Parental Guidance and Supervised Learning by Lizzeri and Siniscalchi 2008 QJE

James J. Heckman and Hanzhe Zhang

Econ 350, Winter 2023

THE UNIVERSITY OF CHICAGO

Abstract

The paper:

1. Proposes a simple theoretical model of *supervised learning*
   - Captures a trade-off between
     (1) Sheltering the child from the consequences of his mistakes and
     (2) Allowing him to learn from experience
   - Characterizes the optimal parenting policy and its comparative-statics properties
2. Uses the model to interpret a number of empirical phenomena relevant to the nature-nurture debate (vs *behavioral genetics*)
   - Interpretations of twin studies
   - Heritability measures

Section I. Introduction

- The nature-nurture debate has been one of the most controversial in the social sciences.
- Behavioral genetics literature provides evidence that parents have little or no effect on a variety of measures of the personality of their children.
- Common perceptions and practice suggest parents do matter.
- The motivating question of the paper: How can we reconcile the evidence and common perceptions?

Behavioral Genetics versus Supervised Learning

- *Behavioral genetics* isolates the effects of genes through several complementary approaches, especially the ACE model.
  - Compare twins raised together by their biological parents with twins reared apart by different adopted families
  - ACE model (essentially a statistical model): twins reared together are as similar as twins reared apart; some studies even show that twins reared together are even less similar.
- *Supervised learning* provides a different interpretation of the same data.
  - Children's characteristics evolve through interactions with the environment and their parents.
  - Every parent faces a trade-off between
    - *sheltering* the child from the consequences of mistakes
    - allowing him to *learn from experience*

THE UNIVERSITY OF CHICAGO

Model Sketch

- The child lives with the parent for T periods and then lives alone for $L - T > 0$ periods.
- Each period, the child performs some task and learns by doing.
- At the end of the period, he receives a signal about the quality of his performance.
- Learning is costly: the worse his performance, the lower the payment in the current period.
- The parent has better information than the child and can help him improve his current performance.
- Helps comes at a cost, as it distorts the child's signal about his performance: making mistake is less severe, but it takes longer for the child to learn and improve his performance in subsequent periods.

Main Results

1. Optimal parenting is *partial sheltering*.
2. The degree of the optimal sheltering is time-varying and depends on the child's bias.
   - Sheltering is greater the more the parent believes that she needs to compensate for her child's perceived inadequacy.
   - In other words, parental intervention *substitutes* for the child's natural ability to perform the task, as perceived by the parent.
   - The more the child and the parent have different priors, the more the parent intervenes.

Main Results

3. Adoptive parents provide more sheltering on average than biological parents because adoptive parents are less similar to their children.

THE UNIVERSITY OF CHICAGO

Literature

This paper's focus is to formally model *supervised learning*, in particular the relationship between parental behavior and its effects on children's learning process, which is almost absent in the literature.

1. *Learning* has been extensively integrated in theoretical models by economists, statisticians, and psychologists; but these studies abstract from the fact that learning takes place under the supervision of parents, caregivers, teachers, advisors, and other experts for a considerable fraction of an individual's life.
   - **?**, **?**, **?**, **?**.
2. *Behavioral genetics* literature typically focuses on personality traits and measures of cognitive achievements and does not typically explore outcomes such as educational attainment or earnings.

Section II. Benchmark Model

- Two agents: a child (he) and his parent (she).
- The child lives with the parent for T periods and then lives alone for $L - T$ periods.
- The correct way to perform a task is M on average.
- The actual correct way is $X_1, \cdots, X_L$, with mean M and precision $p_X = 1/\sigma_X^2$.
- The parent's action is a_t and the child's action is b_t.
- If the correct way to perform the task is X_t, then the child incurs a loss of

$$(X_t + a_t - b_t)^2.$$

Timing

1. A problem arises at the beginning of a period, and the correct answer is distributed according to $N(M, 1/p_X)$.
2. The child has a prior belief $N(M_0, 1/p_0)$ about M, and makes a decision b_t to confront the problem.
3. Observing the child's b_t, the parent chooses a corrective action a_t.
4. The child observes $X_t + a_t$ but not separate components X_t and a_t,

$$\mathbb{E}[X_t | X_1, \cdots, X_{t-1}] = \frac{p_0 M_0 + p_X \sum_{s=1}^{t-1} X_s}{p_0 + (t-1)p_X}.$$

   When parents choose $\{a_t\}$, the child's posterior is

$$\begin{aligned} M_t^a &\equiv \mathbb{E}[X_t | X_1 = x_1 + a_1, \cdots, X_{t-1} = x_{t-1} + a_{t-1}] \\ &= \frac{p_0 M_0 + p_X \sum_{s=1}^{t-1} (x_s + a_s)}{p_0 + (t-1)p_X}. \end{aligned}$$

5. Common objective

Parenting Styles

The expected time-t penalty conditional on the parent's information is

$$\begin{aligned} & \mathbb{E}\left[(X_t + a_t - M^a_{t-1})^2 | X_1, \cdots, X_{t-1}, M\right] \\ = & \underbrace{p_X^{-1}}_{\text{penalty due to uncertainty}} + \underbrace{(a_t - (M^a_{t-1} - M))^2}_{\text{penalty due to parental guidance}} \end{aligned}$$

1. *Letting Go*: $a_t = 0$.
2. *Full Sheltering* ("Italian Mom"): $a_t = M^a_{t-1} - M$
   - This policy maximizes myopic payoff.
3. *Boot Camp*: $a_t = -(M^a_{t-1} - M)[p_0 + (t-1)p_X]/p_X$.
   - The child's posterior M^a_t will be equal to M on average at the end of time $t-1$.

Main Theorem

Theorem 1

The optimal action of the parent at time t is a linear function of the child's bias: $a_t^* = \gamma_t(M_{t-1}^a - M)$. *The intensity of* γ_t

- *is time-varying but deterministic and lies between zero and one,*
- *is decreasing in* δ *and* L, *and*
- *is a weighted average of the intensities of intervention for the Full Sheltering and Boot Camp policies,*

$$\gamma_t = \mu_t \gamma_t^{FS} + (1 - \mu_t)\gamma_t^{BC}.$$

Proof Sketch

Backward induction

1. without parental intervention: period L first, then $L-1$ to $T+1$.
2. with parental intervention: periods T to 1.

Remarks

- **Dynamics of intensity of intervention**: Except for possibly the first few periods, the weight μ_t placed on the Full Sheltering intensity γ_t^{FS} will be decreasing in t for δ relatively low and increasing in t for δ relatively high.
  - If δ is low, the parent places more weight on minimizing current losses than on reducing the child's bias, so a high but eventually decreasing level of sheltering is optimal.
- **Ease of learning**: If p_0 is high or if p_X is low, the child places more weight on his prior M_0 than on observations X_t, and hence learning is slow. Two effects of the ease of learning on γ_t:
  1. straightforward "inertia" effect: if learning is slow, the child benefits less from a reduced bias; an incentive for more sheltering.
  2. "continuation value" effect: if learning is hard, this will be the case not just today but also in the future; an incentive for less sheltering.
- What about the parents?

Relations to Developmental Psychology Literature

- By sheltering the child, the parent helps him perform better in life; this generates a trade-off between sheltering and allowing the child to learn from his mistakes.
- Optimal sheltering is only partial.
- Optimal parenting is responsive to the characteristics of the child.
- The degree of sheltering depends on the difference between the child's characteristics and her own. This feature of the model highlights the need to consider the child-parent pair as an interactive system and not as independent actors.

THE UNIVERSITY OF CHICAGO

Section III. Population Model

- Parents and children are heterogeneous.
- Each child lives with the parent for one period and alone for another period.
- Every child has a prior belief $N(M_0, p_0)$ and every parent has an *initial* belief $N(Z_0, p_Z)$ of M.
- Suppose there is a distribution of M_0 and a distribution of Z_0.
  - Correlations between M_0 and Z_0 reflect genetic relatedness of the individuals under considerations.
  - The initial means of parents, or a parent's initial mean and her adopted child's prior mean are uncorrelated.
  - The prior means of monozygotic twins are perfectly correlated.
- Optimal parental intervention: $a_1 = \gamma(M_0 - Z_0)$.
  - Sheltering will be greater when the parent and the child are a priori more different.
  - Adoptive parents will shelter more than biological parents.

THE UNIVERSITY OF CHICAGO

ACE Model

Decompose the observable characteristic, *phenotype*, into three factors,

$$\underbrace{P}_{\text{phenotype}} = \underbrace{A}_{\text{genotype}} + \underbrace{C}_{\text{shared environment}} + \underbrace{E}_{\text{non-shared environment}}$$

1. *genotype*: the individual's genetic endowment.
2. *shared environment*: factors that affect siblings in the same family.
3. *non-shared environment*: idiosyncratic elements of the phenotype.

ACE model typically assumes that the factors affecting a given individual's phenotype are *mutually independent*:

$$\text{var}[P] = \text{var}[A] + \text{var}[C] + \text{var}[E].$$

Current model mapped to ACE model:

$$\underbrace{M_1}_{P} = \underbrace{\frac{p_0}{p_0 + p_X} M_0}_{A} + \underbrace{\gamma \frac{p_X}{p_0 + p_X} (M_0 - Z_0)}_{C} + \underbrace{\frac{p_X}{p_0 + p_X}}_{E}$$

A and E are genetic, and C captures parenting.

THE UNIVERSITY OF CHICAGO

Interpreting Twin Studies

- Phenotypic correlation between two members of a twin pair:
  - r^{MZT}: monozygotic (identical, MZ) twins reared together.
  - r^{MZA}: monozygotic twins reared apart.
  - r^{DZT}: dizygotic (fraternal, DZ) twins reared together.
- Empirical findings:
  1. $r^{\text{MZA}} > 0.90 \times r^{\text{MZT}}$. In some studies, $r^{\text{MZA}} > r^{\text{MZT}}$.
  2. $r^{\text{MZT}} - r^{\text{MZA}} < r^{\text{MZT}} - r^{\text{DZT}}$. (i.e. $r^{\text{MZA}} > r^{\text{DZT}}$).

Rationalizing Finding 1

$r^{\text{MZA}} > 0.90 \times r^{\text{MZT}}$. In some studies, $r^{\text{MZA}} > r^{\text{MZT}}$.

- *Behavioral genetics*: if genotype and environmental factors are independent, the phenotypic correlation between twins reared together is determined by their common genetic endowment (A) and shared environment (C); on the other hand, for twins reared apart, phenotypic correlation can only be driven by commonality in genetic endowment. If $r^{\text{MZA}} > r^{\text{MZT}}$, then the contribution of common rearing to phenotypic correlation is small.

- *Supervised learning*: Let $p \equiv p_X/(p_0 + p_X)$; v_0, v_{Z_0} population variances of M_0 and Z_0; v_1 and v_{1a} variances of M_1 for children reared by biological and adoptive parents, respectively; $v_X = 1/p_X$ variance of X_t. $r^{\text{MZT}} - r^{\text{MZA}} =$

$$\underbrace{(p\gamma_1)^2 \frac{v_{Z_0}}{v_1}}_{\text{common rearing}} \underbrace{- r_0 \left\{ 2 \frac{v_X + \gamma_1^2 v_{Z_0}}{v_{1a}} [1 - p + p\gamma_1] p^3 \gamma_1 \frac{\sqrt{v_0 v_{Z_0}}}{v_1} \right\}}_{\text{compensating effect due to parental guidance}}$$

- $Corr(M_0, Z_0)$

Intuition for Finding 1 (Compensating Effect)

- Intuition for the compensating effect:
  1. Each child's genetic endowment is correlated with his biological parent but not with that of adoptive parent; $|M_0 - Z_0|$ is likely to be larger for adopted children. Since the parental intervention takes the form of $\gamma_1(M_0 - Z_0)$, adoptive parents provide more sheltering; the parent compensates for the child's perceived shortcomings by providing more sheltering (consistent with empirical evidence).

2. Non-shared environmental influences have fewer opportunities to affect the adopted twins' posterior. On the other hand, because their priors are positively correlated, twins raised by their biological parents will be subject to less sheltering; non-shared environment will have a greater role, and they will end up being less similar to one another.

- Reinterpret ACE model: differential sheltering by biological and adoptive parents implies that, despite being reared apart, adopted twins are in fact subject to a shared environment influence - adoption itself, which compensates for the lack of the direct common-rearing effect.

Intuition for Finding 2

$r^{\text{MZT}} - r^{\text{MZA}} < r^{\text{MZT}} - r^{\text{DZT}}$. (i.e. $r^{\text{MZA}} > r^{\text{DZT}}$).

- *Behavioral genetics*: MZ and DZ twins reared together share the same rearing environment; taking the difference between their phenotypic correlations "cancels out" the common effects of parenting, and so any remaining difference must be due to the fact that MZ twins have the same genetic endowment, whereas DZ twins share only 50%.
- *Supervised learning*: The compensating effect is the key effect that enables the model of supervised learning to generate the correlational pattern.

Numerical Results

For a wide range of parameters, the two findings hold.

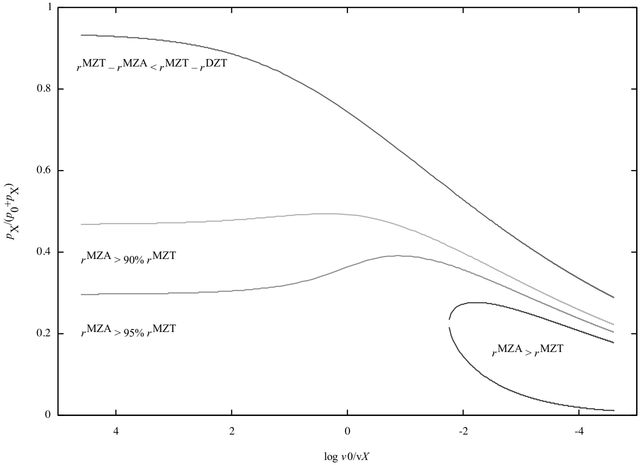

FIGURE II
Twins Reared Together vs. Twins Reared Apart

Auxiliary Predictions of the Model

1. **Differences in Phenotypic Variance**
   - The variance of the posterior mean is larger for adopted children.
     - Adopted children are less exposed to non-shared experiences, and hence don't learn the correct way to perform the task at hand as fast as children reared by their biological parents.
   - cannot be obtained in ACE model because phenotypic variances are the same in all subpopulations.

THE UNIVERSITY OF CHICAGO

2. **Correlation of MZ Twins Adopted Together**
   - Highest correlation arises for MZ twins reared together in an adoptive family.
   - These individuals share the same parents.
   - They are sheltered more than children reared by biological families.
   - These two forces push in the same direction and lead to an increased correlation for twins reared in the same adoptive family.
   - not tested.

THE UNIVERSITY OF CHICAGO

Heritability Measures

- In the ACE model, heritability is measured as $h^2 \equiv \text{var}[A]/\text{var}[P]$ (e.g. heritability of IQ is 69% - 78%). Two standard measures:
  1. the correlation in the measured trait for MZ twins reared apart
  2. difference between the correlations in a measured trait of interest for MZ and DZ twins reared by their biological parents, multiplied by two
- The supervised learning model shows that standard measures of heritability underestimate the effects of parenting.

THE UNIVERSITY OF CHICAGO

Idea 1: Perry Program

- Increase in cognitive skills improvement fades out (caught up by teacher's instructions).
- Increase in non-cognitive skills is persistent.
- The short-run impact of increased cognitive skill makes a long-run impact on non-cognitive skills.
    - The child is exploring the best study strategies; a boosted score because of advantage in cognitive intervention confirms for him that his study strategy is good. He builds his confidence about his study strategies and about his own IQ, which persist even after the real effect goes away.
- It seems that this channel is not explicitly modeled.
- Another implication is that teachers are not teaching enough non-cognitive skills.

Idea 2: Two Parents

- Most papers in the literature have one parent and one child.
- Father and mother play different roles in child's development.
- It is a theoretically and empirically interesting question to see what roles parents play.
- It helps to understand what's missing for child's development in a single-parenthood (in divorced family or out-of-wedlock birth).