Spare the rod? What can we say about the causal effect of spanking?*

C. Andrew Zuppann†

University of Houston

September 25, 2015

Abstract

Over the past 30 years the probability that a mother spanks her child has substantially declined. Over the same time the negative correlation between spanking and outcomes has significantly worsened. I show that the decline in spanking is driven through changes in social stigmas surrounding corporal punishment. Changes in the correlation between spanking and outcomes over time can therefore be used to estimate a causal treatment effect of spanking for households who changed their behavior due to the increasingly negative stigma. I estimate this treatment effect to be significant and generally positive: spanking improves childhood test scores, educational attainment, and labor market outcomes. However, spanking has a negative impact on childhood behavioral problems and non-cognitive measures.

*All remaining errors are my own. Please do not circulate without permission.

†Department of Economics, University of Houston. E-mail: cazuppann@uh.edu

1 Introduction

Many parents in the modern world have a strongly negative response to the idea of spanking their children. But how certain should we be about the negative impact of spanking on childhood development? Many studies have found a consistent negative correlation between corporal punishment and negative outcomes such as test scores, behavioral problems, and educational attainment. Although this is suggestive, interpreting these relationships as causal suffers from selection problems. Families where children are spanked are both observably less stable and lower SES both of which negatively impact children's outcomes. Families that spank might also have unobserved negative shocks that influence the decision to spank and child outcomes. Without a plausible source of exogenous variation, it is impossible to say whether the consistent negative correlations between spanking and outcomes is a causal relationship or driven by selection or unobserved family characteristics.

In this paper I use a battery of statistical and econometric methods to try to shed light on whether spanking is actually harmful for childhood development. I begin by documenting the extent of a negative correlation between spanking and child outcomes in the NLSY79 matched mother-children survey. I then consider three more 'structural' methods to try to control for potential biases that may exist in simple OLS estimates. First, I include controls for mother fixed effects to control for family level unobservable differences in the decision to spank. Second, I estimate the return to spanking using matching on propensity scores as another way to control for unobservable differences across families.

The third approach I take to controlling for the possible bias in OLS is to use control functions to estimate the conditional error terms that cause the bias. At the heart of this method is the fact that the observed correlation between spanking and child outcomes

has changed over time. Making a few admittedly stringent assumptions about why this change occurred over time allows me to estimate a causal impact of spanking on childhood cognitive and non-cognitive outcomes. If there have been changing cultural norms about the appropriateness of physically punishing a child then the change in observed correlations allows identification of the selection into spanking as a function of social norms and observed characteristics. Selection correction can then be applied to estimates of the effect of spanking to identify the causal effect, at least locally.

I use the National Longitudinal Survey of Youth 1979 (NLSY) and the National Longitudinal Survey of Youth - Children and Young Adults datasets to construct measures of parental spanking as well as child cognitive ability and non-cognitive ability. In this survey, mothers are asked about whether they spank their children, the reason for spanking, and how frequently they spank. Additionally, interviewers observe and record whether mothers ever spank the children during the course of the survey. Cognitive skills are measured by Peabody Individual Achievement Tests (PIATs) on skills in mathematics, reading recognition, and reading comprehension. Non-cognitive skills are measured by the Behavioral Problem Index (BPI) and its subindices on anxiety, antisocialness, dependence, headstrongness, hyperactivity, and peer conflicts.

In the cross-section I find a negative relationship between spanking and outcomes - a fact consistent with previous research. Children who are spanked are more likely to have behavioral problems and have lower math and verbal test scores. Including mother/family fixed effects lowers these estimated correlations to roughly 10% of the cross-sectional relationship in test scores, although there is still a strong negative relationship between behavioral problems and spanking. Estimates using propensity score matching on the likelihood of spanking also find a consistently negative impact of spanking on child outcomes.

In my preferred structural estimation strategy I use the fact that the relationship between spanking and outcomes has changed dramatically over time. The negative association between spanking and outcomes has doubled from the early years of early years of the NLSY sample to the later years. Simultaneously, the (age-adjusted) probability that any child is spanked fell by 25 percentage points over this time. I argue that this decline is primarily due to changing stigmas about the appropriateness of corporal punishment within the household. If this nationally changing stigma is unrelated to internal household returns to spanking then I can combine the changing selection into spanking (via the stigma) with the changing return to spanking to estimate the treatment effect of spanking. This estimated effect is only for the group of parents who changed their spanking behavior due to the hypothesized change in stigma. As the probability declined over time the estimated correlation between spanking and outcomes got worse. This is evidence that parents who would have spanked in the early years of the sample but not in the later years had a positive impact on their children via spanking.

Using the changing stigma to estimate the effect of spanking for the types of family who changed their behavior over this time period I find that spanking has both positive and negative effects on children in these families. Spanking improves childhood test scores by roughly 10%. At the same time, spanking has a negative impact on childhood behavioral problems, lowering these measures by 10-20% of a standard deviation.

The paper is organized as follows: Section 2 describes the correlational evidence on spanking and childhood outcomes. Section 3 discusses the NLSY data. Section 4 presents reduced form evidence of spanking and outcomes in my sample. Section 5 discusses econometric issues with the empirical relationships and how I attempt to solve them. Section 6 presents results, and section 7 has a discussion of other possible explanations and con-

cludes.

2 Literature review: The negative associations of spanking

Most previous academic research on spanking has been done by pediatricians and child development psychologists and the vast majority find that spanking is negatively associated with a host of childhood outcomes and continuing into adulthood. To name but a few, researchers have found that spanking is associated with increased anti-social behavior, mental health problems, aggressiveness, decreased cognitive ability during childhood and then increased criminal behavior and depression in adulthood. Gershoff (2002) provides a meta-analysis of many of these surveys and confirms the consistent negative relationship. For other examples of this literature please see McCord (1996), Strauss (2001), and Strauss and Kantor (1994).

Although these studies have found a robust cross-sectional relationship between parental spanking and child outcomes all of them possibly suffer from problems in identifying a causal relationship between spanking and these outcomes.[1] It is well known that families where children are spanked are both observably less stable and have lower SES among other characteristics that negatively influence children's outcomes. It is also likely that there are unobservable factors that are correlated with spanking and decreases in child outcomes. One plausible example of such unobservable factors is a genetic tendency towards aggressiveness. If parents are excessively agressive and likely to spank their children then their children may have tendencies towards aggressiveness that are due to genetic heritability and not spanking.

[1]This point is rarely discussed in the pediatric literature. Baumrind et al. (2002) is a notable exception.

Failing to control for these tendencies would lead simple OLS estimates to overstate the role of spanking in increasing child aggressiveness.

3 Data: The NLSY79 Mother-Children Sample

The National Longitudinal Survey of Youth provides a unique opportunity to evaluate the effects of spanking as it contains detailed questions on parental corporal punishment as well as numerous cognitive and non-cognitive ability measures. We match mothers from the 1979 survey with all their children from the Children and Young Adult survey. Children were surveyed biannually from 1986 to 2010.

The children survey sample has four questions about spanking and corporal punishment in the household. Three questions were asked to mothers about whether they spanked. The questions are "Did <respondent mother> have to spank child in the past week?", "How often was child spanked in the past week?", and "Sometimes children get so angry at their parents that they say things like 'I hate you' or swear in a temper tantrum. What actions would you take if this happened?" Additionally, the interviewer recorded whether or not he/she saw the mother spank or strike the child during the course of the interview. I combine all these questions into a simple binary variable that is 1 if any question indicated that the mother spanked the child.[2]

The matched NLSY mother-child data contains detailed information about childhood cognitive and non-cognitive abilities. Children aged 4 to 14 are given Peabody Individual Achievement Tests (PIATs) that measure cognitive skills in mathematics and reading recognition. To measure non-cognitive abilities, the survey calculates a Behavioral Problem Index

[2]See Appendix XXX for further discussion on these variables. Or maybe put in main text.????

(BPI) and the subindices which measure particular problems including antisocial behaviors, anxiety, dependence, headstrongness, hyperactivity, and social problems. Table 1 presents some summary statistics of the NLSY sample over the entire 1986-2010 time period. Please note that a higher score on the BPI is associated with increased problems.

4 Spanking and childhood outcomes: some patterns

Before attempting to estimate a causal impact of spanking I present reduced form evidence between spanking and outcomes in the NLSY children sample where I consistently find a negatively correlation between spanking and child outcomes.

Table 2 presents the average unconditional math, reading, and behavioral problem scores in the sample broken down by whether the child is spanked. We see that children who aren't spanked have consistently and significantly higher achievement measures compared to children who are spanked and that these effects are between 33% - 50% of a standard deviation.

To more carefully estimate a correlational return I run the following regression for child i in family f at time t:

$$y_{ift} = \alpha + \beta \cdot 1\{spanked\}_{it} + \gamma_f X_f + \gamma_i X_i + \varepsilon_{ift}$$

where y is an outcome of interest, X_f are a set of family (mother) characteristics such as Mother's AFQT score, age at first birth, and race, and X_i are child characteristics (age and gender).

Table 3 presents estimates of β for math, reading, and behavioral problem scores. Con-

trolling for family characteristics does decrease the negative relationship between spanking and outcomes. Of course, interpreting the estimates in Table 3 as causal is problematic as the observed relationships may be plausibly due to either reverse causality or selection on unobserved family characteristics.

As an intuitive preview of how one structural method will be used to estimate the causal impact of spanking for some families it is useful to look at how the return to spanking has changed over time. To do that I estimate the correlation between spanking and child outcomes has changed over time. I allow my estimates of this return to vary between the 'Early' years of the sample and the 'Late' years of the sample where 'Early' is defined as the 1986-1996 waves and 'Late' is defined as the 2000-2010 waves.

Table 4 presents estimates of how the spanking-outcome correlation has changed over time, controlling for children's and mother's ages and other demographic characteristics. We see that in all measures, the return to spanking has significantly declined over time. For instance, the negative relationship between math test scores and spanking has more than doubled over time. Reading scores are lower but not not significantly while behavioral problems have also gotten worse.

Concurrently with this worsening in the correlation between spanking and outcomes the actual incidence of spanking has declined substantially. To calculate this I estimate the following probit specification:

$$Prob\{spanked\}_{ift} = \Phi\left(X_{if}\gamma + \lambda_t + \varepsilon_{ift}\right)$$

where λ_t are year fixed effects. Table 7 presents estimates of the average probability of spanking in a given year with the other covariates at their mean value. The age and

demographic adjusted propensity to spanking has fallen by roughly 25 percentage points between 1986 and 2010.[3]

There are several possible reasons why mothers have changed their parenting behavior so drastically over time. Some possibilities include a change in social norms about the acceptability of spanking, a change in return to spanking, and/or a shift in the general ability of parenting ability over time, among others. I will come back to these later but for the moment let me document the shift in social acceptability of spanking over this time period. To do this I look at survey responses from the General Social Survey (GSS) from 1986 to 2008 which asked the following question to respondents: "Do you strongly agree, agree, disagree, or strongly disagree that it is sometimes necessary to discipline a child with a good, hard spanking?" Figure 8 shows how the approval of spanking has changed over this time frame (again controlling for other demographics of respondents). We see that although general social approval is substantially higher in the GSS than my observed spanking rates in the NLSY there has indeed been a significant decline over the past 25 years in the social acceptability of spanking.

5 Avenues to Causality

The previous section established that OLS estimates of β were negative and that these relationships have significantly worsened over time. In this section I model how simple OLS estimates may be biased and discuss a variety of structural models that allow estimation of the return to spanking.

[3] A similar specification that just looks at the change between 'early' and 'late' periods in the survey finds a decline in the spanking propensity of 23 percentage points.

5.1 Modeling Bias

To model the bias in OLS consider a choice model where parents choose whether to spank based on a choice criterion:[4]

$$s = 1\{v > 0\}$$
$$v = \delta Z + \nu$$

so that parents spank if an unobserved latent variable (V) is positive. Z represents observable characteristics to the researcher and ν is unobserved error. It is standard to assume that ν is normal and so the decision to spank can be estimated via a probit.

Child outcomes are modeled as

$$y_1 = \alpha + \beta_i s_i + \gamma X + \varepsilon_1$$
$$y_0 = \alpha + \gamma X + \varepsilon_0$$

where y_1 is the outcome if parents spank and y_0 is the outcome if parents don't. X are other covariates. Simple OLS estimates are unbiased under the assumption that all the above unobserved errors are uncorrelated.

There are many possible parameters that can be estimated from the above choice system. I will focus on two in particular, the Average Treatment Effect (ATE) and Marginal

[4]This framework closely follows Heckman and Navarro (2004)'s presentation of choice models.

Treatment Effect (MTE) which are defined as

$$\begin{aligned}\text{ATE} &= E(y_1 - y_0|X) \\ \text{MTE} &= E(y_1 - y_0|X, v = 0)\end{aligned}$$

It is well-known that using the mean outcomes of children who are spanked and those who are not lead to unbiased estimates of these effects as covariances between the unobserved effects driving child outcomes and the decision for parents to spank their child. Formally, these biases can be written as

$$\begin{aligned}\text{Bias ATE} = &(E(\varepsilon_1|X, s = 1) - E(\varepsilon_1|X)) \\ &- (E(\varepsilon_0|X, s = 0) - E(\varepsilon_0|X)) \\ \text{Bias MTE} = &(E(\varepsilon_1|X, Z, s = 1) - E(\varepsilon_1|X, Z, v = 0)) \\ &- (E(\varepsilon_0|X, Z, s = 0) - E(\varepsilon_0|X, Z, v = 0))\end{aligned}$$

Bias arise due to differences in the expectation of the unobserved error terms due to differences in conditioning sets. Simply put, parents who spank are different than those who do not.

5.2 Mother fixed effects

One possible option to control for differential parental selection into the decision to spank is to control directly for all parental differences with mother fixed effects. Formally, this

assumption is that

$$E(y_1|X, M, s = 1) = E(y_1|X, M, s = 0)$$
$$E(y_0|X, M, s = 1) = E(y_0|X, M, s = 0)$$

where M is a dummy variable for any particular mother. By conditioning on all family level characteristics we have hopefully controlled for any unobserved shocks that influence both child outcomes and mothers' decisions to spank.

While controlling for family level differences is a good first step towards eliminating the bias in OLS it is likely insufficient. Notably, if there are time varying shocks that influence both children's outcomes and the parents' decision to spank then mother fixed effects will not eliminate the bias. For example if a mother loses her job or experiences some other stressful life event it may be the case that she is more likely to harshly discipline her child. But if the stressful event itself affects children's behavioral problems or test scores then estimates using family fixed effects will still be biased.

5.3 Propensity Score Matching

Another approach in controlling for endogenous selection into the decision to spank is matching on propensity score. This assumes that, conditional on the determinants of the selection equation (and a full support on the probability of spanking) then we can identify

$$E(y_1|X, s = 1, p(S|Z)) = E(y_1|X, s = 0, p(S|Z))$$
$$E(y_0|X, s = 1, p(S|Z)) = E(y_0|X, s = 0, p(S|Z))$$

where $p(S|Z)$ is the propensity for a mother-child pair with characteristics Z to have spanked. Further we need to estimate this propensity to spank using a probability model

$$p(S = 1|X) = \Phi(Z'\Gamma)$$

In practice this method amounts to estimating the probability of spanking using a probit model given mother-child pair characteristics Z, calculating the fitted values from this regression and then including these fitted values in an OLS regression of outcomes on spanking.

Propensity score matching suffers from similar problems as mother fixed effects in that if there are unobserved time-varying shocks that affect the propensity to spank as well as child outcomes then matching estimates will also be biased. This is in addition to well-known problems with matching estimators such that increasing information about the propensity score can actually worsen the matching estimator as the common support between spanking and non-spanking disappears. Additionally, matching estimators, like mother fixed effects, assume that average effects are the same as marginal effects.

5.4 Control Functions

The third approach I use in controlling for the endogeneity of the decision to spank is to use control functions which are semi-parametric estimates of the conditional expectations of both outcomes y_1 and y_0. The semi-parametric model is:

$$E(y_1|X, Z, s = 1) = \alpha + \beta + \gamma X + K_1(P(X, Z))$$
$$E(y_1|X, Z, s = 1) = \alpha + \gamma X + K_0(P(X, Z))$$

where $K_1(\cdot)$ and $K_2(\cdot)$ are fourth degree polynomials in the propensity score. This model is only identified if there is a variable in Z that is not in X. To satisfy this exclusion restriction I include time fixed effects in estimating the propensity score. As documented in Section 4 the propensity to spank has changed dramatically over time. This allows consistent estimation of the control functions and recovery of β.

6 Results

Table 9 presents estimates of the return to spanking with and without mother fixed effects. Interestingly, including mother fixed effects drastically lowers the estimated negative return to spanking for cognitive achievement measures. The effect for both math and reading test scores falls to close to zero and becomes insignificant. This is consistent with a strong negative unobserved selection into the decision to spank. Once we control for some of this selection the OLS estimates increase substantially. However, the estimated effect of spanking on behavioral problems is still negative with mother fixed effects.

Table 10 presents estimates of the return to spanking both with OLS and with propensity score matching. Individuals are matched using nearest neighbors conditional on child gender and age as well as mother's race, age at first birth, and AFQT score. The matching estimates are all lower than OLS estimates, again unsurprising evidence for negative selection effects into spanking. Unlike the estimates using mother fixed effects, estimates that control for the propensity to spank remain significant and negative.

Table 11 presents results using control function methods using time as the excluded variable to estimate the control function. Estimates for baseline OLS, the ATE, and the average MTE for mothers who switched due to the changing propensity over time are pre-

sented. We see that the ATE for cognitive scores is substantially lessened compared to OLS estimates although the negative relationship between spanking and behavioral problems persists.

When we look at the marginal treatment effects for mothers who switched the decision to spank over time we see a striking result: spanking appears to have had a positive effect on their children's cognitive scores! I estimate that if these mothers had continued to spank despite the changing stigma against spanking their children would have roughly 10% higher test scores. Intuitively, this result follows from the reduced form patterns discussed earlier: mothers stopped spanking and the relationship between spanking and outcomes worsened. In the early period, these mothers were contributing 'positively' to the reduced form association and by removing them this association substantially declined.

Although spanking may have had a positive effect for these children's cognitive scores the story is not as surprising for behavioral problems. The estimated mean marginal treatment effect for switchers on behavioral problems is still significantly negative. Indeed, for all approaches, from the reduced form OLS to the more structural estimations there has been a consistently negative relationship between spanking and behavioral problems.

7 Discussion + Future Research

In this paper I have documented the negative cross-selection relationship between spanking and child outcomes using the NLSY79 matched mother-children data. Spanking is negatively related with math and reading test scores as well as child behavioral problems. However, when I control for likely endogenous selection into spanking the evidence against spanking becomes weaker, particularly for test scores. Estimates that control for family

level unobservables and the propensity to spank find a limited effect of spanking on test scores. Using the fact that many mothers have stopped spanking over time and the return to spanking has simultaneously declined I can go further and estimate that, for these mothers, the treatment effect of spanking on test scores was actually *positive*.

It's worthwhile to consider what other possible stories could be explaining these results. There are two other plausible explanations: First, the return to spanking may itself be unstable over time. Second, spanking may be an indicator of other parental decisions. Let me discuss these possibilities and how I may be able to overcome them in turn.

The first alternative explanation is that the relationship between spanking and childhood outcomes could have changed substantially over the past 30 years. If spanking is now worse for children then in the mid 1980s we would see exactly the same relationship in the data: fewer parents would spank and the return would significantly decline. However,

Another alternative explanation for my results is that it is not spanking *per se* that matters for children but rather general parenting skills.The general ability of parents (or knowledge of parenting) may have improved over this time span and I could be incorrectly assigning that to spanking exclusively. If parents have become better parents over time then they may spank less and change other attributes of their parenting. Without controlling for these other attributes of parenting I may be overstating the importance of spanking in any changes over time.

There are two possible I plan to address this possibility using the fact that the NLSY79 children data includes many other measures of parental investment and discipline. First, I can directly include these other measures of parenting ability in my estimation which should hopefully help me isolate the effect of just spanking on outcomes. Second, I can perform

placebo exercises with these other measures of parenting decisions. If changes to general parenting skill are driving my results then we would expect to see a similar pattern in the return to other discipline measures.

References

Diana Baumrind, Robert Lazelere, and Philip Cowan. Ordinary Physical Punishment: Is It Harmful? Comment on Gershoff (2002). *Psychological Bulletin*, 128(4):580–589, July 2002.

Elizabeth Gershoff. Corporal Punishment By Parents and Associated Child Behaviors and Experiences: A Meta-Analysis and Theoretical Review. *Psychological Bulletin*, 128(4): 539–579, July 2002.

James Heckman and Salvador Navarro. Using Matching, Instrumental Variables, and Control Functions To Estimate Economic Choice Models. *Review of Economic Statistics*, 86 (1):30–57, February 2004.

Joan McCord. Unintended Consequences of Punishment. *Pediatrics*, 29, 1996.

Murray A. Strauss. New Evidence for the Benefits of Never Spanking. *Society*, 38:52–60, 2001.

Murray A. Strauss and Glenda Kantor. Corporal Punishment of Adolescents By Parents: A Risk Factor in the Epidemiology of Depression. *Adolescence*, 29, 1994.

Tables

Table 1: NLSY Matched Mother-Children Summary Statistics

	Mean	Std. Dev.
Black	0.28	0.45
Hispanic	0.19	0.39
Female	0.49	0.50
Age	7.59	4.12
Math (percentile)	38.1	27.5
Reading Recognition (percentile)	41.4	29.4
Behavioral Problem Index (percentile)	61.8	29.4
% Spanked	0.31	0.46
Observations	14951	

Data from Children of the NLSY79 sample, 1986-2010. Age is only reported for children with valid test scores and spanking variables.
$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 2: NLSY Matched Mother-Children Summary Statistics

	Not spanked	Spanked	Difference
% Black	0.26	0.36	-0.101***
% Hispanic	0.21	0.19	0.0166***
% Female	0.51	0.45	0.0533***
Age	8.40	5.59	2.810***
Mother's AFQT	-0.33	-0.60	0.273***
Math (percentile)	41.7	27.5	14.22***
Reading Recognition (percentile)	45.1	29.7	15.46***
Behavioral Problems (percentile)	61.1	72.4	-11.31***

Data from Children of the NLSY79 sample, 1986-2010. Age is only reported for children with valid test scores and spanking variables.
$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 3: Conditional correlations of spanking on outcomes

	(1)	(2)
Math	-13.947***	-2.956***
	(0.502)	(0.519)
Reading Recognition	-13.533***	-2.738***
	(0.584)	(0.616)
Behavioral Problems	9.699***	9.640***
	(0.584)	(0.634)
Mother characteristics	No	Yes
N	14620	12823

Data from Children of the NLSY79 sample, 1986-2010. Each cell is the estimate from a separate regression. Mother characteristics are age at first birth, race, and AFQT score. All specifications include controls for child's gender and age. Robust standard errors are clustered at the mother level.
$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 4: How the returns have changed over time

Math :		
	Spanking	-2.193***
		(0.563)
	Late (2000-2010 dummy)	2.972***
		(0.515)
	Spanking X Late	-2.287**
		(1.000)
Reading Recognition:		
	Spanking	-2.680***
		(0.668)
	Late (2000-2010 dummy)	1.569***
		(0.524)
	Spanking X Late	-1.516
		(1.115)
Behavioral Problems:		
	Spanking	8.555***
		(0.708)
	Late (2000-2010 dummy)	-3.844***
		(0.621)
	Spanking X Late	3.272**
		(1.170)
	N	9908

Data from Children of the NLSY79 sample, 1986-2010. Each panel is the estimate from a separate regression for a different outcome variable. Other controls included are mother's age at first birth, race, and AFQT score as well as child's gender and age. Robust standard errors are clustered at the mother level.

$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 5: Summary statistics of mothers/children who spank over time

	Early (1986-1996)	Late (2000-2010)
% Black	0.36	0.37
	(0.48)	(0.48)
% Hispanic	0.19	0.20
	(0.39)	(0.40)
% Female	0.46	0.45
	(0.50)	(0.40)
Average child age	5.37	6.74
	(3.07)	(3.50)
Mother's AFQT	0.64	0.73
	(0.43)	(0.49)
Observations	48418	57573

Data from Children of the NLSY79 sample, 1986-2010.

Table 6: Determinants of spanking propensity over time

	(1)	(2)	(3)
	Whole sample	1986-1996	2000-2010
Child is female dummy	-0.174***	-0.188***	-0.130***
	(0.0185)	(0.0205)	(0.0406)
Black dummy	0.239***	0.209***	0.341***
	(0.0327)	(0.0352)	(0.0632)
Hispanic dummy	-0.106***	-0.115***	0.0133
	(0.0368)	(0.0392)	(0.0675)
AFQT	-0.203***	-0.159***	-0.175***
	(0.0150)	(0.0165)	(0.0288)
N	49132	32072	12597

Data from Children of the NLSY79 sample, 1986-2010. Results are marginal effects from probit regressions. Robust standard errors are clustered at the mother level.
$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 7: NLSY: Adjusted spanking propensity over time

1986	0.39***
1988	0.41***
1990	0.38***
1992	0.35***
1994	0.31***
1996	0.27***
1998	0.25***
2000	0.19***
2002	0.21***
2004	0.19***
2006	0.17***
2008	0.15***
2010	0.15***

Data from Children of the NLSY79 sample, 1986-2010. Estimated probability of spanking controlling for mother and child demographics via probit and fit at mean values.
$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 8: GSS: Approval of spanking over time

1986	0.84***
1988	0.81***
1989	0.79***
1990	0.80***
1991	0.75***
1993	0.75***
1994	0.75***
1996	0.74***
1998	0.76***
2000	0.72***
2002	0.71***
2004	0.72***
2006	0.70***
2008	0.70***

Data from General Social Survey, 1986-2008. Estimated probability of spanking controlling for individual demographics via probit and fit at mean values.
$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 9: Fixed effects estimates of the return to spanking

	(1)	(2)
Math	-2.956***	-0.477
	(0.519)	(0.368)
Reading Recognition	-2.738***	-0.282
	(0.616)	(0.379)
Behavioral Problems	9.640***	4.375***
	(0.634)	(0.311)
Mother characteristics	Yes	No
Mother fixed effects	No	Yes

Data from Children of the NLSY79 sample, 1986-2010. Each cell is the estimate from a separate regression for a different outcome. Mother characteristics are age at first birth, race, and AFQT score. Child characteristics are gender and age. Robust standard errors are clustered at the mother level.

$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 10: Matching estimates of the return to spanking

	Unmatched (OLS)	Matching
Math	-2.956***	-1.528***
	(0.519)	(0.353)
Reading Recognition	-2.738***	-1.971***
	(0.616)	(0.333)
Behavioral Problems	9.640***	8.016***
	(0.634)	(0.328)

Data from Children of the NLSY79 sample, 1986-2010. Each cell is the estimate from a separate regression for a different outcome. OLS estimates control mother's age at first birth, race, and AFQT score as well as child's gender and age. Matching estimates are matched on these variables. Standard errors for OLS estimates are robust and clustered at the mother level. Standard errors for matching estimates are bootstrapped with 200 iterations.

$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 11: Control function estimates of the return to spanking

	OLS	ATE	Mean MTE for switchers
Math	-2.956***	-1.116	2.758***
	(0.519)	(0.616)	(0.851)
Reading Recognition	-2.738***	-1.608*	3.010***
	(0.616)	(0.897)	(0.809)
Behavioral Problems	9.640***	12.896***	7.778***
	(0.634)	(0.939)	(1.220)

Data from Children of the NLSY79 sample, 1986-2010. Each cell is the estimate from a separate estimation for a different outcome. OLS estimates control mother's age at first birth, race, and AFQT score as well as child's gender and age. Standard errors for OLS estimates are robust and clustered at the mother level. Standard errors for control function estimates are bootstrapped with 200 iterations.

$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.