Margaret J, Alice S Carter, and Mary Schwab-Stone. 1996. "Discrepancies Among Mother, Child, and Teacher Reports: Examining the Contributions of Maternal Depression and Anxiety." *Journal of Abnormal Child Psychology* 24 (6):749–765.

Brilli, Ylenia. 2014. "Mother's Time Allocation, Child Care and Child Cognitive Development." *Working Paper* .

Brown, Sarah and Karl Taylor. 2008. "Bullying, Education and Earnings: Evidence from the National Child Development Study." *Economics of Education Review* 27 (4):387–401.

Burwick, Andrew, Heather Zaveri, Lisa Shang, Kimberly Boller, Deborah Daro, and Debra A Strong. 2014. "Costs of Early Childhood Home Visiting: An

Analysis of Programs Implemented in the Supporting Evidence-Based Home Visiting to Prevent Child Maltreatment Initiative." *Mathematica Policy Research Reports* .

Camacho, Adriana. 2008. "Stress and Birth Weight: Evidence from Terrorist Attacks." *American Economic Review* :511–515.

Caplan, HL, SR Cogill, Heather Alexandra, Kay M Robson, R Katz, and R Kumar. 1989. "Maternal Depression and the Emotional Development of the Child." *The British Journal of Psychiatry* 154 (6):818–822.

Card, David. 2012. "Earnings, Schooling, and Ability Revisited." *35th Anniversary Retrospective* 35:111.

Carneiro, P, KT Hansen, and JJ Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44 (2):361–422.

Carneiro, Pedro Manuel, Katrine Vellesen Løken, and Kjell G Salvanes. 2011. "A Flying Start? Maternity Leave Benefits and Long Run Outcomes of Children." IZA discussion paper.

Carrell, Scott E and Mark L Hoekstra. 2010. "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids." *American Economic Journal: Applied Economics* 2 (1):211–228.

Chatterji, Pinka, Margarita Alegria, and David Takeuchi. 2011. "Psychiatric

Disorders and Labor Market Outcomes: Evidence from the National Comorbidity Survey-Replication." *Journal of Health Economics* 30 (5):858–868.

Churchill, Rachel, Theresa HM Moore, Toshi A Furukawa, Deborah M Caldwell, Philippa Davies, Hannah Jones, Kiyomi Shinohara, Hissei Imai, Glyn Lewis, and Vivien Hunot. 2013. "'Third Wave' Cognitive and Behavioural Therapies Versus Treatment as Usual for Depression." *The Cochrane Library* .

Cogill, SR, HL Caplan, Heather Alexandra, Kay Mordecai Robson, and R Kumar. 1986. "Impact of Maternal Postnatal Depression on Cognitive Development of Young Children." *British Medical Journal (Clin Res Ed)* 292 (6529):1165–1167.

Conger, Rand D, Xiaojia Ge, Glen H Elder, Frederick O Lorenz, and Ronald L Simons. 1994. "Economic Stress, Coercive Family Process, and Developmental Problems of Adolescents." *Child Development* 65 (2):541–561.

Conger, Rand D, Lora Ebert Wallace, Yumei Sun, Ronald L Simons, Vonnie C McLoyd, and Gene H Brody. 2002. "Economic Pressure in African American Families: a Replication and Extension of the Family Stress Model." *Developmental Psychology* 38 (2):179.

Conti, Gabriella, Sylvia Frühwirth-Schnatter, James J Heckman, and Rémi Piatek. 2014. "Bayesian Exploratory Factor Analysis." *Journal of Econometrics* 183 (1):31–57.

Cunha, Flavio. 2013. "Investments in Children When Markets Are Incomplete." Mimeo, Rice University.

Cunha, Flavio and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2):31–47.

Cunha, Flavio, James Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers* 57 (2):191–261.

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive skill formation." *Econometrica* 78 (3):883–931.

Cunningham, Peter J. 2009. "Beyond Parity: Primary Care Physicians Perspectives on Access to Mental Health Care." *Health Affairs* 28 (3):w490–w501.

Currie, Janet. 2001. "Early Childhood Education Programs." *The Journal of Economic Perspectives* 15 (2):213–238.

———. 2009. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* 47 (1):87–122.

Currie, Janet and Douglas Almond. 2011. "Human Capital Development Before Age Five." *Handbook of Labor Economics* 4:1315–1486.

Currie, Janet and Mark Stabile. 2006. "Child Mental Health and Human Capital Accumulation: the Case of ADHD." *Journal of Health Economics* 25 (6):1094–1118.

Dahl, Gordon B and Lance Lochner. 2012. "The Impact of Family Income on

Child Achievement: Evidence from the Earned Income Tax Credit." *The American Economic Review* 102 (5):1927–1956.

Dahlen, Heather M. 2016. "The Impact of Maternal Depression on Child Academic and Socioemotional Outcomes." *Economics of Education Review* .

De Los Reyes, Andres, David B Henry, Patrick H Tolan, and Lauren S Wakschlag. 2009. "Linking Informant Discrepancies to Observed Variations in Young Childrens Disruptive Behavior." *Journal of Abnormal Child Psychology* 37 (5):637–652.

De Los Reyes, Andres and Alan E Kazdin. 2005. "Informant Discrepancies in the Assessment of Childhood Psychopathology: a Critical Review, Theoretical Framework, and Recommendations for Further Study." *Psychological bulletin* 131 (4):483.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall. 2014. "Household Choices and Child Development." *The Review of Economic Studies* 81 (1):137–185.

Dohrenwend, Bruce P, Itzhak Levav, Patrick E Shrout, Sharon Schwartz, Guedalia Naveh, Bruce G Link, Andrew E Skodol, and Ann Stueve. 1992. "Socioeconomic Status and Psychiatric Disorders: the Causation-Selection Issue." *Science* 255 (5047):946–952.

Doyle, Orla, Colm P Harmon, James J Heckman, and Richard E Tremblay. 2009. "Investing in Early Human Development: Timing and Economic Efficiency." *Economics & Human Biology* 7 (1):1–6.

Drapeau, Aline, Alain Marchand, and Dominic Beaulieu-Prévost. 2011. "Epidemiology of Psychological Distress." *Mental IllnessesUnderstanding, Prediction and Control* 5:105–134.

Dubin, Jeffrey A and Daniel L McFadden. 1984. "An econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* :345–362.

Duckworth, Angela L, Christopher Peterson, Michael D Matthews, and Dennis R Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology* 92 (6):1087.

Duckworth, Angela L and David Scott Yeager. 2015. "Measurement Matters Assessing Personal Qualities Other than Cognitive Ability for Educational Purposes." *Educational Researcher* 44 (4):237–251.

Duncan, Greg J and Rachel Dunifon. 2012. "Introduction to 'Soft-Skills' and Long-Run Labor Market Success." *Research in Labor Economics* 35:309–312.

Duncan, Greg J, Ariel Kalil, and Kathleen M Ziol-Guest. 2013. "Increasing Inequality in Parent Incomes and Childrens Schooling." *Working Paper* .

Duncan, Greg J and Katherine Magnuson. 2011. "The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. Russell Sage Foundation, 47–69.

Dunning, David, Chip Heath, and Jerry M Suls. 2004. "Flawed Self-assessment

Implications for Health, Education, and the Workplace." *Psychological science in the public interest* 5 (3):69–106.

Durlak, Joseph A, Roger P Weissberg, Allison B Dymnicki, Rebecca D Taylor, and Kriston B Schellinger. 2011. "The Impact of Enhancing Students Social and Emotional Learning: A Meta-analysis of School-based Universal Interventions." *Child development* 82 (1):405–432.

Earls, Marian F et al. 2010. "Incorporating Recognition and Management of Perinatal and Postpartum Depression into Pediatric Practice." *Pediatrics* 126 (5):1032–1039.

Ehrler, David J, J Gary Evans, and Ron L McGhee. 1999. "Extending Big-Five Theory into Childhood: A Preliminary Investigation into the Relationship Between Big-Five Personality Traits and Behavior Problems in Children." *Psychology in the Schools* 36 (6):451–458.

Eisenberg, Daniel, Ezra Golberstein, Janis L Whitlock, and Marilyn F Downs. 2013. "Social Contagion of Mental Health: Evidence from College Roommates." *Health Economics* 22 (8):965–986.

Eissa, Nada and Hilary W Hoynes. 2006. "Behavioral Responses to Taxes: Lessons from the EITC and Labor Supply." In *Tax Policy and the Economy, Volume 20*. The MIT Press, 73–110.

Ettner, Susan L, Richard G Frank, and Ronald C Kessler. 1997. "The Impact of Psychiatric Disorders on Labor Market Outcomes." *Industrial & Labor Relations Review* 51 (1):64–81.

Evans, William N and Craig L Garthwaite. 2014. “Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health.” *American Economic Journal: Economic Policy* 6 (2):258–90.

Farmer, Elizabeth MZ. 1993. “Externalizing Behavior in the Life Course The Transition From School to Work.” *Journal of Emotional and Behavioral Disorders* 1 (3):179–188.

———. 1995. “Extremity of Externalizing Behavior and Young Adult Outcomes.” *Journal of Child Psychology and Psychiatry* 36 (4):617–632.

Fergusson, David M, Michael T Lynskey, and L John Horwood. 1993. “The Effect of Maternal Depression on Maternal Ratings of Child Behavior.” *Journal of Abnormal Child Psychology* 21 (3):245–269.

Frank, Richard G and Ellen Meara. 2009. “The Effect of Maternal Depression and Substance Abuse on Child Human Capital Development.” *Working Paper* .

Frijters, Paul, David W Johnston, and Michael A Shields. 2014. “The Effect of Mental Health on Employment: Evidence from Australian Panel Data.” *Health Economics* 23 (9):1058–1071.

Fronstin, Paul, David H Greenberg, and Philip K Robins. 2005. “The Labor Market Consequences of Childhood Maladjustment.” *Social Science Quarterly* 86:1170–1195.

Gershenson, Seth. 2016. “Linking Teacher Quality, Student Attendance, and Student Achievement.” *Education Finance and Policy* .

Gershenson, Seth, Stephen B Holt, and Nicholas W Papageorge. 2016. “Who Believes in Me? The Effect of Student–teacher Demographic Match on Teacher Expectations.” *Economics of Education Review* 52:209–224.

Ghodsian, M. 1977. “Children’s Behaviour and the BSAG: Some Theoretical and Statistical Considerations.” *British Journal of Social and Clinical Psychology* 16 (1):23–28.

Gregory, Mary K. 1977. “Sex Bias in School Referrals.” *Journal of School Psychology* 15 (1):5–8.

Griffen, Andrew. 2012. “Evaluating the Effects of Child Care Policies on Children’s Cognitive Development and Maternal Labor Supply.” *Working Paper* .

Gupta, Nabanita Datta, Mette Lausten, and Dario Pozzoli. 2016. “Does Mother Know Best? Parental Discrepancies in Assessing Child Behavioral and Educational Outcomes.” *Review of Economics of the Household* :1–19.

Hamre, Bridget K. and Robert C. Pianta. 2006. “Student-Teacher Relationships.” :49–60.

Harris, Katherine M, Christopher Carpenter, and Yuhua Bao. 2006. “The Effects of State Parity Laws on the Use of Mental Health Care.” *Medical Care* 44 (6):499–505.

Heckman, James J. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica* :153–161.

Heckman, James J and Flavio Cunha. 2007. “The Technology of Skill Formation.” *American Economic Review* 97 (2):31–47.

Heckman, James J and Tim Kautz. 2013. “Fostering and Measuring Skills: Interventions that Improve Character and Cognition.” NBER working paper.

Heckman, James J and Dimitriy V Masterov. 2007. “The Productivity Argument for Investing in Young Children.” *Applied Economic Perspectives and Policy* 29 (3):446–493.

Heckman, James J and Stefano Mosso. 2014. “The Economics of Human Development and Social Mobility.” *Annual Review of Economics* 6 (1):689–733.

Heckman, James J, Rodrigo Pinto, and Peter Savelyev. 2013. “Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes.” *American Economic Review* 103 (6):2052–86.

Heckman, James J and Yona Rubinstein. 2001. “The Importance of Noncognitive Skills: Lessons from the GED Testing Program.” *The American Economic Review* 91 (2):145–149.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior.” *The Journal of Labor Economics* 24 (3):411.

Henneberger, Angela K, Donna L Coffman, and Scott D Gest. 2016. “The Effect of Having Aggressive Friends on Aggressive Behavior in Childhood: Using Propensity Scores to Strengthen Causal Inference.” *Social Development* .

Herrick, H. 2015. "The Association of Psychological Distress, Chronic Disease and Health Risk Factors." *North Carolina Department of Health and Human Services Division of Public Health* 5:1–2.

Horwitz, Allan V. 2007. "Distinguishing Distress from Disorder as Psychological Outcomes of Stressful Social Arrangements." *Health* 11 (3):273–289.

Howard, Kimberly S and Jeanne Brooks-Gunn. 2009. "The Role of Home-Visiting Programs in Preventing Child Abuse and Neglect." *The Future of Children* 19 (2):119–146.

Hu, Yingyao. 2008. "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution." *Journal of Econometrics* 144 (1):27–61.

Hu, Yingyao and Susanne M Schennach. 2008. "Instrumental variable treatment of nonclassical measurement error models." *Econometrica* 76 (1):195–216.

Hunot, Vivien, Rachel Churchill, Vanessa Teixeira, and Mauricio Silva de Lima. 2007. "Psychological Therapies for Generalised Anxiety Disorder." *The Cochrane Library* .

Ivcevic, Zorana and Marc Brackett. 2014. "Predicting School Success: Comparing Conscientiousness, Grit, and Emotion Regulation Ability." *Journal of Research in Personality* 52:29–36.

Jackson, C Kirabo. 2012. "Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina." *National Bureau of Economic Research* (No. w18624).

Jackson, Michelle. 2006. "Personality Traits and Occupational Attainment." *European Sociological Review* 22 (2):187–199.

Judge, Timothy A, Beth A Livingston, and Charlice Hurst. 2012. "Do Nice Guys and Gals Really Finish Last? The Joint Effects of Sex and Agreeableness on Income." *Journal of Personality and Social Psychology* 102 (2):390.

Kalil, Ariel, Rebecca Ryan, and Michael Corey. 2012. "Diverging Destinies: Maternal Education and the Developmental Gradient in Time with Children." *Demography* 49 (4):1361–1383.

Kessler, Ronald C, Gavin Andrews, Lisa J Colpe, Eva Hiripi, Daniel K Mroczek, S-LT Normand, Ellen E Walters, and Alan M Zaslavsky. 2002. "Short Screening Scales to Monitor Population Prevalences and Trends in Non-specific Psychological Distress." *Psychological Medicine* 32 (06):959–976.

Kessler, Ronald C, Cindy L Foster, William B Saunders, and Paul E Stang. 1995. "Social Consequences of Psychiatric Disorders, I: Educational Attainment." *American Journal of Psychiatry* 152 (7):1026–1032.

Kessler, Ronald C, Jennifer Greif Green, Lenard A Adler, Russell A Barkley, Somnath Chatterji, Stephen V Faraone, Matthew Finkelman, Laurence L Greenhill, Michael J Gruber, Mark Jewell et al. 2010. "Structure and Diagnosis of Adult Attention-Deficit/Hyperactivity Disorder: Analysis of Expanded Symptom Criteria from the Adult ADHD Clinical Diagnostic Scale." *Archives of General Psychiatry* 67 (11):1168–1178.

Kinsler, Josh, Ronni Pavan et al. 2016. "Parental Beliefs and Investment in Children: The Distortionary Impact of Schools." *Working Paper* .

Lang, Matthew. 2013. “The Impact of Mental Health Insurance Laws on State Suicide Rates.” *Health Economics* 22 (1):73–88.

Lee, Lung-Fei. 1983. “Generalized Econometric Models with Selectivity.” *Econometrica* :507–512.

Levine, Ross and Yona Rubinstein. 2013. “Smart and Illicit: Who Becomes an Entrepreneur and Does it Pay?” NBER working paper.

Lundberg, Shelly. 2013. “The College Type: Personality and Educational Inequality.” *Journal of Labor Economics* 31 (3):421–441.

Mayer, Susan E. 2002. *The Influence of Parental Income on Children's Outcomes*. Wellington, New Zealand: Knowledge Management Group, Ministry of Social Development.

McLeod, Jane D and Karen Kaiser. 2004. “Childhood Emotional and Behavioral Problems and Educational Attainment.” *American Sociological Review* 69 (5):636–658.

McLoyd, Vonnie C. 1990. “The Impact of Economic Hardship on Black Families and Children: Psychological Distress, Parenting, and Socioemotional Development.” *Child Development* 61 (2):311–346.

Mirowsky, John and Catherine E Ross. 2003. *Social Causes of Psychological Distress*. Transaction Publishers.

Moffitt, Robert. 1983. “An Economic Model of Welfare Stigma.” *The American Economic Review* 73 (5):1023–1035.

———. 1992. “Incentive Effects of the US welfare System: A Review.” *Journal of Economic Literature* 30 (1):1–61.

Moffitt, Terrie E, Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J Hancox, HonaLee Harrington, Renate Houts, Richie Poulton, Brent W Roberts, Stephen Ross et al. 2011. “A Gradient of Childhood Self-control Predicts Health, Wealth, and Public Safety.” *Proceedings of the National Academy of Sciences* 108 (7):2693–2698.

Molyneaux, Emma, Louise M Howard, Helen R McGeown, Amar M Karia, and Kylee Trevillion. 2014. “Antidepressant Treatment for Postnatal Depression.” *Advances in Psychiatric Treatment* 20 (6):368–368.

Najman, Jake M, Gail M Williams, Jane Nikles, SUE Spence, William Bor, Michael O’callaghan, Robyne Le Brocque, and Margaret J Andersen. 2000. “Mothers’ Mental Illness and Child Behavior Problems: Cause-Effect Association or Observation Bias?” *Journal of the American Academy of Child & Adolescent Psychiatry* 39 (5):592–602.

Neal, Derek A and William R Johnson. 1996. “The Role of Premarket Factors in Black-White Wage Differences.” *Journal of Political Economy* 104 (5):869–895.

Nemeroff, Robin, Elizabeth Midlarsky, and Joseph F Meyer. 2010. “Relationships Among Social Support, Perceived Control, and Psychological Distress in Late Life.” *The International Journal of Aging and Human Development* 71 (1):69–82.

Papageorge, Nicholas W, Seth Gershenson, and Kyungmin Kang. 2016. "Teacher Expectations Matter." *Working Paper* .

Papageorge, Nicholas W, Victor Ronda, and Yu Zheng. 2017. "The Economic Value of Breaking Bad Misbehavior, Schooling and the Labor Market." *IZA Discussion Paper* (10822).

Parcel, Toby L and Elizabeth G Menaghan. 1988. "Measuring Behavioral Problems in a Large Cross Sectional Survey: Reliability and Validity for Children of the NLS Youth." *Manuscript. The Ohio State University* .

Persson, Petra and Maya Rossin-Slater. 2014. "Family Ruptures, Stress, and the Mental Health of the Next Generation." *Working Paper* .

Peterson, James L and Nicholas Zill. 1986. "Marital Disruption, Parent-Child Relationships, and Behavior Problems in children." *Journal of Marriage and the Family* :295–307.

Prochaska, Judith J, Hai-Yen Sung, Wendy Max, Yanling Shi, and Michael Ong. 2012. "Validity Study of the K6 Scale as a Measure of Moderate Mental Distress based on Mental Health Treatment need and Utilization." *International Journal of Methods in Psychiatric Research* 21 (2):88–97.

Ramey, David M. 2014. *The Social Control of Childhood Behavior via Criminalization or Medicalization: Why Race Matters*. Ph.D. thesis, The Ohio State University.

Rao, Calyampudi Radhakrishna and Sandip Sinharay. 2006. *Handbook of statistics: Psychometrics*, vol. 26. Elsevier.

Ridner, Sheila H. 2004. "Psychological Distress: Concept Analysis." *Journal of Advanced Nursing* 45 (5):536–545.

Rosenquist, J Niels, James H Fowler, and Nicholas A Christakis. 2011. "Social Network Determinants of Depression." *Molecular Psychiatry* 16 (3):273–281.

Ross, Catherine E. 2000. "Neighborhood Disadvantage and Adult Depression." *Journal of Health and Social Behavior* :177–187.

Roy, Andrew Donald. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2):135–146.

Russell, James M, Ernst R Berndt, Robert Miceli, Salvatore Colucci, and AN Grudzinski. 1999. "Course and Cost of Treatment for Depression with Fluoxetine, Paroxetine, and Sertraline." *American Journal of Management Care* 5 (5):597–606.

Samejima, Fumiko. 1969. "Estimation of Latent Ability Using a Response Pattern of Graded Scores." *Psychometrika Monograph Supplement* .

Sciulli, Dario. 2016. "Adult Employment Probabilities of Socially Maladjusted Children." *Journal of Behavioral and Experimental Economics* 60:9–22.

Segal, Carmit. 2013. "Misbehavior, Education, and Labor Market Outcomes." *Journal of the European Economic Association* 11 (4):743–779.

Shepherd, Peter. 2013. "Bristol Social Adjustment Guides at 7 and 11 years." Centre for Longitudinal Studies.

Smith, Steven R. 2007. "Making Sense of Multiple Informants in Child and Adolescent Psychopathology: A Guide for Clinicians." *Journal of Psychoeducational Assessment* 25 (2):139–149.

Spurk, Daniel and Andrea E Abele. 2011. "Who Earns More and Why? A Multiple Mediation Model From Personality to Salary." *Journal of Business and Psychology* 26 (1):87–103.

Stott, Denis Herbert, Emily G Sykes, and NC Marston. 1974. *The Social Adjustment of Children: Manual to the Bristol Social Adjustment Guides*. Hodder & Stoughton.

Urzua, S. 2008. "Racial Labor Market Gaps The Role of Abilities and Schooling Choices." *Journal of Human Resources* 43 (4):919–971.

van Huizen, TM and J Plantenga. 2015. "Universal Child Care and Childrens Outcomes: A Meta-Analysis of Evidence from Natural Experiments." *Working Paper* 15 (13).

Webster-Stratton, Carolyn. 1988. "Mothers' and Fathers' Perceptions of Child Deviance: Roles of Parent and Child Behaviors and Parent Adjustment." *Journal of Consulting and Clinical Psychology* 56 (6):909.

West, Martin R, Matthew A Kraft, Amy S Finn, Rebecca E Martin, Angela L Duckworth, Christopher FO Gabrieli, and John DE Gabrieli. 2015. "Promise and Paradox Measuring Students Non-Cognitive Skills and the Impact of Schooling." *Educational Evaluation and Policy Analysis* .

Willis, Robert J. and Sherwin Rosen. 1979. "Education and Self-Selection." *Journal of Political Economy* 87 (5):S7–S36.

Wolfson, Johanna, Joel H Fields, and Susan A Rose. 1987. "Symptoms, Temperament, Resiliency, and Control in Anxiety-Disordered Preschool Children." *Journal of the American Academy of Child & Adolescent Psychiatry* 26 (1):16–22.

Yeung, W Jean, Miriam R Linver, and Jeanne Brooks-Gunn. 2002. "How Money Matters for Young Children's Development: Parental Investment and Family Processes." *Child Development* :1861–1879.

Youngstrom, Eric, Rolf Loeber, and Magda Stouthamer-Loeber. 2000. "Patterns and Correlates of Agreement Between Parent, Teacher, and Male Adolescent Ratings of Externalizing and Internalizing Problems." *Journal of consulting and clinical psychology* 68 (6):1038.

Curriculum Vitae

Victor Andrea Checchia Ronda received a B.A. in Economics and a B.A. in Mathematics from the New York University in 2011. In that same year, he joined the Ph.D. program in Economics at Johns Hopkins University, where he received the Clarence M. Guggenheimer Award for outstanding performance in his work toward a Ph.D. in 2015. His research interests are in labor economics, health economics and economics of education.

Starting in August 2017, he will join the department of Economics and Business Economics at Aarhus University as Assistant Professor in a position associated with the Center for the Economics of Human Development at the University of Chicago.