Local Distortions in Parental Beliefs over Child Skill

Josh Kinsler, University of Georgia

Ronni Pavan, University of Rochester

September 2018

Abstract

Parental beliefs over child skill have been shown to have a significant impact on investment behavior. In this paper, we examine how parental beliefs form and evolve, focusing on the role of the local environment. Using data from the ECLS-K, we first show that parental beliefs about a child's skill relative to *children of the same age* is distorted by a child's skill relative to *children in the same school*. Parents of children attending schools with low (high) average skills tend to believe their child is higher (lower) in the overall skill distribution than they actually are. Motivated by our descriptive findings, we develop and estimate a formal model of parental learning that is capable of rationalizing these patterns. While we find that parents receive useful information, they tend to misinterpret signals of child and school skill, generating local belief distortions. Using the model, we then investigate how parental beliefs would change if the information set and/or parent perceptions were altered. Finally, we relate parental beliefs and investment, providing insight on how local distortions may impact the skill distribution.

1 Introduction

Across households, there is substantial variation in the productive activities, experiences, and policies that govern a young child's life. For example, some children take music lessons while other children play organized sports. While preferences and constraints guide the decisions parents make, recent evidence suggests that parental beliefs also play an important role. Surveys from developed and developing countries indicate that heterogeneity in parental investment choices is strongly related to heterogeneity in beliefs over objects such as child skill and the productivity of various skill inputs.[1] Because parental investment is a significant input in the production of adolescent skills, learning how parents form beliefs can improve our understanding of child skill development.[2]

In this paper, we take an important step forward in this direction by analyzing how parents form beliefs about the skill of their children and how these beliefs evolve over time. An essential feature of our analysis is to allow the local environment, as measured by schools, to affect the perceptions of parents. Parents observe their children in isolation and in settings with other children. Moreover, parents receive feedback about their children from teachers, coaches, and counselors who tend to evaluate many children from the same local environment. Thus, it is natural to expect that the surrounding community will influence parent perceptions. Because communities and schools tend to be segregated by income, education, and race, local distortions in parental beliefs can have important aggregate consequences for skill development.

Using data from the Early Childhood Longitudinal Study-Kindergarten Class of 1999 (ECLS-K), we first illustrate that the local environment significantly impacts parental perceptions about their children. Three features of the ECLS-K make this possible. First, child skill is measured in every wave of the ECLS-K through survey based skill assessments. Second, parents are asked to rate the skill of their child relative to all similarly aged children. It is important to note that parents never observe the ECLS-K assessment results and thus cannot base their beliefs on these measures. Third, the sampling unit of the ECLS-K is a

[1] See, for example, Cunha *et al.* (2013), Dizon-Ross (2013), Attanasio *et al.* (2018), Boneva & Rauh (2017), and Boneva & Rauh (2018).

[2] Cunha *et al.* (2010) find that measured parental investment accounts for 15% of the variation in educational attainment. Carneiro & Heckman (2003), Cunha & Heckman (2008), and Todd & Wolpin (2007) provide additional supporting evidence on the importance of early childhood investment. Heckman & Mosso (2014) provide a nice summary of the literature.

school, allowing us to generate a noisy measure of local average skill. We find that parental beliefs about a child's skill relative to children of the same age is negatively influenced by the average skill of children *in the same school*. In other words, parents of children in low skill schools tend to think their child is more highly skilled than they actually are. The opposite pattern occurs at high skill schools. We show that this finding is robust to concerns related to misinterpretation of the survey question, measurement error, and unobserved differences across schools.

The descriptive evidence suggests that parents have locally distorted beliefs about the skill of their child relative to the broader population. Moreover, the local distortion in beliefs grows between the fall of kindergarten and the spring of 3rd grade in a statistically significant manner. On the surface, this seems to suggest that parents are not learning about the true skill of their child as they age. However, parental beliefs also become more stable over time, consistent with learning. To reconcile these features of the data we develop a formal model of parental learning and skill evolution.

In the model, parents are uncertain about the skill of their child and the average skill level in their child's school. Each period parents receive signals about these objects, though parents may misinterpret the content of the signals. As an example, parents may receive a report card from the child's teacher and believe it to be a signal of the child's absolute skill when in reality it is a measure of the child's skill relative to her classmates. Child and school average skills evolve over time according to a skill transition equation. We as the econometrician do not observe the signals parents receive, nor the skill level of the children. Instead, we observe parental beliefs and the noisy skill measures generated as part of the ECLS-K survey. The goal of the model is to understand the structure of the signals parents receive, how parents interpret the signals, and how beliefs would change if we were able to alter the information available. We show that the learning model is identified with the available data and then estimate the model by maximum likelihood using a generalized Kalman filter algorithm.

Consistent with the descriptive evidence, estimates from the model indicate that parents misinterpret the information available to them. While parents believe they receive a dedicated signal of child skill, they actually receive a signal that is contaminated by the school average skill level. This leads to the local distortion in beliefs observed in the descriptive analysis. To put the size of the local distortion in context, consider the set of schools that have initial average math test scores at least one standard deviation below the overall mean. Our estimated model predicts that in 3rd grade, 27% of parents in these schools believe their child is above average. If we alter the parameters of the model such that parents interpret the available signals correctly, only 13% of the same 3rd grade parents believe their child is above average, a decline of over 50%. In contrast, providing parents with additional noisy test score measures but otherwise leaving the signal structure unchanged does little to alter parental belief distortions.

While the model of parental learning captures interesting and important patterns in the data, key questions remain. Is our measure of parental beliefs related to investment? If so, do local distortions tend to expand or contract investment gaps across schools? In the final section of the paper we address these questions. We show that parents who believe their child is above average relative to similarly aged children are less likely to exhibit compensatory investment behaviors, such as helping with homework or hiring a tutor. However, parental beliefs are unrelated to most other types of investments that have been shown to be productive, such as reading to a child or enrolling a child in music class. These patterns suggest that the local distortions in parental beliefs may work to exacerbate skill gaps across high and low skill schools. Parents of children in schools with low (high) average skill levels believe their child is doing better (worse) relative to the overall average than they actually are, leading to less (more) compensatory investment.

Our findings contribute to the recent literature on parental investment and beliefs. Cunha *et al.* (2013) is an early and influential paper in this area. The authors elicit subjective expectations about the elasticity of child development with respect to investment from a sample of socioeconomically disadvantaged, pregnant African American women. If the median mother in the survey were given objective elasticities, cognitive skills are predicted to increase between 1% and 5%. Following on the heels of this work is a series of papers that explore parental beliefs about the productivity of different types of skill investments (Attanasio *et al.* (2018)), the optimal timing of investment (Boneva & Rauh (2018)), and the productivity of health investments (Biroli *et al.* (2018)). All of these papers elicit parental beliefs at a point in time, analyze these beliefs, and relate them to contemporaneous investment behavior. Our contribution is instead to model parental belief formation. If beliefs are critically important for parental investment behavior, then it is imperative to understand the factors that influence these beliefs.

The paper that is closest in spirit to our own is Dizon-Ross (2013). Using data from a field experiment in Malawi, Dizon-Ross (2013) finds that parents' perceptions of their children's recent achievement diverges substantially from children's true recent achievement, with an average gap between the two being a full standard deviation. The source of parental distortions in the context of Malawi is slightly different than in our setting. Parents in Malawi are supposed to receive report cards with absolute test scores and the corresponding grade on the standard Malawian scale. However, parents either fail to receive these reports or are unable to understand them. Importantly, providing parents with accurate information about child achievement causes them to alter their allocation of educational investments.

A primary motivation for studying parental beliefs in both the developed and developing world is to understand what drives persistent differences in parental investment behavior across socioeconomic groups. Standard models of parental investment typically focus on the impact of credit constraints or the tradeoff between goods and time investments in children.[3] More recently, researchers have considered how large a role heterogeneity in preferences and beliefs might play.[4] Our paper adds to this literature by illustrating how local environments can contribute to belief heterogeneity.

Finally, our work is related to the broader education literature on subjective beliefs and learning. Since Manski (2004), researchers have augmented education investment models with expectations data in an effort to separately identify preferences from beliefs.[5] Other work has explicitly modeled the learning process concurrent with own educational investment choices.[6] Here we focus on parental learning about child skill, but leave the dynamic

[3]Caucutt & Lochner (2012) and Cunha (2013) estimate dynamic models of investment focusing on the role of credit constraints. These papers build on earlier work by Becker & Tomes (1986). Boca *et al.* (2014) and Bernal (2008) estimate dynamic models of investment focusing on the labor supply and time allocation problem facing parents.

[4]Caucutt *et al.* (2017) explore theoretically whether uncertainty and parental bias can help explain key stylized facts regarding early investment, parental income, and achievement. Cunha (2015) utilizes heterogeneity in elicited beliefs and preferences to understand investment gaps.

[5]Recent examples include Attanasio & Kaufman (2014), Stinebrickner & Stinebrickner (2014b), Wiswall & Zafar (2015), Wiswall & Zafar (2019), and Arcidiacono *et al.* (2017).

[6]Stinebrickner & Stinebrickner (2014a) and Arcidiacono *et al.* (2016) incorporate ability learning in a model of college dropout, while Arcidiacono (2004) models learning in the context of major choice.

investment problem parents face for future work.

The remainder of the paper is as follows. Section 2 describes the ECLS-K data in detail. In Section 3 we illustrate that parental beliefs are locally distorted. Sections 4 and 5 present the parental learning model and results. The link between beliefs and investment is investigated in Section 6. Section 7 concludes.

2 Description of ECLS-K Data

We use the Early Childhood Longitudinal Study, Kindergarten Class of 1999 (ECLS-K) to explore how local environments can impact the formation of parental beliefs. The ECLS-K is a longitudinal study that surveys a nationally representative sample of parents, children, teachers, and school administrators in the fall of kindergarten, spring of kindergarten, 1st, 3rd, 5th, and 8th grades. 21,409 children distributed across 1,018 schools are included in the initial kindergarten sample. Information about a child's home, school, and classroom environments is collected. We focus our analysis on data collected prior to 5th grade.[7]

In each round of the survey, student math and reading skills are evaluated. The ECLS-K assesses skills that are typically taught and developmentally important, and the assessment frameworks are derived from national and state standards. The cognitive assessments are two-stage adaptive tests; all children begin a subject area test with a routing test, which is then followed by a second-stage form. The two-stage, adaptive assessment format helps ensure that children are tested with a set of items most appropriate for their level of achievement and minimized the potential for floor and ceiling effects. For each wave of the survey, we standardize the Item Response Theory Scale Scores from the reading and math assessments and utilize these as unbiased measures of child skill.[8] Note that parents never

[7] We restrict the sample to earlier grades for a few reasons. First, attrition in the ECLS-K is considerable, an issue we discuss further below. Second, a key variable, how parents believe their child compares to their classmates is not available beginning in fifth grade. Third, limiting the number of periods facilitates estimation of the parental learning model. Finally, since standardized testing does not typically begin until the end of third grade, the timeframe we consider is one where parents are less likely to be informed. Note, however, that even in fifth grade beliefs about overall skill are influenced by measures of local skill.

[8] Item Response Theory uses the pattern of right, wrong, and omitted responses to the items actually administered in a test and the difficulty, discriminating ability, and "guess-ability" of each item to place each child on a continuous ability scale. The items in the routing test, plus a core set of items shared among the different second stage forms, made it possible to establish a common scale. It is then possible to estimate the score the child would have achieved if all of the items in all of the test forms had been administered.

observe the ECLS-K scores and thus cannot use them to learn about the skill of their child.

Parental beliefs about a child's skill relative to children of the same age are elicited in the fall of kindergarten and in the spring of 1st and 3rd grade. The precise wording of the question is as follows: "Does your child learn, think, and solve problems better, as well, slightly less well, or much less well than other children his/her age?" In the spring of 1st and 3rd grade, parents are also asked to compare the math and reading skills of their child to the math and reading skills of the other children in their child's class. Here parents are asked, "Compared to other children in your child's class, how well do you think he/she is doing in school this spring in math? Do you think he/she is doing much worse, a little worse, about the same, a little better, or much better?" A similar question is asked for reading. These classroom based questions are useful for demonstrating that parents understand the difference between local and global comparisons and utilize different reference points to assess their child's skills.

In addition to parental beliefs, the survey also contains basic demographic and socioeconomic variables, such as race, gender, parental education, and family income in the fall of kindergarten. Another important feature of the data is the ability to group respondents together in schools. In the fall of kindergarten, we observe approximately twenty-one survey respondents per sampled school. We create proxies for the average math and reading skill in each school by averaging the student-level standardized IRT scores by school and grade.[9]

While the survey data is incredibly rich, the challenge in working with the ECLS-K is the high level of attrition. This is particularly problematic in our setting since we want to maintain a reasonable number of students in each school so that our proxies for school skill levels are informative. In the fall of kindergarten, there are 21,409 sampled children. Any kindergarten student who lacks a school identifier or is missing test scores and parental beliefs is excluded from the sample in kindergarten and all future grades. We pursue a similar strategy for 1st and 3rd grade observations, eliminating students who lack key information. This entails dropping a significant number of students since attrition is common between each round of interviews. Approximately 4,000 students attrit between

[9]The ECLS-K also contains a set of variables related to parental investment decisions. These are useful for illustrating that parental beliefs do in fact matter for choices. The investment variables are discussed in more detail in in Section 6.

kindergarten and 1st grade, with additional attrition of approximately 3,000 students between 1st and 3rd grade. Finally, we calculate the number of valid student observations available for each school-grade combination. The first time a student is associated with a school-grade combination with fewer than five students, we drop that observation and all subsequent observations associated with that student. This eliminates relatively few kindergarten students, but approximately 1,500 observations in both 1st and 3rd grade.

Our final sample contains 20,957, 19,759, 15,072, and 11,258 students in the fall of kindergarten, spring of Kindergarten, spring of 1st grade, and spring of 3rd grade respectively. Table 1 provides means for the key variables discussed above for each grade in our sample. The first few rows of the table indicate that attrition is not entirely random since the sample becomes increasingly white and wealthier as measured by income in the fall of kindergarten. Also, the number of students per school declines considerably as a result of attrition. The parental belief variables indicate a significant skewness. More than 30% of the sample think their child thinks and solves problems better than other children his/her age, while only 7% think their child thinks and solves problems slightly less well or much less well than other children. Similar patterns are observed when parents are asked to compare their child to other children in the same class.

3 Evidence of Local Distortions in Parental Beliefs

In this section, we present descriptive evidence that parental beliefs about a child's skill relative to similarly aged children is impacted by the skill level of his/her schoolmates. Throughout the analysis, our primary measure of parental beliefs comes from the survey question that asks parents to asses a child's ability to learn, think, and solve problems relative to children of a similar age. Parents are given four options when answering this question, however, almost all parents respond that their child is either better or as good as similarly aged children. Thus, for most of this section we treat parental beliefs as if they are binary, with a one indicating that they believe their child is above average. We refer to this binary variable as parental global beliefs.

Table 2 illustrates the relationship between parents' beliefs about whether a child is above average and school average math scores using a linear probability model. The first column excludes the school average score, and simply relates parental beliefs to a child's math score. The results indicate a strong relationship between parental beliefs and a child's score, despite the fact that parents do not observe these measures directly. A one-standard deviation increase in a child's ECLS-K math score is associated with a 16 percentage point increase in the probability that a parent thinks the child is above average. In the second column of Table 2, we add as a regressor the average standardized IRT math score of the other children in a child's school and grade. This measure is calculated within-sample, and thus represents a noisy measure of school skills. The coefficient on the school average score is negative and statistically significant at a 1% level, suggesting that parental beliefs are negatively correlated with average school skills.

However, this result could be spurious since school average math scores are likely correlated with a host of other student characteristics that might also influence parental beliefs. The results in columns (3) and (4) of Table 2 assuage this concern by controlling for a host of student and family characteristics, such as reading scores, race, income, and maternal education. In column (3) the controls are included linearly, while in column (4) all of the controls, including own math and reading score are interacted in a flexible manner.[10] We continue to find a strong negative and significant relationship between school average math scores and parental beliefs.[11] Utilizing the results from column (3), a one standard deviation increase in school average math score is associated with a 4.4 percentage point reduction in the probability that a parent thinks their child is above average. At the mean, this represents a 13% decline in the likelihood of a parent reporting that a child is above average.

The ECLS-K math assessment is not necessarily designed to measure how well a child learns, thinks, and solve problems. To investigate whether the local distortion in parental beliefs cited above is robust to this concern, we replace the school average math score with the school average reading score. The results, listed in Table 3, are almost identical to the results obtained with math scores. School average reading scores have a significant

[10]We include all pairwise interactions among controls other than the school average score and third order polynomials in own math and reading scores. We do not report the linear coefficients from this specification since they are difficult to interpret in isolation.

[11]Because we find that adding non-linear controls yields no significant change in the school average coefficient, in all of our robustness checks we focus on a model with linear controls. However, all results are robust to the inclusion of 3rd-order polynomials in test scores.

negative impact on parental global beliefs, conditional on a child's own scores and a host of demographic characteristics. It appears as if school skill measures, generally construed, are strongly related to parental beliefs about the skill of a child. If parents were fully aware of how their child compared to the population average, then the local average should have no impact on parental global beliefs.[12]

3.1 Robustness of Local Distortions in Parental Beliefs

While the results in Tables 2 and 3 are consistent with our hypothesis that parental beliefs are locally distorted, there are alternative explanations for a link between global beliefs and school average scores. In this section we pursue many of these alternative theories.

Local Distortion or Parent Misinterpretation?

A key concern regarding the results in Tables 2 and 3 is that parents may simply be misinterpreting the global beliefs question. While we interpret the question as a comparison between the child and the average child in the population, parents may believe they are being asked to compare their child to other children in the child's class. As a result, the school average test score should influence parental responses. However, we can investigate this concern directly since in the 1st and 3rd grade waves of the survey, parents are asked to compare their child's math and reading skills to other children in their child's class. If parents are misinterpreting the global beliefs question, we would expect similar responses to the two types of belief questions and for own and school average test scores to have similar relationships with the two belief measures.

The simplest way to determine if parents interpret these questions differently is to compare responses. Using the detailed global and local belief responses (better, as well, slightly less well, much less well), we find that more than 40% of respondents answer the questions differently. This is true whether we use the local beliefs questions focused on math or reading. While this suggests that at least some of the respondents interpret the questions differently, the relationship between global beliefs and school average test scores could still be driven by the 60% of respondents who provide the same answer to the global

[12] For the regressions listed in Tables 2 and 3, we treat attrition as if it were random. However, we can replicate these regressions using only students who remain in the sample through third grade with no appreciable change in the patterns.

and local beliefs questions. However, Table 4 indicates that this is not the case. We repeat our basic global belief regression using only 3rd grade survey responses and split the sample according to whether respondents answer the global and local belief questions differently in 1st grade.[13] The idea behind splitting the sample in this manner is that parents who answer the local and global belief questions differently in 1st grade likely interpreted these questions differently. As a result, we would expect them to continue to interpret the local and global belief questions differently when surveyed in 3rd grade. The results indicate that the impact of school average scores on global beliefs in 3rd grade is large and significant regardless of whether respondents answered the local and global beliefs questions identically in 1st grade. In fact, we cannot reject that the impact of school average scores on global beliefs are equal across the samples.

Another way to assess whether parents are simply misinterpreting the global beliefs question is to examine whether the determinants of local and global beliefs are similar. In other words, do own and school average scores impact the two types of belief outcomes similarly? We investigate this question by regressing both local and global beliefs on test scores for 1st and 3rd graders.[14] Table 5 presents the results of these regressions. In columns (1) and (2), we regress an indicator for whether parents report a child is doing much better than their classmates in math (reading) on own and school average math (reading) scores. In columns (3) and (4), we use as the dependent variable whether a child learns, thinks, and solve problems better than children of a similar age.[15] The impact of own and school average scores on local and global beliefs are quite different, again indicating that parents respond to these questions differently. Moreover, for the local beliefs regressions in columns (1) and (2), we see that that the coefficients on own and school average score have similar magnitudes and opposite signs. In other words, increasing both own scores and school average scores by one unit has essentially no impact on local beliefs. This is exactly what one would expect since only the deviation from the school average is informative about how a child compares to his classmates.

Ultimately, the evidence from Tables 4 and 5 is consistent with parents understanding

[13] We use the difference in the belief questions in first grade to split the sample since we don't want to select the sample based on the contemporaneous outcome.

[14] We limit the sample to 1st and 3rd graders since local beliefs are not available in other waves.

[15] We do not report results including the full set of controls as it becomes more difficult to make comparisons across the local and global belief regressions in the presence of many regressors.

that they are being asked two different questions and utilizing different reference points to assess their child's skills. However, school average scores may have an impact on global beliefs for other reasons unrelated to local distortions. We investigate some of these below.

Measurement Error

One potential problem affecting the global belief regressions in Tables 2 and 3 is measurement error. If a student's own test score is sufficiently noisy, then the school average might enter the belief regressions significantly since it may also act as a noisy measure of a student's underlying skill. Note that if this were the case we would expect the school average score to positively influence beliefs. A student who attends a school with a high average test score would tend to have higher unobserved skills. Thus, we might expect the local distortion to become even stronger if we are able to minimize measurement error in our test score measures.

To address this concern we re-estimate our baseline global belief regression instrumenting for child and school average math scores with child and school average reading scores. Each subject score provides a noisy measure of a child's underlying cognitive skill. As long as the noise in each measure is uncorrelated, using one measure to instrument the other eliminates any attenuation bias. Relative to OLS, the coefficient on the school average math score is approximately 25% larger in absolute value when using reading scores as instruments.

Other approaches for mitigating measurement error are also available. Rather than use one subject score to instrument the other subject score, we can simply average the two scores and use this as our measure of skill. For most of the remaining robustness checks this is the approach we take. We replace math (and reading) scores with a standardized average of math and reading scores to measure child skill. Additionally, we can use lag scores, either math, reading, or an average of the two, to instrument for contemporaneous scores in the belief regressions. We find that the size of the local distortion is again approximately 25% larger than the simple OLS regression results indicate.[16]

Heterogenous Reference Points

[16] Additional results available upon request.

The results thus far strongly suggest that parental beliefs about a child's skill relative to children of a similar age are influenced by comparisons between the child and his/her schoolmates. However, an alternative interpretation could be that parents have heterogenous reference points for similarly aged children. For example, parents in California don't compare their child to children in Texas since they won't compete for the same colleges or future jobs. If this is the case, then a child's standing in the local skill distribution may be more important for parents than the child's placement in the overall distribution. We investigate whether this type of behavior can explain the patterns we observe by including additional skill averages based on groupings broader than the school. The two groupings we consider are based on socioeconomic characteristics and geography. For socioeconomic groupings, we find the average skill for students of the same race and gender with similar family incomes. For groupings based on geography, we construct average skill by census region and whether the child lives in an urban, suburban, or rural community. Table 6 illustrates that when these averages are included in our baseline regressions, the coefficients on own skill and school average skill are essentially unchanged. This suggests that the school average is not picking up a more "local" comparison then all children of a similar age.

School Sorting

Our baseline global belief regressions (Tables 2 and 3) rely primarily on cross-school variation in average skill to identify the local distortion effect. However, one possibility is that parents who send their children to schools with high (low) average skills are simply more (less) pessimistic. To explore this possibility, we exploit *within-school* heterogeneity in average *classroom* skill to see if the local distortion we document above exists at this micro level. There are a number of factors that will likely make it difficult to identify a distortion in parental global beliefs associated with variation in average skill across classrooms within a school. First, a child's current classmates do not represent all the classmates a child has encountered or is likely to encounter in a school. Since it is not clear exactly what set of students parents use as the local reference point, the across classroom variation in skill within a school is likely to have a weaker relationship with parental global beliefs. Second, our measure of classroom average skill will be much noisier than our measure of school average skill since we only observe a handful of children in a particular class. Despite these two issues, much of the information parents receive regarding their child's performance flows from the child's teacher. If teachers use within-class comparisons to evaluate children, then any across classroom differences in average skill are likely to impact parental beliefs.

Table 7 shows that indeed classroom deviations are significant predictors of parental global beliefs even when we control for school fixed effects. Column (1) regresses global beliefs on classroom average skills and own skills without school fixed effects simply to illustrate that the classroom average is less strongly related to beliefs as compared to school average skills. In column (2) we incorporate school fixed effects, and while the coefficient on classroom average skill declines, it remain negative and statistically significant. Finally, we estimate the school fixed effect model using reading scores as instruments for math scores to minimize measurement error. The classroom level distortion increases by more than 40% relative to column (2). Thus, even when we condition on a child's own skill and the school they attend, classroom level skill comparisons influence whether a parent believes their child is above average relative to all children of a similar age. To mitigate further the noisiness of the classroom measures, we repeat the specifications from the first three columns of Table 7, but only include classrooms where we observe at least 5 students. The relationship between classroom average skills and global beliefs increases by more than 20% in columns (5) and (6) relative to columns (2) and (3).

An alternative approach to mitigate concerns related to unobserved heterogeneity across schools is to examine how beliefs vary when students move schools. The challenge with this strategy is that the ECLS-K survey is designed to follow schools, not students, and thus when a student leaves a school they generally exit the sample. However, there are a small number of students who switch schools within sample. Using these school switchers we find that parents of students who switch to higher skill schools are more likely to revise their global beliefs downward relative to students who switch to lower skill schools. While the estimated coefficients using school switchers are not statistically significant, the general patterns are consistent with parents using local comparisons to inform global beliefs.[17]

[17]Further details available upon request.

3.2 Heterogeneity in Parental Belief Distortions

The impact of school average skill on parental beliefs should be smaller among parents who are well informed. In particular, if parents know where a child's school fits in the overall skill distribution, then the local comparison is directly informative about a child's skill relative to the overall average.

The first three columns of Table 8 show how the impact of school skills on global beliefs varies with maternal education, family income, and whether the child has an older sibling that attended the same school. Well educated, high income families may have more options for choosing a school, and thus possibly obtain more knowledge about the relative ranking of various schools when making their choice. Also, parents of children who have older siblings attending the same school have had more time to collect information about a school's relative performance. The results presented in Table 8 are consistent with the idea that highly educated and wealthy parents tend to have global beliefs that are less distorted than other types of households, but there exists no sibling effect. As an example, the impact school average skill has on global beliefs is approximately 35% smaller for college graduate mothers relative to mothers with a high school degree or less.

Taking the arguments above one step further, one might expect the distortion in global beliefs to vary with the average skill level in a school. The final two columns of Table 8 list estimates from our basic global belief regression, splitting the sample according to whether school average skill is greater than or less than zero. We fail to reject that the local distortion varies across these samples. A similar result occurs if we compare the local distortion for schools in the top and bottom 25th percentiles of the school average skill distribution. So while certain types of families appear to be more informed than others, the phenomena of using local information to infer something about the broader population is not limited to one segment of the school distribution.[18]

[18]There is also no evidence that the distortion varies according to the skill of the child. Results available upon request.

4 Parental Learning and the Evolution of Beliefs

The evidence reported in the previous section suggests that parents have distorted beliefs about the skill of their child relative to the broader population. Yet, how do these local distortions arise and evolve? Interestingly, if we estimate the reduced-form global belief regression by grade, we find that the local distortion in global beliefs is actually growing over time in a statistically significant manner. The coefficient on the school average test score is equal to -0.064, -0.083, and -0.104 in kindergarten, first grade, and third grade respectively. On the surface, this seems to suggest that parents are not learning about the true relative skill of their child as they age. However, we also find evidence that parental global beliefs are becoming more stable over time, consistent with learning. The auto-correlation coefficient in a regression of global beliefs on lag global beliefs increases significantly from 0.33 to 0.38 between first and third grade.[19] How can we reconcile these two features of the data?

In this section we develop a model of parental learning and skill evolution that is capable of matching these salient features. The model will specify the parental learning process including the information parents receive, how parents interpret information, and the evolution of child skill. We then show formally the features in the data that identify the key model parameters. Finally, we discuss model estimation.

4.1 The Model

Parents are uncertain about the skill of their child and the skill of the average child in the local school, but are assumed to know the overall average skill level in the economy (which we normalize to zero).[20] A more general approach would allow for parents to be uncertain about all three skill levels. However, as we illustrate in Appendix A, the ECLS-K only captures parental beliefs about skill differences and therefore we cannot separately identify uncertainty about the three skill levels. We need to make one normalization, which for

[19]The estimated increase understates the true degree of increasing persistence since the time gap between when current beliefs and lag beliefs are elicited is 25% longer in third grade.

[20]As we show in Appendix A, this setup is isomorphic to a model where parents are certain about the skill level of their child, but uncertain about the average skill level in the local school and the average skill level in the overall population. Note that this also means that our learning model is essentially equivalent to a model of learning about relative skills.

the purposes of this paper is without loss of generality. In a more comprehensive model that included parental investment for example, it would be useful to separately identify beliefs about child, school, and population skill levels. Yet, as long as parents care about the relative performance of their child, the information distortions we identify are still critical for understanding parental behavior. Parents likely have direct preferences over the relative skill of their child for various reasons, including any direct utility parents receive from comparing their child to others, tournament features of the higher education, marriage, and labor markets, and/or uncertainty about the long-term returns to skill.

4.1.1 Setup

Child i's life begins in kindergarten ($t = 1$) and she attends primary school for T periods. Children attend different schools (indexed by j) and are assumed to never change schools. Children are also differentiated by a time invariant index of observable family characteristics x_{ij}. The average level of this variable in school j is $x_j = \frac{1}{N_j} \sum x_{ij}$, where N_j is the number of students observed in school j. The overall average of observables is normalized to zero in the population.

The initial average skill level in the population is normalized to zero. The initial average skill level in school j is determined according to:

$$a_{j1} = \beta_{j1} x_j + \epsilon_{j1} \quad \text{where } \epsilon_{j1} \sim N(0, \pi_j). \tag{1}$$

The presence of x_j in the average skill level of school j allows schools with higher levels of parental income and education to have higher levels of initial skill. The initial skill level for child i is determined according to:

$$a_{ij1} = \beta_{i1}(x_{ij} - x_j) + a_{j1} + \epsilon_{ij1} \quad \text{where } \epsilon_{ij1} \sim N(0, \pi_i). \tag{2}$$

The skill level for child i varies from the average skill level in school j both through observed and unobserved channels. It is also useful to define $u_{ij1} = a_{ij1} - a_{j1}$, or the deviation between child i's skill and the average skill level in school j. In the learning model presented below, we find it simpler to work with the school deviation (u_{ij1}) as opposed to the level of child i's skill (a_{ij1}).

While parents do not observe a_{j1} and a_{ij1}, we assume that parents observe perfectly their own index of characteristics, x_{ij}, and the average index of characteristics in school j, x_j.[21] Prior to receiving signals in the first period, parents form priors over their child's skill and the average skill level in school j based entirely on these indices and their knowledge of the overall distributions. The priors are given below,

$$\begin{pmatrix} \hat{a}_{j1} \\ \hat{u}_{ij1} \end{pmatrix} = \begin{pmatrix} \mu + \beta_{j1} x_j \\ \beta_{i1}\left(x_{ij} - x_j\right) \end{pmatrix} \tag{3}$$

with a perceived mean squared error given by

$$\Sigma_1 = \begin{pmatrix} \Sigma_1^j & \Sigma_1^{ij} \\ \Sigma_1^{ij} & \Sigma_1^i \end{pmatrix} = \begin{pmatrix} \pi_j & 0 \\ 0 & \pi_i \end{pmatrix} \tag{4}$$

It is important to note that $\hat{a}_{j1}$ should be indexed by i, since the parents of each child receive different signals and thus have different beliefs. For ease of exposition we suppress the i index in most of our formulations. However, there will be times where it is useful to distinguish $\hat{a}_{j1}$ for a particular student i, which we accomplish in the following manner $\hat{a}_{j_i 1}$.

The distribution of the school and individual skill shocks ensures that skills will be normally distributed in the initial period. Throughout the model we make additional normality assumptions with respect to signal noise and transition shocks. These assumptions greatly simplify our learning model and approach to estimation and identification.

4.1.2 Skill Signals and Parental Updating

Parents begin each period t with prior beliefs over the school average skill level ($\hat{a}_{jt}$) and the difference between their child's skill level and the school average ($\hat{u}_{ijt}$). Associated with these priors is a level of precision, given by Σ_t. During period t, parents update their priors after receiving one signal about the average skill level in school j and one signal about the

[21]The index is constructed as a weighted sum of family income, maternal education, child gender, and child race. The weights are derived from a regression of math scores on these observable measures in the fall of kindergarten.

skill level of their child. A critical feature of our model will be the difference between the true content of the signals and parents' interpretation of the signals. This asymmetry will allow for both parental learning and consistent local distortions, features we document in our data.

Parents believe they receive signals taking the following form,

$$\begin{aligned} S^J_{ijt} &= \tilde{\alpha}_{jt} a_{jt} + \tilde{v}^J_{ijt} && (5) \\ S^I_{ijt} &= \tilde{\alpha}_{it} a_{ijt} + \tilde{v}^I_{ijt} = \tilde{\alpha}_{it} u_{ijt} + \tilde{\alpha}_{it} a_{jt} + \tilde{v}^I_{ijt} && (6) \end{aligned}$$

while the true signals are instead given by

$$\begin{aligned} S^J_{ijt} &= \alpha_{jt} a_{jt} + v^J_{ijt} && (7) \\ S^I_{ijt} &= \alpha_{it} a_{ijt} + \tilde{d}_{jt} a_{jt} + v^I_{ijt} = \alpha_{it} u_{ijt} + d_{jt} a_{jt} + v^I_{ijt}, && (8) \end{aligned}$$

where $\tilde{\alpha}_{ct} = \alpha_{ct} + D_{ct}$ for $c \in \{i, j\}$. v_{ijt} are independent, normally distributed shocks with means equal to zero and variances equal to one.[22] The second half of equations (6) and (8) simply restate the signals in terms of the deviation between child i's skill and the average skill level in school j.

The first signal parents receive, S^J_{ijt}, is a dedicated signal of the average skill level in school j, and reflects the notion that parents collect independent information about their child's school. We allow parents to inflate or deflate the true informativeness of this signal through D_{jt}. Parents interpret the second signal they receive, S^I_{ijt}, as a dedicated measure of child i's skill. However, their interpretation is at odds with the truth in two ways. First, the school average skill level may impact S^I_{ijt} directly through $\tilde{d}_{jt}$. As an example, parents might interpret a child's report card strictly as a measure of the child's skill relative to the population average, when it fact it has a local component.[23] Second, we allow parents to inflate or deflate the true informativeness of S^I_{ijt} through D_{it}. Intuitively, D_{it} and D_{jt} will

[22] $\tilde{v}^J_{ijt}$ and $\tilde{v}^J_{ijt}$ incorporate additional terms related to the distortions on child and school skill. However, when parents update beliefs, they treat $\tilde{v}^J_{ijt}$ and $\tilde{v}^J_{ijt}$ as if they have the same distribution as v^J_{ijt} and v^J_{ijt}.

[23] This is consistent with what we observe in the ECLS-K. In the survey, teachers are asked to assess the skill of each child relative to similarly aged children. We find that similar to parents, teacher reports of child skill are negatively impacted by the average skill in the child's school. Moreover, we find that teacher reports are highly correlated with parental beliefs. Because the precise communication between parents and teachers is unobserved, we do not use the teacher assessments directly in the model.

affect the rate at which parents update their beliefs, while d_{jt} can generate the local belief distortions documented in Section 3.

The signal structure outlined above will allow for both learning and persistent distortions, but it does not allow for heterogeneity in these distortions. Yet, in the data we find that higher income and more educated households have smaller local distortions in global beliefs relative to lower socioeconomic status households (see Table 8). To accommodate this feature of the data, we allow for observed heterogeneity in the signal parameters related to x_{ij} and x_j. In particular, we assume that $\alpha_{ct} = \check{\alpha}_{ct} \exp\left(\phi_j^{\alpha} x_j + \phi_i^{\alpha} x_{ij}\right)$ for $c \in \{j, i\}$. We model observed heterogeneity in D_{ct} for $c \in \{j, i\}$ and d_{jt} in a similar fashion. Heterogeneity in signal parameters can arise through two sources. First, higher socioeconomic status households may be better able to parse signals of similar content. Second, the signals received by higher socioeconomic households may be of differential quality. As a result of this heterogeneity, the weights in the signal equations are individual specific. In the exposition below we suppress this heterogeneity, though we allow for this flexibility in estimation.

While the signals S_{ijt}^J and S_{ijt}^I are observed by parents, they are unobservable to the econometrician. The appropriate way to view our modeling of these signals is as a composite of multiple sources of information parents collect. Parents can obtain information through their own research, interactions with school personnel such as teachers and administrators, interactions with their child's friends, or interactions with friends and family. We are agnostic about the precise sources of this information, though we believe teachers play a critical role (see footnote 23).

Upon receiving $S_{ijt} = \left[S_{ijt}^J \ \ S_{ijt}^I\right]'$, parents update their beliefs regarding a_{jt} and u_{ijt}. This is a straightforward application of the updating step in a standard Kalman filter. The equation below shows how beliefs evolve following receipt of these signals:

$$\begin{pmatrix} \hat{a}''_{jt} \\ \hat{u}''_{ijt} \end{pmatrix} = \begin{pmatrix} \hat{a}_{jt} \\ \hat{u}_{ijt} \end{pmatrix} + \Sigma_t H_t^T \left(H_t \Sigma_t H_t^T + 1\right)^{-1} \left(S_{ijt} - H_t \begin{pmatrix} \hat{a}_{jt} \\ \hat{u}_{ijt} \end{pmatrix}\right) \tag{9}$$

where

$$H_t = \begin{bmatrix} \alpha_{jt} + D_{jt} & 0 \\ \alpha_{it} + D_{it} & \alpha_{it} + D_{it} \end{bmatrix}$$

and H_t^T is its transpose.[24] The key feature of the above equation is the disconnect between how parents interpret the signal and the true content of the signal. The updating equation uses the parents interpretation of the signal, embedded in H_t. However, S_{ijt} will reflect the true signals, and this discrepancy can lead to distortions in parental beliefs.

At the end of period t, parental beliefs are fully characterized by $\hat{a}''_{jt}$, $\hat{u}''_{ijt}$, and Σ''_t. We use the double prime notation to indicate that these are the beliefs after incorporating the information from the two time t signals. Note that parental beliefs for the skill level of child i can be calculated as $\hat{a}''_{ijt} = \hat{a}''_{jt} + \hat{u}''_{ijt}$.

4.1.3 Skill Transitions and Parental Forecasting

Skill levels change as children age and progress through school. Thus, parents are attempting to learn about objects that are evolving over time. In this section we describe how child i's skill level and the average skill level in school j transition from one period to the next. We also discuss how parents incorporate the evolution of skill into their beliefs about these objects.

We assume that the skill level for child i transitions according to

$$a_{ijt} = \mu_{0t} + \beta_{jt} x_j + \beta_{it} \left(x_{ij} - x_j \right) + \rho_{it} a_{ijt-1} + \rho_{jt} a_{jt-1} + \epsilon_{jt} + \epsilon_{ijt} \tag{10}$$

where ϵ_{jt} is a common school level shock and ϵ_{ijt} is a student specific skill shock. These are assumed to be normally distributed with means equal to zero and standard deviations defined by σ_{jt} and σ_{it} respectively. The skill equation is similar to a value-added model in the sense that unobserved skill from the previous period, a_{ijt-1}, influences skill at time t. We also allow a_{jt-1} to enter directly to capture heterogeneity in school and peer quality. Time invariant observed characteristics x_j and x_{ij} also influence skill evolution, proxying for unobserved input quantity and quality.[25]

[24] The perceived mean square error is updated according to $\Sigma''_t = \Sigma_t - \Sigma_t H_t^T \left(H_t \Sigma_t H_t^T + 1 \right)^{-1} H_t \Sigma_t$.

[25] We have experimented with a skill transition function that allows parental beliefs to impact skill accumulation directly, but prefer this simpler approach. OLS estimates of an extended version of Equation (10) indicate that parental global and local beliefs have a positive and significant influence on skill transitions. However, once we allow for non-linear impacts of past scores and eliminate measurement error using reading to instrument for math, parental beliefs have small and insignificant effects. Thus, if we include parental beliefs in the skill transition equation, we would also want to allow for non-linear test score impacts. The computational costs associated with this outweigh the benefit, especially since the focus is on belief formation.

The form of the skill transition function for child i has implications for how the average skill in school j transitions. By construction, $a_{jt} = E_j(a_{ijt})$, where E_j denotes the expected value in school j. We realize that schools have a finite number of students and as a result the realized school average might differ from the expected value. However, in practice the two objects will be very close and this simplification does not have an impact on the estimates of our model.[26] Defining other school averages similarly, we also know that $E_j(x_{ij}) = x_j$, $E_j(a_{ijt-1}) = a_{jt-1}$, and $E_j(\epsilon_{ijt}) = 0$. Thus, the transition for the average skill level in school j is given by

$$a_{jt} = \mu_{0t} + \beta_{jt}x_j + (\rho_{it} + \rho_{jt})\, a_{jt-1} + \epsilon_{jt}. \tag{11}$$

Using the skill equation for child i in school j along with the skill transition for the average skill level in school j, we can define how the difference between the two evolves. The transition for the local skill deviation is given by the difference between Equations (10) and (11), shown below

$$u_{ijt} = \beta_{it}\,(x_{ij} - x_j) + \rho_{it}u_{ijt-1} + \epsilon_{ijt}. \tag{12}$$

Parents understand that skills evolve according to Equations (10)-(12), but do not observe the precise level of skill at anytime, nor the shocks to skill transition. However, using these equations parents can forecast the time t skill level for child i and the average skill level in school j given their beliefs about these objects in period $t-1$. In other words, parents will generate time t beliefs, $\hat{a}_{jt}$ and $\hat{u}_{ijt}$, by taking conditional expectations of equations (11) and (12) given their perceived information set. We denote these parental expectations by E_i. Taking the expectation E_i over equations (11) and (12) yields the following formulas for parental predictions of time t skills,

$$\hat{a}_{jt} = \mu_{0t} + \beta_{jt}x_j + (\rho_{it} + \rho_{jt})\, \hat{a}''_{jt-1} \tag{13}$$

$$\hat{u}_{ijt} = \beta_{it}\,(x_{ij} - x_j) + \rho_{it}\hat{u}''_{ijt-1}. \tag{14}$$

These objects characterize parental beliefs at the start of period t, and will eventually be updated when time t signals are received.

[26] We have estimated versions of our model where the expected value was replaced by the actual school average and the estimated parameters are largely unchanged.

The signal and belief updating process combined with the skill evolution and forecasting process generates a recursive structure to parental belief formation. We rely on this recursive structure when estimating the model using a Kalman filter based likelihood approach. However, before presenting our estimation strategy, we discuss how the key parameters of the model are identified.

4.2 Identification of Parental Learning Model

We focus our discussion of identification on the parameters describing the signal structure outlined in Section 4.1.2: α_{jt}, α_{it}, d_{jt}, D_{jt}, and D_{it}. The basic approach is to show how using a single cross-section, the observed relationships between beliefs, skills, and lag beliefs identify the parameters of interest. In Appendix B we provide the detailed algebra that underlies the arguments we make in this section.

For any generic period t, we observe whether parents believe that their child is above average relative to similarly aged children in the population and similarly aged children in child i's school. Denote these beliefs $\mathbb{1}(\hat{u}''_{ijt} + \hat{a}''_{jt} > k_{1t})$ and $\mathbb{1}(\hat{u}''_{ijt} > k_{2t})$, where the primes indicate that these are beliefs after receiving period t signals.[27] In addition to contemporaneous belief indicators, we assume that the following objects are also observable for all i and j: school average skill (a_{jt}), the difference between the skill level of child i and school average skill (u_{ijt}), lag global beliefs ($\hat{a}''_{jt-1} + \hat{u}''_{ijt-1}$), and lag local beliefs ($\hat{u}''_{ijt-1}$).[28] While this assumption makes the exposition of our identification argument simpler, it is at odds with the data in two respects. First, skills are actually unobserved, though multiple noisy measures are available. Second, we do not observe continuous measures of lag beliefs, instead we observe indicators for whether lags of global and local beliefs are above certain thresholds. At the end of this section we argue that the identification of the signal parameters is robust to these limitations.

To derive the model implied relationships between beliefs, skills, and lag beliefs, we substitute the expressions for the true skill signals, Equations (7) and (8), into the parental updating formula, Equation (9). This allows us to relate $\hat{a}''_{jt}$ and $\hat{u}''_{ijt}$ to a_{jt}, $\hat{a}_{jt}$, u_{ijt}, and

[27] For estimation we allow beliefs to be directly influenced by x_{ij} and x_j. Our identification approach can be applied conditional on these values.

[28] Under this assumption, the skill transition parameters and prior distributions are immediately identified. We assume these parameters are known when identifying the signal structure.

$\hat{u}_{ijt}$ as functions of the signal parameters and unknown mean square error parameters for parental beliefs at the start of period t. While we don't observe $\hat{a}_{jt}$ and $\hat{u}_{ijt}$ directly, we can construct them using Equations (13) and (14) and knowledge of the lagged belief measures and skill evolution parameters.

The signal parameters can then be identified using the coefficients from the following reduced-form regression,

$$\begin{aligned} \text{Prob}(\hat{u}''_{ijt} + \hat{a}''_{jt} > k_{1t}) &= 1 - \Phi\left(\tilde{k}_{1t} - \Gamma_{1t}\hat{a}_{jt} - \Gamma_{2t}\hat{u}_{ijt} - \Gamma_{3t}a_{jt} - \Gamma_{4t}u_{ijt}\right) \\ \text{Prob}(\hat{u}''_{ijt} > k_{2t}) &= 1 - \Phi\left(\tilde{k}_{2t} - \Gamma_{5t}\hat{a}_{jt} - \Gamma_{6t}\hat{u}_{ijt} - \Gamma_{7t}a_{jt} - \Gamma_{8t}u_{ijt}\right) \end{aligned} \tag{15}$$

where k_{1t} and k_{2t} are the cutoffs parents use to decide whether the gap between their child and the relevant comparison group is large enough to claim that the child is above average. $\tilde{k}_{1t}$ and $\tilde{k}_{2t}$ are the same cutoffs scaled by the standard deviation of the unobserved components of beliefs. There are eight Γ's and eight parameters to be identified: α_{jt}, α_{it}, d_{jt}, D_{jt}, D_{it}, Σ_t^j, Σ_t^i, and Σ_t^{ij}. The Σ_t parameters are functions of the previous period's signal and skill transition parameters, but because we want to rely on a single cross-section we treat them as unknown. While we are unable to solve this complicated system analytically, we can show numerically that the parameters of interest are identified. Further details are provided in Appendix B.

At the start of the identification section we made two assumptions that contradict the data. First, we assumed that skills are observable to the econometrician, though we only observe noisy measures of these skills. Consider the measurement system given below,

$$\begin{aligned} m_{ijt} &= a_{ijt} + e_{ijt}^m \\ r_{ijt} &= \mu_t^r + \lambda_t^r a_{ijt} + e_{ijt}^r \end{aligned}$$

where m_{ijt} and r_{ijt} are observable ECLS-K math and reading scores respectively. We can essentially use an instrumental variables type approach to estimate the skill transition and belief equation parameters, using one noisy measure to proxy for the other. Second, we only have binary measures of lagged beliefs, not the continuous measures of lag beliefs we utilize in the analysis above. However, as discussed in Heckman (1981), as long as the unobserved components of beliefs are iid, we can identify the Γ's by employing a multivariate probit model. In our setting, the unobserved components of beliefs are functions of the noise in the parental signals, which we assume are uncorrelated both over time and across signals.

4.3 Estimation

We estimate the model by maximum likelihood using a generalized Kalman filter algorithm. The Kalman filter is a useful way to model the joint distribution and evolution of a set of unobserved variables when noisy measures of these variables are available each period. Importantly, our model is non-standard in two key dimensions: a reliance on both discrete and continuous measures and the interdependence of state variables among children in the same school. In this section we describe the broad structure of our estimation approach in the face of these challenges, but leave the derivation of the detailed likelihood function to Appendix C.

The unit of analysis in our model is a school. Each school is inhabited by N_j students who are described by their skill levels and parental beliefs. All children in school j face the same school average skill level and experience common unobserved shocks to this quantity. The state space for school j at the end of period t is composed of both student level and school level unobservables. There are three state variables per student within school j, u_{ijt}, $\hat{u}''_{ijt}$, and $\hat{a}''_{jt}$. These describe the true skill level and actual parental beliefs for student i. There is an additional state variable common to all students in school j, a_{jt}. This state variable reflects the true average skill level in school j. The total number of state variables in school j is given by $3 \times N_j + 1$.

At the end of any period t, it is possible to characterize the mean and variance of the unobserved state vector in school j as a function of the prior distribution of the unobserved state variables and the model parameters governing parental signals. The precise structure of these matrices is presented in Appendices C.1 and C.2, however, a key feature of the model is the covariance structure describing the correlations in the parental belief state variables. Parents form $\hat{u}''_{ijt}$ and $\hat{a}''_{jt}$ based on signals that are in part a function of a common unobserved factor, a_{jt}. As a result, $\hat{u}''_{ijt}$ and $\hat{a}''_{j_it}$ will be correlated with $\hat{u}''_{i'jt}$ and $\hat{a}''_{j_{i'}t}$ for $i \neq i'$, where $\hat{a}''_{j_it}$ reflects beliefs about average skill in school j for the parents of child i. This correlation will be important when calculating the likelihood of the observed parental belief measures, as it will require estimating a joint probability.

While we are able to characterize the joint distribution of the state space as a function of the signal and skill transition parameters, we cannot observe child skill, school skill, and parental beliefs. However, in each period t we observe noisy measures of these objects. In particular, for each student we observe a math score, a reading score, whether parents believe their child is above average relative to similarly aged children in the population, and whether parents believe their child is above average relative to similarly aged children in the same school. We assume these measures for student i in school j and period t are generated according to:

$$\begin{aligned} m_{ijt} &= a_{ijt} + e^m_{ijt} \\ r_{ijt} &= \mu^r_t + \lambda^r_t a_{ijt} + e^r_{ijt} \\ \mathbb{1}(\hat{u}''_{ijt} + \hat{a}''_{jt} + \beta^G_i x_{ij} + \beta^G_j x_j &> k_{1t}) \\ \mathbb{1}(\hat{u}''_{ijt} + \beta^L_i x_{ij} + \beta^L_j x_j &> k_{2t}) \end{aligned} \tag{16}$$

where e^m_{ijt} and e^r_{ijt} are mean zero, normal random variables with variances equal to σ^m_t and σ^r_t respectively. The math score sets the scale of unobserved skill and we allow the mean and loading factor for the reading score to vary. The final two measures are indicators for parental beliefs. We allow parental beliefs to be directly impacted by individual and school observables. This implicitly allows the relevant cutoff point for being "above average" to vary across parents based on observable characteristics.

The likelihood contribution associated with school j is simply the joint probability of observing the $4 \times N_j \times T$ test score and belief measures conditional on a given parameter vector. However, it is infeasible to work with this joint probability, so we decompose the likelihood in the following manner,

$$L_j = \prod_{t=1}^{T} p_{t|t-1}\left(h_{jt}|h_j^{t-1}\right)$$

where h_{jt} is a vector of all the measures from school j at time t and $p_{t|t-1}\left(h_{jt}|h_j^{t-1}\right)$ is then the probability of observing the time t measures conditional on the full history of measures prior to time t, h_j^{t-1}. To compute these conditional probabilities we rely on a non-linear state space approach that is akin to a generalized Kalman filter. The non-linearity in our algorithm arises because the belief related data is a step function of the underlying state variables. In Appendix C.3, we derive closed form expressions for updating the conditional distribution of the state space in response to the discrete belief measures. The overall likelihood is then simply the product of the school specific likelihoods.

Our methodology is an approximation to a standard Kalman filter in the sense that we maintain the assumption that the conditional expectation of the unobserved state variables is normally distributed despite the presence of discrete measures. While this is not exactly true, we find that this approximation does not introduce biases in our estimation.[29] Cunha *et al.* (2010) make a similar assumption about the normality of the conditional expectation of the unobserved state variables in the presence of a non-linear state space transformation. While Cunha *et al.* (2010) rely on the unscented transform to update the mean and variance of the conditional state space, we are able to calculate the updated mean and variance using precise closed form expressions. Our approach is a better approximation than the unscented Kalman filter, but it can only be applied when the non-linearity arises because discrete measures (with an underlying normally distributed latent variable) are available.

5 Results

We estimate the parental learning model using the ECLS-K data discussed in Section 2 and analyzed in Section 3.[30] There are four periods in our estimated model, corresponding to: (1) the fall of kindergarten, (2) the spring of kindergarten, (3) the spring of 1st grade, and (4) the spring of 3rd grade. Below, we present and discuss the parameter estimates for the learning model. We follow this with an assessment of model fit and perform some counterfactual exercises.

[29]We perform some simple Monte Carlo exercises where we simulate data from the true model and then estimate the parameters using our approximation approach. The parameters are precisely estimated and do not suffer from any significant biases.

[30]Relative to the descriptive analysis, the data is altered slightly to accommodate the learning model. First, we collapse student observable characteristics such as gender, race, and income into a single index using weights from a regression of kindergarten math scores on observable characteristics. An average index of observables for each school is then constructed within sample. Second, we exclude observations where a student switches school in a non-structural fashion. A structural transition occurs when an entire first grade class moves to a new school. In these cases, we simply treat the students as if they remain in the same school. For non-structural moves, the student is treated as if they attrited from the sample. In general, we assume that attrition occurs randomly. These assumptions simplify the estimation procedure.

5.1 Parameter Estimates

Table 9 presents the estimates related to child and school average skill. The initial distribution of these objects is determined by the first four parameters in the table. To help put these parameters in context, note that the standard deviation of the school observable index (x_j) is 0.33 and the standard deviation of the child deviation from the school index ($x_{ij} - x_j$) is 0.35. Using these values plus the parameters it is straightforward to show that the initial across-school variance of average skill is 0.35, with approximately half of this attributed to observable factors. The within-school variation in child skill is 0.60, with approximately 90% of the variability related to unobserved child specific factors. As a result, parents face significantly more uncertainty regarding their child's skill relative to the school average then they face regarding the school average itself.

The remaining parameters in Table 9 describe the evolution of child skill across grades. To reduce the number of parameters, we restrict the impacts of observables and skills to be proportional across periods according to ϕ_t. For example, $\beta_{jt} = \phi_t \beta_{j2}$ for $t \in \{3, 4\}$. β_{it}, ρ_{jt}, ρ_{it} for $t \in \{3, 4\}$ are defined similarly. The estimates indicate that current skills are the best predictor for next periods skill, though school average skills and child observables are also positively related to next period skills. Shocks to skill evolution are larger in the spring of first and third grades, consistent with the gap in time being greater across these periods. The school-wide shock to skill evolution is generally of the same magnitude as that of the child specific shocks. These shocks create additional uncertainty for parents as they attempt to learn about child and school average skill.

The parameters of the signal equations drive parental learning and are presented in Table 10. Note that because we do not observe local and global beliefs every period (nor do we observe lag beliefs in the initial period), we constrain the signal parameters to be the same across periods 1 through 3 and allow them to change only for the final period. The upper panel describes the parameters for the signal of school average skill. The estimates indicate that the school specific signal is rather uninformative, containing mostly noise. The true loading on the school average is close to zero in the first few periods, and only slightly above zero in the final period. These estimates suggest that, at least early on, parents do not receive useful information about school average skill beyond what they can infer from the average observables in a school. However, parents believe they are receiving useful information. The large values for D_{jt} indicate that parents put a lot of faith in these signals, despite their lack of informativeness. In the next section we illustrate further the role that the D's play in parental learning.

The parameters governing the child specific skill signal are presented in the middle panel of Table 10. Recall that parents believe they are receiving a dedicated signal of child skill, however, we allow for the possibility that the school average influences the signal directly. $\tilde{d}_{jt}$ captures this direct effect and is estimated to be approximately -0.4. This implies that the true child specific signal is much more informative about the deviation from the school average (loading equal to approximately 0.65) than it is about the school average (loading equal to 0.25). The large values for D_{it} again indicate that parents are putting significantly more faith in the signals than they actually deserve.

The bottom panel of Table 10 shows the parameters that govern heterogeneity in skill signals. These parameters scale the α, D, and $\tilde{d}$ coefficients according to x_{ij} and x_j. The idea is to allow observably different parents to receive different quality signals. The only place where heterogeneity appears to matter is in the loading factor associated with school average skill in the child skill signal. It appears that higher individual and school observables tend to scale down this loading factor.[31]

Finally, Table 11 lists the parameter estimates for our measurement equations. The left panel contains the estimates for the math and reading score measures, while the right panel contains the estimates for the belief measures. The math and reading score measures are both highly informative of the underlying skill of a child, with the math measure being slightly less noisy. For an average child ($x_{ij} = 0$) in an average school ($x_j = 0$), the belief cutoffs indicate that for parents to claim their child is actually above average they must believe the child's skill is approximately 10% of a standard deviation larger than the true average. The coefficients on the individual and school x's have little impact on global beliefs. However, for local beliefs we find that being in a high x_j school directly reduces the likelihood of reporting that a child is above the local average.

[31] Recall that the effective loading factor is equal to $\tilde{d}_{jt}\exp(\phi_i^d x_{ij} + \phi_j^d x_j)$.

5.2 Model Fit and Counterfactuals

While the parameter estimates appear to be consistent with our initial descriptive evidence, in this section we replicate the key regressions from Section 3 using our estimated model. This allows us to illustrate how well the model fits the patterns described earlier. We also explore how beliefs would relate to child skill in various counterfactual scenarios.

Table 12 replicates our basic global belief regression under various scenarios. Each row of the table is a separate regression where the dependent variable is an indicator for whether the parent reports that their child is above average relative to similarly aged children. The coefficients of interest are associated with child skill and school average skill. All regressions also control for grade effects and individual and school observable indices. The top panel of the table utilizes noisy math scores to proxy for child and school average skill. The bottom panel instead uses true child skill, which is only feasible when using data simulated from our model. The first row of results is generated using the ECLS-K data and is similar to our descriptive analysis. Conditional on child skill, a higher school average skill makes it less likely that parents believe their child is above average. The second row runs the identical regression but uses data that was generated from our model.[32] Our model is able to replicate well the basic distortion pattern observed in the data.

The next three rows in the top panel of Table 12 illustrate how the local distortion in parental beliefs changes when the parameters of the model are altered. In each case we modify the parameters and simulate new student and school data. First, we set $\tilde{d}_{jt} = 0$. This means that the child skill signal is in fact a dedicated signal of a_{ijt}, aligning parent perceptions of the signal with reality. When this occurs, we find that the school average skill level continues to have an effect on parental beliefs, except now it is positive. This occurs because the math scores utilized to construct child and school skill are noisy and therefore school averages act as an additional proxy for a child's underlying skill. In the "No Distortions" row, we eliminate all the parental distortions by setting $\tilde{d}_{jt}$, D_{it}, and D_{jt} equal to zero. Relative to simply setting $\tilde{d}_{jt} = 0$, there is little change in the skill coefficients. However, setting D_{it} and D_{jt} equal to zero does impact parental beliefs. Figure 1 displays

[32] For our simulation, we match the number of schools in the sample and assume that there are 20 children per school. However, we create attrition in our simulated data to match the attrition we observe in the sample. Thus, we approximate the number of children observed per school across grades.

the distribution of prediction errors parents make when forecasting child skill across the four periods in our model. Prediction errors are defined as the difference between parental global beliefs and true child skill. The graph illustrates that the D parameters act to mitigate learning. When the D's are eliminated, the distribution of prediction errors shrinks from the fall of kindergarten to the spring of 3rd grade. This does not occur in the model or when only the distortion related to $\tilde{d}_{jt}$ is eliminated. In the row labeled "Perfect Knowledge", we assume that parents can observe child skill directly. In this case there is no distortion, but because we are using noisy measures the school average continues to have a large positive effect.

In the final row of the top panel of Table 12, we explore how the local distortion in parental beliefs might change were parents made aware of their child's math and reading test scores. As children progress past 3rd grade, standardized testing becomes increasingly prevalent. Moreover, testing is more common for recent cohorts of children. We find that when parents are able to observe test scores the coefficient on the school average declines by about 10%, while the coefficient on the child's math score increases by 25%. Parents are better informed relative to the baseline model, but the distortion remains large and significant. This result is driven in part by the assumption that parents continue to overweight the unobserved signal of child skill. However, we do not know how these weights might change, nor how the structure of the unobserved signal might change were test scores made available to parents. At a minimum, simply providing parents with unbiased scores will not eliminate local distortions.

In the bottom panel of Table 12, we repeat the same regressions but use the actual skill measures rather than noisy math scores. This is only possible using simulated data, and thus there is no row titled "Data". The general pattern across these four regressions are similar to the upper panel. The baseline model generates a large and significant local distortion. This distortion is larger than it was in the top panel because measurement error has been eliminated. Once we set $\tilde{d}_{jt} = 0$, the local distortion is completely eliminated. The average skill level in a child's school has no impact on whether parents believe their child is above average relative to similarly aged children. When parents are perfectly informed, child skill is the only regressor that has any predictive power. Finally, when test scores are made available the local distortion shrinks relative to the baseline model, but remains large.

In our descriptive analysis, we showed that high income and higher educated households exhibited smaller local distortions in parental beliefs. Table 13 shows that our model is able to replicate this heterogeneity. High observable type families tend to put more weight on child skill and less weight on school average skill in the global belief regression. A one standard deviation increase in x_{ij} reduces the coefficient on the school average score by approximately 30%. When distortions are eliminated the school average math score continues to matter because the skill measures are noisy. However, the impact of the school average no longer varies with parent characteristics since it is an equally good signal of child skill for high and low x_{ij} parents.

Finally, as motivation for our parental learning model we discussed the idea that parents appear to be learning while at the same time local distortions worsen. Our model captures these basic facts. First, using our simulated data we find that the dispersion of beliefs declines by approximately 25% between the first and final period. This occurs despite the unobserved shocks to skill evolution. Second, Table 14 shows that our model is able to replicate the increase in belief distortions over time. For both the data and model, the local distortion in parental beliefs increases in 3rd grade relative to the fall of kindergarten. In addition we find that as time passes the coefficient on the child's own score is increasing. When we eliminate distortions, the impact of the school average math score is positive in all periods, but has a diminishing effect.

The above discussion illustrates that our model is able to fit the patterns in the data and that eliminating distortions in how parents acquire information greatly mitigates distortions in parental beliefs. To put in context the degree to which signal distortions impact parental beliefs, consider the share of parents who believe their child is above average in low skill schools. In schools that are at least one standard deviation below average in initial test scores, 27% of 3rd grade parents believe their child is above average. When distortions are eliminated, only 13% of 3rd grade parents believe their child is above average, a more than 50% decline. An alternative approach for assessing the size of the distortion is to examine the differences between continuous measures of parental beliefs about child skill and actual child skill. For 3rd grade parents in schools initially scoring at least one standard deviation below average, the gap between beliefs and truth is 0.82, or about 80% of a standard deviation in skill. On average, these parents believe their child is 0.8 standard deviations higher in the skill distribution than they actually are. This gap is reduced to 0.23 when parental distortions are removed. In addition to reducing the size of the prediction errors, eliminating signal distortions also shrinks the dispersion in parental prediction errors by 18%. Parents are significantly better informed in our no distortion counterfactual relative to the baseline.

Yet, a key question remains. Is parental investment related to our measure of beliefs? If so, do local distortions tend to shrink or expand investment gaps? The next section speaks to these questions.

6 Parental Beliefs and Investment

While the magnitude of systemic biases in parental assessments of child skill is interesting in its own right, it takes on significantly more importance if parents act on these distorted beliefs. In particular, if parents view skill investments as complementary to child skill, then local distortions in parental beliefs will tend to mitigate skill differences across schools. Parents of students in low (high) skill schools will believe their child is more (less) highly skilled than they actually are and invest more (less). However, if parents view skill investments as a substitute for child skill, the exact opposite occurs and local distortions will exacerbate skill gaps across schools.

We can address this issue directly since parents of the children in the ECLS-K are surveyed each wave about the various activities they perform with their child. These activities include, for example, how often they read books to the child, go to the museum, and help the child with homework. When exploring the link between global beliefs and each of these activities, we always relate the contemporaneous measure of the activity with a lagged measure of global beliefs. We do this because many of the activity questions ask parents to look retrospectively over the past month or past school year. These past activities could directly affect parental beliefs at the time of the survey.

Tables 15 and 16 contain the results of this analysis. Table 15 shows the impact of parental beliefs on a series of categorical measures of investment, where the range of responses is typically the number of days a week the activity occurs. Table 16 instead shows the impact of parental beliefs on a series of binary measures of parental investment. Each cell in the left panel of both tables represents a separate OLS regression of investment on beliefs where only the coefficient on lagged global beliefs is shown. The rows indicate the different measures of investment utilized as outcomes, while the sample and controls vary across columns (1)-(3). In particular, column (3) controls for the entire vector of lagged investment in the hopes of controlling for permanent unobserved heterogeneity in parental behavior.[33] The right panel of both tables relates the magnitude of the coefficient from column (3) to the scale of the investment measure. The rows are ordered according to the impact parental beliefs has on the investment outcome.

The first thing to notice across the two tables is that parental beliefs have the largest impact on two types of investment that we would consider compensatory. When parents believe their child is above average relative to similarly aged children, they are less likely to help with homework and less likely to hire a tutor.[34] Conditional on all lagged investment measures and lagged child test scores, the number of times per week parents help with homework decreases by 9% of a standard deviation and the likelihood of hiring a tutor declines by 17% when parents believe their child is above average.[35]

The second takeaway from the two tables is the general lack of an effect of parental beliefs on activity type investments that have been previously documented to be beneficial to skill development. For example, activities such as reading books, playing games, or attending a concert are not related to parental beliefs once we control for lagged investment behavior and lag test scores. In other words, it appears that unobserved heterogeneity across households drives these behaviors, not beliefs about child skill. Parental beliefs appear to have a positive and significant impact on two activity type investments, discussing nature/science and visiting a museum. However, the change in the impact of beliefs across columns (2) and (3) suggests that these effects may be spurious. When we don't control for lagged investment in column (2), parental global beliefs have a positive and significant

[33] When we condition on lagged investment, we lose observations from the earlier periods since lagged investment is not observed. Column (2) shows the impact of lagged beliefs using the same sample as in column (3), but without the lag investment measures.

[34] In addition to the documented controls, the homework regression also includes indicators for the number of times per week the child does homework at home. The homework results are robust to including school fixed effects, or controlling directly for the homework policy of the teacher.

[35] With only 2 or 3 observations available per student and discrete outcomes, fixed effects estimates of the investment equations yield noisy and imprecise estimates.

impact on many types of investment behaviors. However, the magnitude and significance of these effects declines dramatically when we move to column (3). This is also true for nature/science activities and visiting a museum. The magnitude of the belief coefficient declines by at least 50% when lagged investments are included. This suggests that if additional controls were available the impact would shrink further. The homework and tutor investment behaviors do not display the same pattern when lagged investment measures are included.

The descriptive analysis presented in Tables 15 and 16 suggests that local distortions might work to expand skill gaps across schools. Parents of children in schools with low average skill levels believe their child is doing better relative to the overall average than they actually are. As a result, there could be less pressure to intervene and ensure basic skills are being developed. At schools with high average skill levels, local comparisons lead parents to believe their child is less skilled relative to the overall average, potentially leading to increased investment. However, it is important to note that within our framework it is not possible to determine the precise impact eliminating local distortions would have on parental investment. As mentioned in Section 4.1, in order to make specific predictions we would need to understand parental preferences over local relative skill, global relative skill, and absolute skill. Furthermore, we would need to understand better the precise source of the uncertainty in relative skill, essentially going beyond what the current data allows us to identify. In particular, it would be important to obtain information on parental beliefs over the average level of performance in the whole population. Finally, in order to understand how belief distortions impact skill gaps, we would also need to consistently estimate the impact each parental investment behavior has on child skill.

7 Conclusion

In this paper we present evidence that parental beliefs about a child's skill relative to similarly aged children are distorted by a child's skill relative to children in the same school. Our model shows that these patterns are consistent with a parental learning model where parents misinterpret the signals they receive. In particular, parents believe they receive dedicated signal of overall skill, when in fact the signal is contaminated by the local average skill level. The model also illustrates how parental belief distortions change when we modify the parents' information set. As the previous section highlights, local belief distortions can have important consequences for parental investment. Parents of low skill children who attend schools where average skill is also low perform fewer remedial type investments than parents of similarly able children who attend schools where average skill is high. Because of the tendency for students and families to sort into schools and neighborhoods, low skill children are more likely to attend schools where average skill is also low. As a result, the distortion in parental beliefs generated by local skill comparisons can lead to underinvestment for low skill children.

Our paper complements recent work illustrating parental misinformation about child skill and development. Cunha *et al.* (2013) finds that socioeconomically disadvantaged, pregnant African American women have biased beliefs regarding the productivity of parental investment. Closer to our paper, Dizon-Ross (2013) finds that parents in Malawi significantly overstate their child's skill and when given more accurate information choose more remedial type investments to help their children. The elicitation of parental beliefs in these papers is cleaner than in our setup since the survey/experiments employed were designed precisely for this reason. However, by using the ECLS-K we are able to explore in greater detail the nature and source of parental distortions. As a result we are able to gain additional insight into policies capable of ameliorating these distortions.

The finding that parent beliefs about a child's relative skill are distorted by the local distribution also connects our paper to the broader peer effects literature. In an effort to estimate the impact of peers, researchers often estimate the impact average classroom skill has on individual test score outcomes. It is generally not clear the channel through which average peer skill operates, but the typical interpretation is that it works through in-school behaviors of either the teacher or students themselves. Our paper suggests that average peer skill also matters for individual outcomes through its impact on parental investment.

Parental investment in children, particularly at young ages, has been shown to a be a key input into skill development. As a result, it is imperative that we understand the key determinants of these investment decisions. Parental beliefs can play a significant role in these choices since it is difficult for parents to discern their child's skill and the marginal benefit of a myriad of potential inputs. We show in this paper that parental beliefs can be inaccurate in a systematic manner that can may have important aggregate impacts. However, significant work remains. In addition to the limitations discussed at the end of Section 6, parental beliefs about the returns to investment and beliefs about non-cognitive skills are also likely important and they may reinforce or mitigate biased beliefs over child cognitive skill. Moreover, embedding parental beliefs into a dynamic model of investment that accounts for borrowing constraints and the trade-offs between goods and time investments would be extremely informative. These additional constraints may temper the impact of beliefs or exacerbate them depending on the relationships between beliefs and other family characteristics.

References

Arcidiacono, Peter. 2004. Ability sorting and the returns to college major. *Journal of Econometrics*, **121**(1-2), 343–375.

Arcidiacono, Peter, Aucejo, Esteban, Maurel, Arnaud, & Ransom, Tyler. 2016. *College Attrition and the Dynamics of Information Revelation.* Working Paper.

Arcidiacono, Peter, Hotz, V. Joseph, & Romano, Teresa. 2017. *Recovering ex ante returns and preferences for occupations using subjective expectations data.* Working Paper.

Attanasio, Orazio, & Kaufman, Katja. 2014. Education Choices and Returns to Schooling: Intra-household Decision Making, Gender and Subjective Expectations. *Journal of Development Economics*, **109**, 203–216.

Attanasio, Orazio, Boneva, Teodora, & Rauh, Christopher. 2018. *Parental Beliefs about Returns to Different Types of Investments in School Children.* Working Paper.

Becker, Gary S., & Tomes, Nigel. 1986. Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, **4**(3), S1–S39.

Bernal, Raquel. 2008. The Effect of Maternal Employment and Child Care on Children's Cognitive Development. *International Economics Review*, **49**(4), 1173–1209.

Biroli, Pietro, Boneva, Teodora, Raja, Akash, & Rauh, Christopher. 2018. *Parental Beliefs about Returns to Child Health Investments.* Working Paper.

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

Boneva, Teodora, & Rauh, Christopher. 2017. *Human Capital Production and Parental Beliefs.* Working Paper.

Boneva, Teodora, & Rauh, Christopher. 2018. Parental Beliefs about Returns to Educational Investments-The Later the Better? *Journal of the European Economics Association.*

Carneiro, Pedro, & Heckman, James J. 2003. Human Capital Policy. *In:* Heckman, James J., Krueger, Alan B., & Friedman, Benjamin M. (eds), *Inequality in America: What Role for Human Capital Policies*. Cambridge, MA: MIT Press.

Caucutt, Elizabeth M., & Lochner, Lance. 2012. *Early and Late Human Capital Investments, Borrowing Constraints, and the Family*. NBER Working Paper 18493.

Caucutt, Elizabeth M., Lochner, Lance, & Park, Youngmin. 2017. Correlation, Consumption, Confusion, or Constraints: Why do Poor Children Perform so Poorly? *Scandinavian Journal of Economics*, **119**(1), 102–147.

Cunha, Flavio. 2013. *Investments in Children When Markets are Incomplete*. Working Paper.

Cunha, Flavio. 2015. *Gaps in Early Investments in Children*. Working Paper.

Cunha, Flavio, & Heckman, James J. 2008. Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, **43**(4), 738–782.

Cunha, Flavio, Heckman, James J., & Schennach, Susanne M. 2010. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, **78**(3), 883–931.

Cunha, Flavio, Elo, Irma, & Culhane, Jennifer. 2013. *Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation*. NBER Working Paper 19144.

Dizon-Ross, Rebecca. 2013. *Parents' Perceptions and Children's Education: Experimental Evidence from Malawi*. Working Paper.

Genz, Alan. 1992. Numerical Computation of Multivariate Normal Probabilities. *Journal of Computational and Graphical Statistics*, **1**(2), 141–149.

Heckman, James J. 1981. Statistical Models for Discrete Panel Data. *Chap. 3, pages 114–178 of:* Manski, C., & McFadden, D. (eds), *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge: MIT Press.

Heckman, James J., & Mosso, Stefano. 2014. The Economics of Human Development and Social Mobility. *Annual Review of Economics*, **6**, 689–733.

Manski, Charles F. 2004. Measuring Expectations. *Econometrica*, **72**(5), 1329–1376.

Stinebrickner, Ralph, & Stinebrickner, Todd R. 2014a. Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model. *Journal of Labor Economics*, **32**(3), 601–644.

Stinebrickner, Ralph, & Stinebrickner, Todd R. 2014b. A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. *The Review of Economic Studies*, **81**(1), 426–472.

Todd, Petra E., & Wolpin, Kenneth I. 2007. The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, **1**(1), 91–136.

Wiswall, Matthew, & Zafar, Basit. 2015. Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, **82**(2), 791–824.

Wiswall, Matthew, & Zafar, Basit. 2019. Preference for the workplace, human capital and gender. *Quarterly Journal of Economics*.

Table 1: Summary Statistics

	Fall-K	Spring-K	1st	3rd
% White	0.55	0.56	0.57	0.60
Log Income	10.51	10.53	10.55	10.59
Mom has BA	0.22	0.22	0.23	0.24
Children per School	21.1	20.0	16.0	12.6
Comparisons to children of same age				
Above Average	0.34		0.31	0.34
Below Average	0.08		0.08	0.09
Comparisons to children in same class				
Above Average, Math			0.37	0.35
Below Average, Math			0.05	0.08
Above Average, Reading			0.40	0.36
Below Average, Reading			0.10	0.10
Child Math Score	0.00	0.01	0.01	0.04
SD(Child Math Score)	1.00	1.00	0.99	0.99
School Average Math Score	0.00	0.01	0.01	0.04
SD(Average Math Score)	0.52	0.52	0.50	0.53
N	20,957	19,759	15,072	11,258

Data include students in kindergarten, 1st, and 3rd grade from the Early Childhood Longitudinal Study, Kindergarten Class of 1999. Data cleaning and sample restrictions are described in Section 3. Text for the questions pertaining to how a child compares to others can also be found in Section 3.

Table 2: Parental Global Beliefs and School Math Scores

"Does your child learn, think, and solve problems better" than other children his/her age?"

	(1)	(2)	(3)	(4)
School Math		-0.052	-0.082	-0.084
		(0.005)	(0.006)	(0.006)
Child Math	0.156	0.167	0.106	
	(0.002)	(0.003)	(0.004)	
Child Reading			0.085	
			(0.004)	
Log(Income_K)			0.003	
			(0.004)	
Black			0.011	
			(0.008)	
Hispanic			0.057	
			(0.008)	
Asian			0.019	
			(0.013)	
Other non-white			0.022	
			(0.011)	
Male			-0.028	
			(0.005)	
Mom has some college			0.039	
			(0.006)	
Mom has BA			0.085	
			(0.008)	
Grade effects	Y	Y	Y	Y
Flexible controls	N	N	N	Y
N	39,092	39,092	39,092	39,092
R^2	0.109	0.111	0.137	0.147

*,** Indicates coefficients significant at a 1% and 5% level respectively. The unit of observation is fall kindergarten, 1st, and 3rd grade students from the Early Childhood Longitudinal Study, Kindergarten Class of 1999 (ECLS-K). All models are estimated by OLS. Standard errors are clustered at the student level and are reported in parentheses. The test score regressors are standardized IRT math and reading scores from the ECLS-K. School Math is the grade specific school average math score calculated within sample, excluding a student's own score. Log(Income_K) is the log of family income measured in the fall of kindergarten. Flexible controls indicates that all pairwise interactions between controls other than school average are included along with 3rd degree polynomials in math and reading scores.

Table 3: Parental Global Beliefs and School Reading Scores

"Does your child learn, think, and solve problems better" than other children his/her age?"

	(1)	(2)	(3)	(4)
School Reading		-0.041	-0.071	-0.075
		(0.006)	(0.006)	(0.006)
Child Reading	0.153	0.161	0.088	
	(0.003)	(0.003)	(0.004)	
Child Math			0.101	
			(0.004)	
Ln(Income$_K$)			0.002	
			(0.004)	
Black			0.021	
			(0.008)	
Hispanic			0.061	
			(0.008)	
Asian			0.024	
			(0.013)	
Other non-white			0.026	
			(0.011)	
Male			-0.026	
			(0.005)	
Mom some college			0.038	
			(0.006)	
Mom has BA			0.084	
			(0.008)	
Grade effects	Y	Y	Y	Y
Flexible controls	N	N	N	Y
N	39,092	39,092	39,092	39,092
R^2	0.105	0.106	0.136	0.146

*,** Indicates coefficients significant at a 1% and 5% level respectively. The unit of observation is fall kindergarten, 1st, and 3rd grade students from the Early Childhood Longitudinal Study, Kindergarten Class of 1999 (ECLS-K). All models are estimated by OLS. Standard errors are clustered at the student level and are reported in parentheses. The test score regressors are standardized IRT math and reading scores from the ECLS-K. School Reading is the grade specific school average reading score calculated within sample, excluding a student's own score. Ln(Income$_K$) is the log of family income measured in the fall of kindergarten. Flexible controls indicates that all pairwise interactions between controls other than school average are included along with 3rd degree polynomials in math and reading scores.

Table 4: Parental Global Beliefs by Consistency in Lagged Local and Global Beliefs

"Does your child learn, think, and solve problems better than other children his/her age?"

	Lag Beliefs		Lag Beliefs	
	Global=Local$_M$	Global≠Local$_M$	Global=Local$_R$	Global≠Local$_R$
	(1)	(2)	(3)	(4)
School Math	-0.115	-0.091		
	(0.014)	(0.017)		
School Reading			-0.103	-0.074
			(0.014)	(0.016)
Own Scores	Y	Y	Y	Y
Demographics	Y	Y	Y	Y
N	5,526	3,715	5,313	3,941
R^2	0.156	0.155	0.173	0.128

*,** Indicates coefficients significant at a 1% and 5% level respectively. The unit of observation is 3rd grade students from the Early Childhood Longitudinal Study, Kindergarten Class of 1999 (ECLS-K). All models are estimated by OLS. Heteroskedastic robust standard errors are reported in parentheses. Own scores include both math and reading standardized IRT scores. School averages are grade specific and are calculated within sample excluding own scores. Local$_M$ and Local$_R$ are parental beliefs regarding how a child compares to classmates in math and reading respectively. The Global=Local column includes only students whose parents provided the same answer to the global and local belief question in 1st grade.

Table 5: Comparing School Score Effects on Local and Global Beliefs

	"Compared to other children in your child's class, how well is your child doing in school this spring in..."		*"Does your child learn, think, and solve problems better than other children his/her age?"*	
	Math	Reading		
	(1)	(2)	(3)	(4)
School Math	-0.126		-0.064	
	(0.007)		(0.007)	
Child Math	0.147		0.173	
	(0.003)		(0.003)	
School Reading		-0.093		-0.061
		(0.007)		(0.007)
Child Reading		0.107		0.176
		(0.004)		(0.003)
N	23,043	22,769	23,043	22,769
R^2	0.077	0.041	0.118	0.121

*,** Indicates coefficients significant at a 1% and 5% level respectively. The unit of observation is 1st and 3rd grade students from the ECLS-K. All models are estimated by OLS. Standard errors are clustered at the student level and are reported in parentheses. The test score regressors are standardized IRT math and reading scores from the ECLS-K. School averages are grade specific and are calculated within sample excluding own scores.

Table 6: Parental Global Beliefs: Heterogeneous Reference Points

"Does your child learn, think, and solve problems better" than other children his/her age?"

	(1)	(2)	(3)
School Skill	-0.078	-0.077	-0.077
	(0.006)	(0.006)	0.006
Socioeconomic Group Average Skill		-0.014	
		(0.016)	
Location Based Average Skill			-0.011
			(0.018)
Child Skill	0.175	0.175	0.175
	(0.003)	(0.003)	(0.003)
Grade Effects	Y	Y	Y
Demographics	Y	Y	Y
N	39,092	39,092	39,092
R^2	0.136	0.136	0.136

*,** Indicates coefficients significant at a 1% and 5% level respectively. The unit of observation is fall kindergarten, 1st and 3rd grade students from the ECLS-K. All models are estimated by OLS. Standard errors are clustered at the student level and are reported in parentheses. Own skill is the standardized average of IRT math and reading scores from the ECLS-K. School skill is the leave-out average skill in a student's school. Socioeconomic average skill is the leave-out average skill among students from the same income quartile, race, and gender. Location based average skill is the leave-out average skill among students from the same census region and population density (central city, large town, rural). All averages are constructed within sample.

Table 7: Parental Global Beliefs: School Fixed Effects

"Does your child learn, think, and solve problems better" than other children his/her age?"

	(1)	(2)	(3)	(4)	(5)	(6)
Classroom Average Skill	-0.051	-0.035	-0.050	-0.060	-0.047	-0.059
	(0.004)	(0.005)	(0.008)	(0.006)	(0.009)	(0.014)
Child Skill	0.176	0.179	0.252	0.174	0.175	0.249
	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.006)
School Fixed Effects	N	Y	Y	N	Y	Y
Classes with ≥ 5 students	N	N	N	Y	Y	Y
Instrument for Skill	N	N	Y	N	N	Y
Demographics	Y	Y	Y	Y	Y	Y
N	37,064	37,064	37,064	25,437	25,437	25,437
R^2	0.133	0.172	0.132	0.130	0.173	0.133

*,** Indicates coefficients significant at a 1% and 5% level respectively. The unit of observation is fall kindergarten, 1st and 3rd grade students from the ECLS-K. All models are estimated by OLS. Standard errors are clustered at the student level and are reported in parentheses. In columns (1)-(2) and (4)-(5), child and school skill is based on the standardized average of IRT math and reading scores from the ECLS-K. In columns (3) and (6), child and school skill is based on the standardized math score. Classroom skill is the leave-out average skill in a student's class.

Table 8: Heterogenous Local Distortions in Parental Global Beliefs

"Does your child learn, think, and solve problems better" than other children his/her age?"

				School Skill	
				≤ 0	> 0
	(1)	(2)	(3)	(4)	(5)
School Skill	-0.086	-0.305	-0.082	-0.086	-0.070
	(0.008)	(0.056)	(0.008)	(0.011)	(0.010)
School Skill x SomeColl	0.000				
	(0.012)				
School Skill x BA	0.028				
	(0.013)				
School Skill x Ln(Income$_K$)		0.021			
		(0.005)			
School Skill x Old Sibling			0.003		
			(0.010)		
Grade Effects	Y	Y	Y	Y	Y
Child Skill	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y
N	39,092	39,092	37,434	18,975	20,117
R^2	0.136	0.137	0.138	0.111	0.144

*,** Indicates coefficients significant at a 1% and 5% level respectively. The unit of observation is fall kindergarten, 1st, and 3rd grade students from the ECLS-K. All models are estimated by OLS. Standard errors are clustered at the student level and are reported in parentheses. School skill averages are calculated within sample and exclude own scores.

Table 9: Estimates for Skill Related Parameters

Parameter	Description	Estimate	Standard Error
	Initial Skill Distributions		
β_{j1}	Impact of observables on school skill	1.218	(0.081)
π_j	St. Dev. of unobserved school skill component	0.433	(0.015)
β_{i1}	Impact of observables on child skill	0.683	(0.016)
π_i	St. Dev. of unobserved child skill component	0.738	(0.004)
μ_{01}	Average skill, FK	-0.007	(0.030)
	Skill Transition Equation		
μ_{02}	Average shift in skill, SK	-0.009	(0.008)
μ_{03}	Average shift in skill, S1	-0.005	(0.010)
μ_{04}	Average shift in skill, S3	-0.010	(0.010)
β_j	Impact of school observables	-0.029	(0.020)
β_i	Impact of child observables	0.054	(0.006)
ρ_j	Impact of school skill	0.068	(0.011)
ρ_i	Impact of child skill	0.964	(0.006)
ϕ_3	Spring 1st Scale factor	0.929	(0.010)
ϕ_4	Spring 3rd Scale factor	0.990	(0.012)
σ_{j2}	St.Dev. of school level skill shock, SK	0.155	(0.005)
σ_{j3}	St.Dev. of school level skill shock, S1	0.181	(0.007)
σ_{j4}	St.Dev. of school level skill shock, S3	0.184	(0.008)
σ_{i2}	St.Dev. of child specific skill shock, SK	0.131	(0.007)
σ_{i3}	St.Dev. of child specific skill shock, S1	0.233	(0.005)
σ_{i4}	St.Dev. of child specific skill shock, S3	0.188	(0.009)

Section 4 defines each of the parameters in detail. FK indicates the fall of kindergarten. SK, S1, and S3 indicate the spring of kindergarten, 1st, and 3rd grade respectively.

Table 10: Estimates for Signal Related Parameters

Parameter	Description	Estimate	Standard Error
	School Skill Signal Parameters		
α_{j1-3}	Loading on school average skill: FK, SK, S1	0.000	(0.003)
D_{j1-3}	Parental distortion of above loading: FK, SK, S1	1.075	(0.095)
α_{j4}	Loading on school average skill: S3	0.097	(0.052)
D_{j4}	Parental distortion of above loading: S3	3.176	(0.232)
	Child Skill Signal Parameters		
α_{i1-3}	Loading on child skill deviation: FK, SK, S1	0.622	(0.013)
D_{i1-3}	Parental distortion of above loading: FK, SK, S1	3.952	(0.226)
$\tilde{d}_{j1-3}$	Loading on school average skill: FK, SK, S1	-0.388	(0.027)
α_{i4}	Loading on child skill deviation: S3	0.656	(0.032)
D_{i4}	Parental distortion of above loading: S3	4.319	(0.373)
$\tilde{d}_{j4}$	Loading on school average skill: S3	-0.428	(0.060)
	Signal Heterogeneity Parameters		
ϕ_i^α	Heterogeneity in loadings thru x_{ij}	0.014	(0.045)
ϕ_j^α	Heterogeneity in loadings thru x_j	0.038	(0.063)
ϕ_i^D	Heterogeneity in loading distortions thru x_{ij}	-0.087	(0.039)
ϕ_j^D	Heterogeneity in loading distortions thru x_j	0.030	(0.046)
ϕ_i^d	Heterogeneity in d_{jt} thru x_{ij}	-0.368	(0.124)
ϕ_j^d	Heterogeneity in d_{jt} thru x_j	-0.207	(0.156)

Section 4 defines each of the parameters in detail. FK indicates the fall of kindergarten. SK, S1, and S3 indicate the spring of kindergarten, 1st, and 3rd grade respectively.

Table 11: Measurement System Estimates

Math				*Local Beliefs*			
St.Dev. Noise	σ_1^m	0.419	(0.003)	Cutoffs	k_{1-3}^L	0.085	(0.014)
	σ_2^m	0.445	(0.002)		k_4^L	0.099	(0.008)
	σ_3^m	0.520	(0.003)				
	σ_4^m	0.484	(0.005)	Coeffs x_{ij}	β_{i1-3}^L	0.193	(0.013)
					β_{i4}^L	0.108	(0.013)
Reading							
Mean Shifts	μ_1^r	-0.041	(0.007)	Coeffs x_j	β_{j1-3}^L	-0.619	(0.057)
	μ_2^r	-0.028	(0.006)		β_{j4}^L	-0.306	(0.042)
	μ_3^r	-0.015	(0.006)				
	μ_4^r	-0.001	(0.006)	*Global Beliefs*			
				Cutoffs	k_{1-2}^G	0.105	(0.005)
Loading Factors	λ_1^r	0.895	(0.006)		k_3^G	0.098	(0.005)
	λ_2^r	0.912	(0.006)		k_4^G	0.081	(0.005)
	λ_3^r	0.969	(0.008)				
	λ_4^r	0.951	(0.009)	Coeffs x_{ij}	β_{i1-2}^G	0.035	(0.008)
					β_{i3}^G	0.023	(0.008)
St.Dev. Noise	σ_1^r	0.590	(0.002)		β_{i4}^G	0.049	(0.009)
	σ_2^r	0.591	(0.002)				
	σ_3^r	0.565	(0.004)	Coeffs x_j	β_{j1-2}^G	-0.003	(0.013)
	σ_4^r	0.552	(0.005)		β_{j3}^G	-0.044	(0.013)
					β_{j4}^G	-0.071	(0.016)

Standard errors in parentheses. Section 4 defines each of the parameters in detail. Periods 1-4 map to the fall of kindergarten, the spring of kindergarten, the spring of 1st grade, and the spring of 3rd grade.

Table 12: Model Fit and Counterfactuals

"Does your child learn, think, and solve problems better" than other children his/her age?"

	Child Skill	School Average Skill	Demographics & Grade Effects
Proxy skill using noisy math score			
Data	**0.159**	**-0.070**	**Y**
Model	0.161	-0.059	Y
$\tilde{d}_{jt} = 0$	0.158	0.046	Y
No Distortions	0.141	0.045	Y
Perfect Knowledge	0.293	0.106	Y
Model + Test Scores	0.201	-0.054	Y
Utilize true skill			
Model	0.234	-0.146	Y
$\tilde{d}_{jt} = 0$	0.220	-0.004	Y
No Distortions	0.199	-0.005	Y
Perfect Knowledge	0.412	0.014	Y
Model + Test Scores	0.271	-0.137	Y

Each row is a separate OLS regression. The dependent variable is an indicator for whether parents believe their child is above average relative to similarly aged children. The upper panel uses noisy, but observable math scores to proxy for child skill and school average skill. The bottom panel uses the true skill level based on simulated data. Simulated data is used in all rows but the first, where ECLS-K data is utilized. "Model" is simulated data using our baseline estimates. "No Distortions" indicates that $\tilde{d}_{jt}$, D_{it}, and D_{jt} are all zero. "Perfect Knowledge" indicates that parents know the skill level of their child and the school average skill level. "Model + Test Scores" allows parents to observe noisy test score measures in addition to distorted signals.

Table 13: Model Fit and Counterfactuals, Heterogeneous Learning

"Does your child learn, think, and solve problems better"
than other children his/her age?"

	Data	Model	No Distortions
Child Math	0.156	0.160	0.140
Child Math x x_{ij}	0.020	0.032	0.034
School Average Math	-0.074	-0.060	0.045
School Average Math x x_{ij}	0.045	0.042	0.009
Demographics	Y	Y	Y
Grade Effects	Y	Y	Y

Each column is a separate OLS regression. The dependent variable is an indicator for whether parents believe their child is above average relative to similarly aged children. "Data" uses the ECLS-K data for the regression. "Model" utilizes simulated data from our baseline estimates. "No Distortions" indicates that $\tilde{d}_{jt}$, D_{it}, and D_{jt} are all zero.

Table 14: Model Fit and Counterfactuals, Variability by Grade

"Does your child learn, think, and solve problems better"
than other children his/her age?"

	Data	Model	No Distortions
Child Math	0.150	0.153	0.112
Child Math x Grade=1	0.008	0.005	0.039
Child Math x Grade=3	0.027	0.024	0.068
School Average Math	-0.056	-0.055	0.072
School Average Math x Grade=1	-0.007	0.005	-0.034
School Average Math x Grade=3	-0.042	-0.024	-0.058
Demographics	Y	Y	Y
Grade Effects	Y	Y	Y

Each column represents is a separate OLS regression. The dependent variable is an indicator for whether parents believe their child is above average relative to similarly aged children. "Data" uses the ECLS-K data for the regression. "Model" utilizes simulated data from our baseline estimates. "No Distortions" indicates that $\tilde{d}_{jt}$, D_{it}, and D_{jt} are all zero.

Table 15: Parental Investments, Categorical Measures

Outcome	Coefficient on Lag Belief (1)	(2)	(3)	% of SD(Outcome) based on (3)
Help Homework	-0.064*	-0.096*	-0.101*	9.0%
Nature/Science	0.114*	0.117*	0.042*	5.1%
Chores	0.011	0.013	-0.026*	3.1%
Family Dinner	0.089*	0.096*	0.044	2.6%
Sing Songs	0.086*	0.085*	0.017	1.7%
Read Books	0.033*	0.035*	-0.016	1.6%
Family Breakfast	0.010	0.037	-0.032	1.3%
Tell Stories	0.083*	0.084*	0.010	1.1%
Practice Skill	0.042*	0.030	-0.004	0.5%
Play Games	0.050*	0.050*	0.004	0.5%
Home Crafts	0.061*	0.061*	0.002	0.3%
Build Something	0.053*	0.054*	-0.003	0.3%
Lag Student Scores	Y	Y	Y	
Demographics	Y	Y	Y	
Lag Investments	N	N	Y	

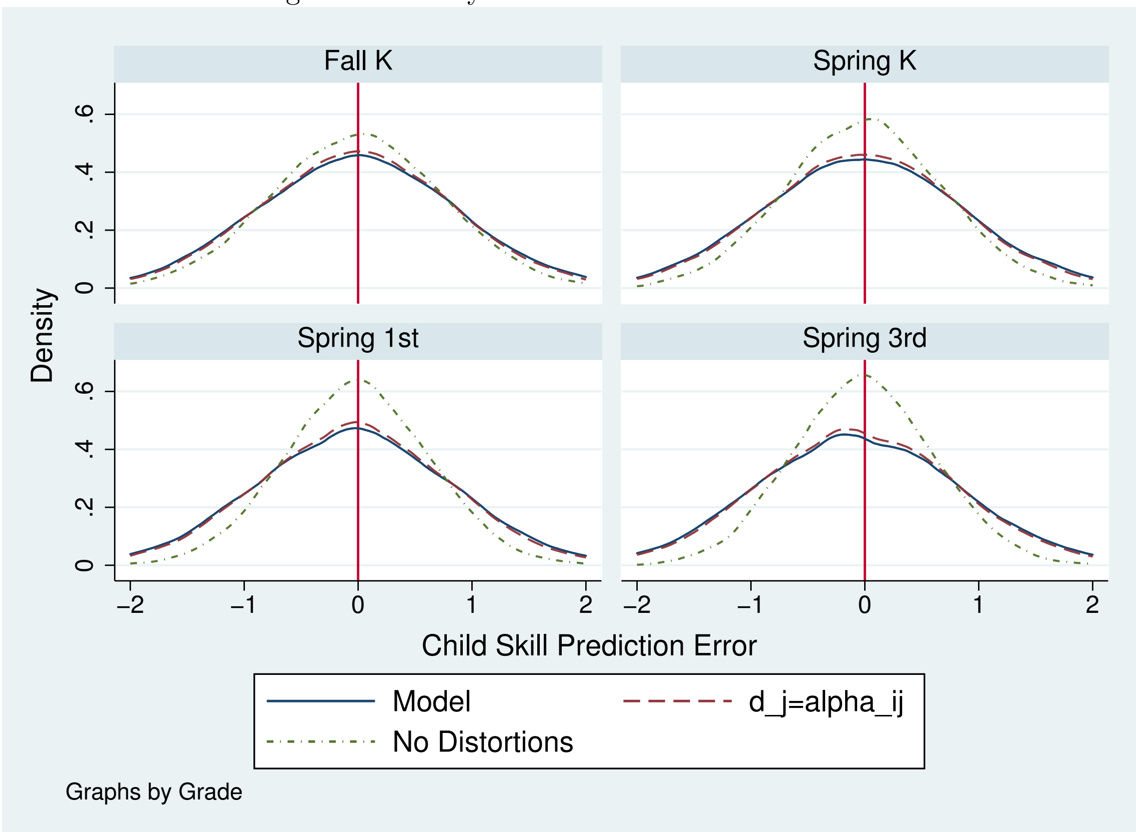
* Indicates coefficients significant at a 5% level. Each number in columns (1)-(3) reflect a separate regression of the outcome on the parental global belief indicator. All columns control for child demographics and lag math and reading scores. Column (3) also controls for lags of all categorical and binary investment measures. Column (2) uses the same sample as column (3) but without lag investment controls. The final column reports the ratio of the lag belief coefficient in column (3) to the standard deviation of the outcome variable.

Table 16: Parental Investments, Binary Measures

Outcome	Coefficient on Lag Belief (1)	(2)	(3)	% of Mean(Outcome) based on (3)
Hire Tutor	-0.028*	-0.025*	-0.022*	17.1%
Visit Museum	0.038*	0.050*	0.025*	7.5%
Art Activity	0.014*	0.018*	0.007	6.1%
Music Activity	0.018*	0.024*	0.008	5.1%
Visit Zoo	0.024*	0.031*	0.015	5.0%
Drama Activity	0.019*	0.023*	0.009	4.1%
Dance Activity	0.019*	0.019*	0.007	3.9%
Attend Concert	0.025*	0.032*	0.009	2.2%
Sport Activity	-0.003	0.004	0.007	1.1%
Attend Game	0.002	0.011	0.005	1.0%
Visit Library	0.014	0.017	0.003	0.6%
Club Activity	0.002	0.005	-0.001	0.3%
Lag Student Scores	Y	Y	Y	
Demographics	Y	Y	Y	
Lag Investments	N	N	Y	

* Indicates coefficients significant at a 5% level. Each number in columns (1)-(3) reflect a separate regression of the outcome on the parental global belief indicator. All columns control for child demographics and lag math and reading scores. Column (3) also controls for lags of all categorical and binary investment measures. Column (2) uses the same sample as column (3) but without lag investment controls. The final column reports the ratio of the lag belief coefficient in column (3) to the mean of the outcome variable.

Figure 1: Density of Child Skill Prediction Errors



Appendices

A Non-Identification of Child, School, and Population Skill

Ideally we would like to identify parental beliefs about their own child's skill, the average level of skill in their child's school, and the average level of skill in the population of similarly aged children. To put us in the best possible scenario for identifying these three objects, we assume that we observe child skill and that parental belief measures are continuous, although not necessarily centered at the truth. Dynamics are not essential since the same logic can be applied period by period. Skills are assumed to be normally distributed. While this assumption is not crucial, it allows us to fully characterize the problem using only the mean and variance of each random variable.

For a given cohort of children, assume that the world is characterized by:

1. Overall average ability, a
2. Average ability in child i's school j, $a_j \sim N(a, \sigma_1)$
3. Child i's ability, $a_{ij} \sim N(a_j, \sigma_2)$

Parents cannot observe any of these objects, but form beliefs over them based on their own information set: $\hat{a}_{ij} = E(a_{ij}|I_{ij})$, $\hat{a}_j = E(a_j|I_{ij})$, and $\hat{a} = E(a|I_{ij})$. We assume that these beliefs have the following shape:

$$\begin{aligned}
\hat{a} &\sim N(\mu_0 + \delta_0 a + \delta_1 a_j + \delta_2 a_{ij}, \gamma_0) \\
\hat{a}_j &\sim N(\mu_1 + \beta_0 a + \beta_1 a_j + \beta_2 a_{ij}, \gamma_1) \\
\hat{a}_{ij} &\sim N(\mu_2 + \lambda_0 a + \lambda_1 a_j + \lambda_2 a_{ij}, \gamma_2).
\end{aligned}$$

This representation is consistent with parental information arising from a series of noisy linear signals of the underlying skills. Identifying the three distributions of beliefs requires identifying the 15 parameters describing the means and the variances of the above normal distributions.

True parental beliefs, represented by $\hat{a}$, $\hat{a}_j$, and $\hat{a}_{ij}$, are unobserved. Instead, the ECLS-K asks parents whether their child is above average relative to their classmates and relative to similarly aged children. For this exercise, we assume that ECLS-K parents actually quantify the gap between their child and the relevant average. When observed beliefs are instead binary, it reduces the amount of information we can identify, but does not change its basic nature. Under these assumptions we **observe**:

$$\begin{aligned} \text{Local beliefs:} \qquad \hat{a}_{ij} - \hat{a}_j &= E\left(a_{ij} - a_j | I_{ij}\right) \\ \text{Global beliefs:} \qquad \hat{a}_{ij} - \hat{a} &= E\left(a_{ij} - a | I_{ij}\right) \end{aligned}$$

Define $\hat{a} - E\left(\hat{a}|a, a_j, a_{ij}\right) = u$, $\hat{a}_j - E\left(\hat{a}_j|a, a_j, a_{ij}\right) = u_j$, and $\hat{a}_{ij} - E\left(\hat{a}_{ij}|a, a_j, a_{ij}\right) = u_{ij}$. Using these deviations along with the posterior beliefs defined above, we can rewrite local and global **observed** beliefs in the following manner:

$$\begin{aligned} \hat{a}_{ij} - \hat{a}_j &= (\mu_2 - \mu_1) + (\lambda_0 - \beta_0)\, a + (\lambda_1 - \beta_1)\, a_j + (\lambda_2 - \beta_2)\, a_{ij} + u_{ij} - u_j \\ \hat{a}_{ij} - \hat{a} &= (\mu_2 - \mu_0) + (\lambda_0 - \delta_0)\, a + (\lambda_1 - \delta_1)\, a_j + (\lambda_2 - \delta_2)\, a_{ij} + u_{ij} - u \end{aligned}$$

Abilities are observed, an assumption consistent with the availability of multiple test scores in the ECLS-K. Thus, with linear regression methods we can identify the two constants, the four coefficients on a_j, a_{ij}, and the two variances $\gamma_2 + \gamma_1$ and $\gamma_2 + \gamma_0$, for a total of 8 coefficients. This number is smaller than 15, and therefore we cannot identify the full distribution of parental beliefs over the three unobservables.

Although there are several normalizations that yield identification, it is quite natural to assume that parents know the overall average of skills and that beliefs are centered, i.e. $(\delta_0 = 1, \delta_1 = 0, \delta_2 = 0, \gamma_0 = 0)$ and $(\mu_0 = 0, \mu_1 = 0, \mu_2 = 0)$.[36] Imposing these normalizations, observable beliefs become:

$$\begin{aligned} \hat{a}_{ij} - \hat{a}_j &= (\lambda_0 - \beta_0)\, a + (\lambda_1 - \beta_1)\, a_j + (\lambda_2 - \beta_2)\, a_{ij} + u_{ij} - u_j \\ \hat{a}_{ij} - \hat{a} &= (\lambda_0 - 1)\, a + \lambda_1 a_j + \lambda_2 a_{ij} + u_{ij} \end{aligned}$$

where each parameter is now just identified. Therefore, we conclude that given our data, we can assume without loss of generality that parents know the average skill in the population.

[36] An alternative and isomorphic strategy is to assume that parents observe their own child's skills perfectly.

The above derivation shows that the questions in the ECLS-K (in their most ideal form) are not sufficient to simultaneously identify child, school, and population skill levels. However, we can achieve identification by normalizing, for example, the population average skill level. This remains true even when using the less than ideal data on skills and beliefs available in the ECLS-K. For the remainder of the paper, we therefore assume without loss of generality that parents know the average skill in the population. The next section describing the identification of the the signaling model starts from this premise.

B Identification of Signaling Model

In this appendix, we first show how to derive Equation (15), the reduced-form equation used for identification of the learning parameters, and then describe our numerical approach to solving the system of equations underlying the coefficients of Equation (15). As discussed in section 4.2, we assume that child skill and lagged continuous beliefs are observed. These assumptions simplify the exposition and are innocuous. With multiple noisy measures of child skill we can instrument one for the other to achieve identification. Moreover, we can identify the belief equations derived below with lagged discrete belief measures following Heckman (1981). Importantly, these concepts also apply to the skill transition function, meaning that the skill evolution parameters are directly identified.

Prior to deriving the key identifying equation for the signal parameters, it is useful to describe our basic approach. Although we could potentially use the dynamic structure of the model to aid in identification, our strategy is to use only the cross-sectional information contained in time t to identify all the learning parameters relative to time t. Therefore, in this section we derive a mapping between observables and parameters for a generic period t and discuss how that mapping can be inverted to identify our parameters of interest.

B.1 Derivation of Equation(15)

For a generic period t we assume the following objects are observed: $\mathbb{1}(\hat{u}''_{ijt} + \hat{a}''_{jt} > k_{1t})$, $\mathbb{1}(\hat{u}''_{ijt} > k_{2t})$, a_{jt}, u_{ijt}, $\hat{a}''_{jt-1} + \hat{u}''_{ijt-1}$, and $\hat{u}''_{ijt-1}$. Recall that the primes indicate parental beliefs after receiving skill signals in the subscripted time period. With knowledge of $\hat{a}''_{jt-1}$ and $\hat{u}''_{ijt-1}$, we can derive $\hat{a}_{jt}$ and $\hat{u}_{ijt}$ using the known parameters of the skill evolution function. Our first goal is then to derive the precise mapping between posterior beliefs in period t, $\mathbb{1}(\hat{u}''_{ijt} + \hat{a}''_{jt} > k_{1t})$ and $\mathbb{1}(\hat{u}''_{ijt} > k_{2t})$, and a_{jt}, u_{ijt}, $\hat{a}_{jt}$ and $\hat{u}_{ijt}$. The coefficients describing the linear mapping between these variables will be the building blocks of our identification.

School Signal: To incorporate period t signals into beliefs, we take a sequential approach. The order of the updating does not matter so we begin with the school signal, given in

Equation (7) and repeated below for convenience:

$$S^J_{ijt} = \alpha_{jt} a_{jt} + v^J_{ijt}.$$

Recall that parents misinterpret this signal according to Equation (5), repeated below for convenience

$$S^J_{ijt} = \tilde{\alpha}_{jt} a_{jt} + \tilde{v}^J_{ijt}.$$

Using a scalar version of Equation (9), it is straightforward to show that parental beliefs after receiving the school signal are given by:

$$\begin{pmatrix} \hat{a}'_{jt} \\ \hat{u}'_{ijt} \end{pmatrix} = \begin{pmatrix} \hat{a}_{jt} \\ \hat{u}_{ijt} \end{pmatrix} + \begin{pmatrix} \frac{(\alpha_{jt}+D_{jt})\Sigma^j_t}{(\alpha_{jt}+D_{jt})^2\Sigma^j_t+1} \\ \frac{(\alpha_{jt}+D_{jt})\Sigma^{ij}_t}{(\alpha_{jt}+D_{jt})^2\Sigma^j_t+1} \end{pmatrix} \left(S^J_{ijt} - \begin{pmatrix} (\alpha_{jt}+D_{jt}) & 0 \end{pmatrix} \begin{pmatrix} \hat{a}_{jt} \\ \hat{u}_{ijt} \end{pmatrix} \right).$$

Note that this formula embeds the idea that parents misinterpret the school signal since D_{jt} alters the updating equation. The above formula also contains Σ^j_t and Σ^{ij}_t. These terms are part of the covariance matrix $\Sigma_t = \begin{pmatrix} \Sigma^j_t & \Sigma^{ij}_t \\ \Sigma^{ij}_t & \Sigma^i_t \end{pmatrix}$ describing the precision of parental beliefs at the start of period t. These variances and covariances are functions of the skill transition function and signal parameters from periods $t-1$ and earlier.

We can substitute for the true school signal, S^J_{ijt}, and express beliefs as a function of the true school average skill,

$$\begin{pmatrix} \hat{a}'_{jt} \\ \hat{u}'_{ijt} \end{pmatrix} = \begin{pmatrix} \frac{1}{(\alpha_{jt}+D_{jt})^2\Sigma^j_t+1}\hat{a}_{jt} \\ \hat{u}_{ijt} - \frac{(\alpha_{jt}+D_{jt})^2\Sigma^{ij}_t}{(\alpha_{jt}+D_{jt})^2\Sigma^j_t+1}\hat{a}_{jt} \end{pmatrix} + \begin{pmatrix} \frac{\alpha_{jt}(\alpha_{jt}+D_{jt})\Sigma^j_t}{(\alpha_{jt}+D_{jt})^2\Sigma^j_t+1} \\ \frac{\alpha_{jt}(\alpha_{jt}+D_{jt})\Sigma^{ij}_t}{(\alpha_{jt}+D_{jt})^2\Sigma^j_t+1} \end{pmatrix} a_{jt} + \begin{pmatrix} \frac{(\alpha_{jt}+D_{jt})\Sigma^j_t}{(\alpha_{jt}+D_{jt})^2\Sigma^j_t+1} \\ \frac{(\alpha_{jt}+D_{jt})\Sigma^{ij}_t}{(\alpha_{jt}+D_{jt})^2\Sigma^j_t+1} \end{pmatrix} v^J_{ijt}.$$

Because the coefficients on beliefs and skills will ultimately get quite complicated, we find it simpler to define an intermediate set of variables. Thus we re-write the above expression as

$$\begin{pmatrix} \hat{a}'_{jt} \\ \hat{u}'_{ijt} \end{pmatrix} = \begin{pmatrix} \lambda_{1t}\hat{a}_{jt} \\ \hat{u}_{ijt} - \lambda_{2t}\hat{a}_{jt} \end{pmatrix} + \begin{pmatrix} \lambda_{3t} \\ \lambda_{4t} \end{pmatrix} a_{jt} + \begin{pmatrix} \lambda_{5t} \\ \lambda_{6t} \end{pmatrix} v^J_{ijt}$$

where $\lambda_{1t} = \frac{1}{(\alpha_{jt}+D_{jt})^2\Sigma^j_t+1}$, $\lambda_{2t} = \lambda_{1t}(\alpha_{jt}+D_{jt})^2\Sigma^{ij}_t$, $\lambda_{3t} = \lambda_{1t}\alpha_{jt}(\alpha_{jt}+D_{jt})\Sigma^j_t$, $\lambda_{4t} = \lambda_{3t}\frac{\Sigma^{ij}_t}{\Sigma^j_t}$, $\lambda_{5t} = \frac{\lambda_{3t}}{\alpha_{jt}}$, and $\lambda_{6t} = \frac{\lambda_{4t}}{\alpha_{jt}}$.

Parents also update the precision of their beliefs following receipt of the school signal. Using the equation outlined in Footnote 16, it is straightforward to show,

$$\begin{pmatrix} \Sigma_t^{j\prime} & \Sigma_t^{\prime} \\ \Sigma_t^{\prime} & \Sigma_t^{i\prime} \end{pmatrix} = \begin{pmatrix} \frac{\Sigma_t^{j}}{(\alpha_{jt}+D_{jt})^2\Sigma_t^{j}+1} & \frac{\Sigma_t^{ij}}{(\alpha_{jt}+D_{jt})^2\Sigma_t^{j}+1} \\ \frac{\Sigma_t^{ij}}{(\alpha_{jt}+D_{jt})^2\Sigma_t^{j}+1} & \Sigma_t^{i} - \frac{(\alpha_{jt}+D_{jt})^2\Sigma_t^{ij}\Sigma_t^{ij}}{(\alpha_{jt}+D_{jt})^2\Sigma_t^{j}+1} \end{pmatrix}.$$

Student Signal: Parents believe that the next signal they receive is given by Equation (6), repeated below,

$$S_{ijt}^{I} = \tilde{\alpha}_{it}a_{ijt} + v_{ijt}^{I} = \tilde{\alpha}_{it}u_{ijt} + \tilde{\alpha}_{it}a_{jt} + \tilde{v}_{ijt}^{I}$$

while the true signal is instead given by Equation (8), repeated below

$$S_{ijt}^{I} = \alpha_{it}a_{ijt} + \tilde{d}_{jt}a_{jt} + v_{ijt}^{I} = \alpha_{it}u_{ijt} + d_{jt}a_{jt} + v_{ijt}^{I}.$$

Again, using a scalar version of Equation (9), it is straightforward to show that parental beliefs after receiving the student signal are given by:

$$\begin{pmatrix} \hat{a}_{jt}^{\prime\prime} \\ \hat{u}_{ijt}^{\prime\prime} \end{pmatrix} = \begin{pmatrix} \hat{a}_{jt}^{\prime} \\ \hat{u}_{ijt}^{\prime} \end{pmatrix} + \begin{pmatrix} \frac{(\alpha_{it}+D_{it})\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})\Sigma_t^{\prime}}{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}+2(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1} \\ \frac{(\alpha_{it}+D_{it})\Sigma_t^{\prime}+(\alpha_{it}+D_{it})\Sigma_t^{i\prime}}{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}+2(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1} \end{pmatrix} \left(S_{ijt}^{I} - (\alpha_{it}+D_{it})\,\hat{a}_{jt}^{\prime} - (\alpha_{it}+D_{it})\,\hat{u}_{ijt}^{\prime}\right)$$

where the prior is now parental beliefs after incorporating the school signal. Combining the terms associated with $\hat{a}_{jt}^{\prime}$ and $\hat{u}_{ijt}^{\prime}$ yields

$$\begin{aligned} \begin{pmatrix} \hat{a}_{jt}^{\prime\prime} \\ \hat{u}_{ijt}^{\prime\prime} \end{pmatrix} &= \begin{pmatrix} \frac{(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1}{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}+2(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1}\hat{a}_{jt}^{\prime} - \frac{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}}{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}+2(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1}\hat{u}_{ijt}^{\prime} \\ -\frac{(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}}{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}+2(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1}\hat{a}_{jt}^{\prime} + \frac{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1}{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}+2(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1}\hat{u}_{ijt}^{\prime} \end{pmatrix} \\ &+ \begin{pmatrix} \frac{(\alpha_{it}+D_{it})\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})\Sigma_t^{\prime}}{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}+2(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1} \\ \frac{(\alpha_{it}+D_{it})\Sigma_t^{\prime}+(\alpha_{it}+D_{it})\Sigma_t^{i\prime}}{(\alpha_{it}+D_{it})^2\Sigma_t^{j\prime}+(\alpha_{it}+D_{it})^2\Sigma_t^{i\prime}+2(\alpha_{it}+D_{it})^2\Sigma_t^{\prime}+1} \end{pmatrix} S_{ijt}^{I}. \end{aligned}$$

Again, it is useful to define some intermediate variables so that we can express the above equation in a simpler fashion. We rewrite beliefs according to,

$$\begin{pmatrix} \hat{a}''_{jt} \\ \hat{u}''_{ijt} \end{pmatrix} = \begin{pmatrix} \lambda_{7t}\hat{a}'_{jt} - \lambda_{8t}\hat{u}'_{ijt} \\ -\lambda_{9t}\hat{a}'_{jt} + \lambda_{10t}\hat{u}'_{ijt} \end{pmatrix} + \begin{pmatrix} \lambda_{11t} \\ \lambda_{12t} \end{pmatrix} S^I_{ijt}$$

where

$$\begin{aligned}
\lambda_{7t} &= \frac{(\alpha_{it}+D_{it})^2\,\Sigma^{i\prime}_t + (\alpha_{it}+D_{it})^2\,\Sigma'_t + 1}{(\alpha_{it}+D_{it})^2\,\Sigma^{j\prime}_t + (\alpha_{it}+D_{it})^2\,\Sigma^{i\prime}_t + 2\,(\alpha_{it}+D_{it})^2\,\Sigma'_t + 1} \\
\lambda_{8t} &= \frac{(\alpha_{it}+D_{it})^2\,\Sigma^{j\prime}_t + (\alpha_{it}+D_{it})^2\,\Sigma'_t}{(\alpha_{it}+D_{it})^2\,\Sigma^{j\prime}_t + (\alpha_{it}+D_{it})^2\,\Sigma^{i\prime}_t + 2\,(\alpha_{it}+D_{it})^2\,\Sigma'_t + 1} \\
\lambda_{9t} &= \frac{(\alpha_{it}+D_{it})^2\,\Sigma'_t + (\alpha_{it}+D_{it})^2\,\Sigma^{i\prime}_t}{(\alpha_{it}+D_{it})^2\,\Sigma^{j\prime}_t + (\alpha_{it}+D_{it})^2\,\Sigma^{i\prime}_t + 2\,(\alpha_{it}+D_{it})^2\,\Sigma'_t + 1} \\
\lambda_{10t} &= \frac{(\alpha_{it}+D_{it})^2\,\Sigma^{j\prime}_t + (\alpha_{it}+D_{it})^2\,\Sigma'_t + 1}{(\alpha_{it}+D_{it})^2\,\Sigma^{j\prime}_t + (\alpha_{it}+D_{it})^2\,\Sigma^{i\prime}_t + 2\,(\alpha_{it}+D_{it})^2\,\Sigma'_t + 1} \\
\lambda_{11t} &= \frac{\lambda_{8t}}{(\alpha_{it}+D_{it})} \\
\lambda_{12t} &= \frac{\lambda_{9t}}{(\alpha_{it}+D_{it})}.
\end{aligned}$$

The final step is to substitute for the true student signal, S^I_{ijt}, using Equation (8) and $\hat{a}'_{jt}$ and $\hat{u}'_{ijt}$ from the previous derivation. After some manipulation, we arrive at

$$\begin{aligned}
\hat{a}''_{jt} &= (\lambda_{7t}\lambda_{1t} + \lambda_{8t}\lambda_{2t})\,\hat{a}_{jt} - \lambda_{8t}\hat{u}_{ijt} + (\lambda_{7t}\lambda_{3t} - \lambda_{8t}\lambda_{4t} + \lambda_{11t}d_{jt})\,a_{jt} \\
&+ \lambda_{11t}\alpha_{it}u_{ijt} + (\lambda_{7t}\lambda_{5t} - \lambda_{8t}\lambda_{6t})\,v^J_{ijt} + \lambda_{11t}v^I_{ijt}
\end{aligned}$$

and

$$\begin{aligned}
\hat{u}''_{ijt} &= -(\lambda_{9t}\lambda_{1t} + \lambda_{10t}\lambda_{2t})\,\hat{a}_{jt} + \lambda_{10t}\hat{u}_{ijt} + (\lambda_{10t}\lambda_{4t} - \lambda_{9t}\lambda_{3t} + \lambda_{12t}d_{jt})\,a_{jt} \\
&+ \lambda_{12t}\alpha_{it}u_{ijt} + (\lambda_{10t}\lambda_{6t} - \lambda_{9t}\lambda_{5t})\,v^J_{ijt} + \lambda_{12t}v^I_{ijt}.
\end{aligned}$$

Beliefs at the end of period t are now expressed as functions of objects we know ($\hat{a}_{jt}$, $\hat{u}_{ijt}$, a_{jt}, and u_{ijt}) and independent, normally distributed signal noise (v^J_{ijt} and v^I_{ijt}). λ_{1t} through λ_{12t} are complicated combinations of the underlying parameters of interest (plus Σ^j_t, Σ^i_t, and Σ^{ij}_t).

Belief Equations: The final step is to convert the time t, continuous parental beliefs into probabilities associated with discrete belief responses. We observe $\mathbb{1}(\hat{u}''_{ijt} + \hat{a}''_{jt} > k_{1t})$, an indicator that parents believe their child's skill is above the overall skill average. We also observe $\mathbb{1}(\hat{u}''_{ijt} > k_{2t})$, an indicator that parents believe their child's skill is above the school average skill level.

To construct the probability of observing these discrete responses, we first derive the expression for $\hat{u}''_{ijt} + \hat{a}''_{jt}$, which is given below:

$$
\begin{aligned}
\hat{u}''_{ijt} + \hat{a}''_{jt} \;\; &= \;\; (\lambda_{7t}\lambda_{1t} + \lambda_{8t}\lambda_{2t} - \lambda_{9t}\lambda_{1t} - \lambda_{10t}\lambda_{2t})\,\hat{a}_{jt} + (\lambda_{10t} - \lambda_{8t})\,\hat{u}_{ijt} \\
&+ \;\; ((\lambda_{7t}\lambda_{3t} - \lambda_{8t}\lambda_{4t} + \lambda_{11t}d_{jt}) + (\lambda_{10t}\lambda_{4t} - \lambda_{9t}\lambda_{3t} + \lambda_{12t}d_{jt}))\,a_{jt} + (\lambda_{11t}\alpha_{it} + \lambda_{12t}\alpha_{it})\,u_{ijt} \\
&+ \;\; ((\lambda_{7t}\lambda_{5t} - \lambda_{8t}\lambda_{6t}) + (\lambda_{10t}\lambda_{6t} - \lambda_{9t}\lambda_{5t}))\,v^J_{ijt} + (\lambda_{11t} + \lambda_{12t})\,v^I_{ijt}
\end{aligned}
$$

The variance of $\hat{u}''_{ijt} + \hat{a}''_{jt}$ conditional on $\hat{a}_{jt}$, $\hat{u}_{ijt}$, a_{jt}, and u_{ijt} is given by,

$$
Var(\hat{u}''_{ijt} + \hat{a}''_{jt}|\hat{a}_{jt}, \hat{u}_{ijt}, a_{jt}, u_{ijt}) = ((\lambda_{7t}\lambda_{5t} - \lambda_{8t}\lambda_{6t}) + (\lambda_{10t}\lambda_{6t} - \lambda_{9t}\lambda_{5t}))^2 + (\lambda_{11t} + \lambda_{12t})^2
$$

where we make use of the fact that the variances of v^J_{ijt} and v^I_{ijt} are one. Using the two previous expressions, we can express the probability that a parent believes their child is above average relative to similarly aged children in the population:

$$
\text{Prob}(\hat{u}''_{ijt} + \hat{a}''_{jt} > k_{1t}) = 1 - \Phi\left(\tilde{k}_{1t} - \Gamma_{1t}\hat{a}_{jt} - \Gamma_{2t}\hat{u}_{ijt} - \Gamma_{3t}a_{jt} - \Gamma_{4t}u_{ijt}\right)
$$

where

$$\begin{aligned}
\Gamma_{1t} &= \frac{(\lambda_{7t}\lambda_{1t} + \lambda_{8t}\lambda_{2t} - \lambda_{9t}\lambda_{1t} - \lambda_{10t}\lambda_{2t})}{\sqrt{V_{Gt}}} \\
\Gamma_{2t} &= \frac{(\lambda_{10t} - \lambda_{8t})}{\sqrt{V_{Gt}}} \\
\Gamma_{3t} &= \frac{((\lambda_{7t}\lambda_{3t} - \lambda_{8t}\lambda_{4t} + \lambda_{11t}d_{jt}) + (\lambda_{10t}\lambda_{4t} - \lambda_{9t}\lambda_{3t} + \lambda_{12t}d_{jt}))}{\sqrt{V_{Gt}}} \\
\Gamma_{4t} &= \frac{(\lambda_{11t}\alpha_{it} + \lambda_{12t}\alpha_{it})}{\sqrt{V_{Gt}}} \\
\tilde{k}_{1t} &= \frac{k_{1t}}{\sqrt{V_{Gt}}} \\
V_{Gt} &= ((\lambda_{7t}\lambda_{5t} - \lambda_{8t}\lambda_{6t}) + (\lambda_{10t}\lambda_{6t} - \lambda_{9t}\lambda_{5t}))^2 + (\lambda_{11t} + \lambda_{12t})^2 .
\end{aligned}$$

Using a similar approach, we can derive the probability that a parent believes their child's skill is above average relative to their classmates. The variance of $\hat{u}''_{ijt}$ conditional on $\hat{a}_{jt}$, $\hat{u}_{ijt}$, a_{jt}, and u_{ijt} is given by,

$$Var(\hat{u}''_{ijt}|\hat{a}_{jt}, \hat{u}_{ijt}, a_{jt}, u_{ijt}) = (\lambda_{10t}\lambda_{6t} - \lambda_{9t}\lambda_{5t})^2 + \lambda_{12t}^2$$

The relevant probability is now given by

$$\text{Prob}(\hat{u}''_{ijt} > k_{2t}) = 1 - \Phi\left(\tilde{k}_{2t} - \Gamma_{5t}\hat{a}_{jt} - \Gamma_{6t}\hat{u}_{ijt} - \Gamma_{7t}a_{jt} - \Gamma_{8t}u_{ijt}\right)$$

where

$$\begin{aligned}
\Gamma_{5t} &= \frac{(\lambda_{9t}\lambda_{1t} + \lambda_{10t}\lambda_{2t})}{\sqrt{V_{Lt}}} \\
\Gamma_{6t} &= \frac{\lambda_{10t}}{\sqrt{V_{Lt}}} \\
\Gamma_{7t} &= \frac{(\lambda_{10t}\lambda_{4t} - \lambda_{9t}\lambda_{3t} + \lambda_{12t}d_{jt})}{\sqrt{V_{Lt}}} \\
\Gamma_{8t} &= \frac{\lambda_{12t}\alpha_{it}}{\sqrt{V_{Lt}}} \\
\tilde{k}_{2t} &= \frac{k_{2t}}{\sqrt{V_{Lt}}} \\
V_{Lt} &= (\lambda_{10t}\lambda_{6t} - \lambda_{9t}\lambda_{5t})^2 + \lambda_{12t}^2.
\end{aligned}$$

The newly defined Γ_{1t} through Γ_{8t} can be directly identified using probit regressions. We have just shown that these objects are known functions of the underlying parameters: α_{jt}, α_{it}, d_{jt}, D_{jt}, D_{it}. Notice that these objects are also functions of the elements of the covariance matrix Σ_t, which is a function of the signal parameters from time $t-1$. However, we only want to use a single cross-section to identify the time t parameters. This means that we also need to identify the elements of Σ_t. Using Γ_{1t} to Γ_{8t} (8 elements), we need to identify α_{jt}, α_{it}, d_{jt}, D_{jt}, D_{it} and Σ_t (8 elements). Ideally, we would like to invert the mapping between the 8 observed coefficients and the 8 unknown parameters.[37] Unfortunately this mapping is highly non-linear in the underlying parameters and an exact solution cannot be found. We discuss our numerical approach in the next section.

B.2 Numerical Identification

Although we are unable to solve analytically for α_{jt}, α_{it}, d_{jt}, D_{jt}, D_{it}, and Σ_t as functions of Γ, we can show numerically that this system of equations yields a unique solution. The basic approach is to choose values for α_{jt}, α_{it}, d_{jt}, D_{jt}, D_{it}, and Σ_t and then construct the complicated Γ coefficients using the formulas provided above. Finally, we use an optimization procedure to search for values of α_{jt}, α_{it}, d_{jt}, D_{jt}, D_{it}, and Σ_t that generate "estimated" values for Γ that are as close to the truth as possible. We repeat this process for 1,000 different parameter vectors, and always find that the optimization procedure identifies the true values for the parameters of interest. The remainder of this section describes how one iteration of this process works.

The first step is to choose values for α_{jt}, α_{it}, d_{jt}, D_{jt}, and D_{it}. The natural starting point is to set them close to the values that we estimate. However, because these are estimates, we perturb these values each iteration with random shocks. Thus our starting

[37] We do not incorporate k_{1t} and k_{2t} into this system of equations as they simply introduce two additional parameters and two additional equations. If the eight Γ coefficients identify the parameters of interest, then the estimates of $\tilde{k}_{1t}$ and $\tilde{k}_{2t}$ will identify k_{1t} and k_{2t}.

values each iteration are given by:

$$\begin{aligned}
\ln \alpha_{jt} &= \ln(0.6 + 0.1z_1) \\
\ln \alpha_{it} &= \ln(0.025 + 0.005z_2) \\
d_{jt} &= -0.4 + 0.05z_3 \\
D_{jt} &= 4.0 + 0.25z_4 \\
D_{it} &= 2.0 + 0.25z_5
\end{aligned}$$

where z_k's are iid draws from a standard normal distribution.[38] Note that we do not choose the Σ_t parameters freely since they are functions of the underlying signal, transition, and initial skill distribution parameters. We generate values for Σ_t that would be consistent with the true values Σ_t would take at the start of period two given the randomly selected signal parameters.[39] Once the signal and MSE parameters are set, we construct true values for Γ using the formulas presented in the previous section.

The next step is to behave as if the only thing we observe is Γ and attempt to recover the parameters of interest. We first choose a set of starting values for each of the parameters according to the following:

$$\begin{aligned}
\ln \dot{\alpha}_{jt} &= \ln \alpha_{jt} + 0.5z_6 \\
\ln \dot{\alpha}_{it} &= \ln \alpha_{it} + 0.35z_7 \\
\dot{d}_{jt} &= d_{jt} + 0.3z_8 \\
\dot{D}_{jt} &= D_{jt} + 0.5z_9 \\
\dot{D}_{it} &= D_{it} + 0.5z_{10} \\
\ln \dot{\Sigma}_t^j &= \ln \Sigma_t^j + 0.2z_{11} \\
\ln \dot{\Sigma}_t^i &= \ln \Sigma_t^i + 0.2z_{12} \\
\dot{\Sigma}_t^{ij} &= \Sigma_t^{ij} + 0.2z_{13}
\end{aligned}$$

[38] α_{jt} and α_{it} are constrained to be positive.

[39] Calculating values for Σ_t requires not just signal parameters, but values for the transition parameters and initial skill parameters. Again, we choose values similar to our estimate: $\pi_j = 0.2$, $\pi_i = 0.5$, $\rho_i = 0.95$, $\rho_j = 0.1$, $\sigma_i = 0.05$, and $\sigma_j = 0.05$.

where again the z_k's are iid draws from a standard normal distribution, and the dot indicates that these are starting values and not the true parameters. Given a set of values for the parameters we can generate $\hat{\Gamma}$ using the formulas from the previous section. We then search for the set of parameter values that minimizes $\sum_{i=1}^{8}(\Gamma_i - \hat{\Gamma}_i)^2$. This search restricts α_i, α_j, Σ_t^j, and Σ_t^i to be positive, and Σ_t^{ij} to be negative.

Because our starting values can be quite distant from the truth, we do not always reach the true values after completion of the first search algorithm. After each iteration of the search algorithm, we check the value of our objective function. If it is more than 1e-10, we draw a new set of starting values and re-estimate. Eventually we always obtain an objective function with a value less than 1e-10, and the resulting parameter values are always within 0.1% of the true values.

In the above numerical exercise, we illustrate how to identify the signal parameters for a given period t. For estimation, we need to impose further restrictions since we do not observe beliefs in every period, nor do we observe lagged beliefs in the initial period. In periods where we do not observe beliefs, the outcome variable in the reduced form regression is missing and thus we cannot identify the eight relevant coefficients. In the first period, we do not observe lag beliefs and thus can only identify four reduced form coefficients. These four coefficients are not enough to identify the period one signal parameters (α_{j1}, α_{i1}, d_{j1}, D_{j1}, D_{i1}).[40] We resolve these issues by assuming that the signal parameters are time invariant over the first three model periods, but allow the signal parameters to vary in the final period.

Finally, when illustrating that we can identify the time t parameters numerically, we assume that the true value for α_j is strictly positive. Estimates of the model using the ECLS-K data, however, indicate that the estimated value for α_j is very close to zero over the first three periods. The issue is that the above numerical identification argument breaks down when $\alpha_j = 0$, meaning that the signal parameters from the first three periods are not identified. Yet, we can resolve this problem relying on knowledge of the period four mean square error parameters. In period 4, α_j is significantly larger than zero, implying that we can identify the signal parameters and Σ_4. Using the production function parameters and Σ_4, we can recover what Σ_3'' must have been. This essentially provides us with three

[40] Σ_1 is pinned down by the unconditional distribution of skill.

additional coefficients to match when identifying the period three signal and mean square error (Σ_3) parameters since Σ_3'' is a function of exactly these parameters. Thus, we can again show numerical identification, including identifying the fact that $\alpha_j = 0$ in period three.

C Estimation

In this appendix we derive the likelihood equation for the estimation of our parental learning model. In order to compute the conditional expectation of our observations, we use a non-linear state space approach. We characterize the conditional distribution of our unobserved state variables, derive what the likelihood function is once we condition on those state variables and then integrate them away. Although the model is non-linear (the belief-related data is a step function of the underlying state variables), we approximate the conditional expectations using linear projections in the spirit of a generalized Kalman filter approach. Using a Monte Carlo analysis, we show that this approximation method works quite well, with limited evidence of significant biases in the estimates using simulated data.[41]

We proceed in several steps. First, we describe how to derive the joint distribution of the unobserved state variables at the end of the first period. We then discuss how the state space transitions between any two periods. Then, we discuss our measurement system and construct the recursive algorithm for our approximated Kalman filter. Finally, we derive the decomposition of the likelihood function. Our focus is on a model with homogenous signals. At the close of this section we briefly discuss how our approach is altered when the signal loading factors and signal distortions are heterogeneous with respect to x_{ij} and x_j.

C.1 Joint Density of State Space for $t = 1$

The unit of analysis in our model is a school. Each school is inhabited by N_j students who are described by their skill levels and parental beliefs. All children in school j face the same school average skill level and experience common unobserved shocks to this quantity. The state space for school j at the end any period t is composed of both student level and school level unobservables. There are three state variables per student within school j, u_{ijt}, $\hat{u}''_{ijt}$, and $\hat{a}''_{jt}$. These describe the true skill level and actual parental beliefs for student i after receiving time t skill signals. There is an additional state variable common to all students in school j, a_{jt}. This state variable reflect the true average skill level in school j. The total number of state variables in school j is given by $3 \times N_j + 1$.

The first step is to derive the joint distribution of the unobserved state variables in

[41] Results available upon request.

school j for $t = 1$. Once we obtain these distributions, will be able to calculate the probability of observing a particular set of measures for period $t = 1$. Our strategy is to separately describe the structure of each state variable at the end of period one, and then jointly characterize these variables.

As discussed in Section 4.1.1, we assume that the initial draws of school average skill and individual skill are determined according to

$$\begin{aligned} a_{j1} &= \beta_{j1}x_j + \epsilon_{j1} \text{ where } \epsilon_{j1} \sim N(0, \pi_j) \\ a_{ij1} &= \beta_{i1}(x_{ij} - x_j) + a_{j1} + \epsilon_{ij1} \text{ where } \epsilon_{ij1} \sim N(0, \pi_i), \end{aligned}$$

where we also define $u_{ij1} = a_{ij1} - a_{j1}$. Parents form priors over these objects according to,

$$\begin{pmatrix} \hat{a}_{j1} \\ \hat{u}_{ij1} \end{pmatrix} = \begin{pmatrix} \mu + \beta_{j1}x_j \\ \beta_{i1}\left(x_{ij} - x_j\right) \end{pmatrix}$$

with a perceived mean squared error given by

$$\Sigma_1 = \begin{pmatrix} \Sigma_1^j & \Sigma_1^{ij} \\ \Sigma_1^{ij} & \Sigma_1^i \end{pmatrix} = \begin{pmatrix} \pi_j & 0 \\ 0 & \pi_i \end{pmatrix}$$

where $\hat{u}_{ij1} = \hat{a}_{ij1} - \hat{a}_{j1}$.

Prior to the end of period 1, parents receive two signals regarding school average skill and individual skill. The structure of these signals is described in Section 4.1.2. Using these signals, parents update their priors. Following the steps in Section B.1, it is straightforward to show that the posterior for parental beliefs in period one can be expressed as:

$$\begin{pmatrix} \hat{a}''_{j1} \\ \hat{u}''_{ij1} \end{pmatrix} = \begin{pmatrix} \delta_1\mu_j\left(x\right) + \delta_3\mu_i\left(x\right) \\ \delta_2\mu_j\left(x\right) + \delta_4\mu_i\left(x\right) \end{pmatrix} + \begin{pmatrix} \delta_5 \\ \delta_6 \end{pmatrix} a_{j1} + \begin{pmatrix} \delta_7 \\ \delta_8 \end{pmatrix} u_{ij1} + \begin{pmatrix} \delta_9 v^I_{ij1} + \delta_{11} v^J_{ij1} \\ \delta_{10} v^I_{ij1} + \delta_{12} v^J_{ij1} \end{pmatrix}$$

where the δ's are functions of the underlying parameters, $\mu_j\left(x\right) = \left(\mu + \beta_{j1}x_j\right)$, and $\mu_i\left(x\right) = \beta_{i1}\left(x_{ij} - x_j\right)$. To derive the joint density, we need to fully characterize the distribution of the above beliefs. It is straightforward to see that they are normally distributed, with a mean vector given by:

$$E\begin{pmatrix} \hat{a}''_{j1} \\ \hat{u}''_{ij1} \end{pmatrix} = \begin{pmatrix} (\delta_1+\delta_5)\,\mu_j(x) + (\delta_3+\delta_7)\,\mu_i(x) \\ (\delta_2+\delta_6)\,\mu_j(x) + (\delta_4+\delta_8)\,\mu_i(x) \end{pmatrix}.$$

To arrive at the above equation we make use of the fact that in the first period $E(a_{j1}) = \mu_j(x)$ and $E(u_{ij1}) = \mu_i(x)$. Below we present the full covariance structure for parental beliefs in school j.

We can now describe the joint distribution of the unobserved state variables for school j in period 1:

$$\left(\underbrace{\begin{matrix} u_{1j1} \\ \hat{u}''_{1j1} \\ \hat{a}''_{j_1 1} \\ \vdots \\ u_{Nj1} \\ \hat{u}''_{Nj1} \\ \hat{a}''_{j_N 1} \\ a_{j1} \end{matrix}}_{\theta_{j1}}\right) \sim N\left(\underbrace{\begin{pmatrix} \mu_1(x) \\ (\delta_2+\delta_6)\,\mu_j(x) + (\delta_4+\delta_8)\,\mu_1(x) \\ (\delta_1+\delta_5)\,\mu_j(x) + (\delta_3+\delta_7)\,\mu_1(x) \\ \vdots \\ \mu_N(x) \\ (\delta_2+\delta_6)\,\mu_j(x) + (\delta_4+\delta_8)\,\mu_N(x) \\ (\delta_1+\delta_5)\,\mu_j(x) + (\delta_3+\delta_7)\,\mu_N(x) \\ \mu_j(x) \end{pmatrix}}_{\mu_{j1}}, \underbrace{\begin{pmatrix} \Omega_{11} & \cdots & \Omega_{1N} & \Omega_1^{j\prime} \\ \vdots & \ddots & & \vdots \\ \Omega_{1N} & & \Omega_{NN} & \Omega_N^{j\prime} \\ \Omega_1^j & \cdots & \Omega_N^j & \Omega^j \end{pmatrix}}_{\Omega_1}\right)$$

where

$$\Omega_{ll} = \begin{pmatrix} \pi_i & \delta_8\pi_i & \delta_7\pi_i \\ \delta_8\pi_i & \delta_6^2\pi_j + \delta_8^2\pi_i + \delta_{10}^2 + \delta_{12}^2 & \delta_5\delta_6\pi_j + \delta_7\delta_8\pi_i + \delta_9\delta_{10} + \delta_{11}\delta_{12} \\ \delta_7\pi_i & \delta_5\delta_6\pi_j + \delta_7\delta_8\pi_i + \delta_9\delta_{10} + \delta_{11}\delta_{12} & \delta_5^2\pi_j + \delta_7^2\pi_i + \delta_9^2 + \delta_{11}^2 \end{pmatrix} \forall\, l \in \{1, N\},$$

$$\Omega_{lk} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \delta_6^2\pi_j & \delta_6\delta_5\pi_j \\ 0 & \delta_6\delta_5\pi_j & \delta_5^2\pi_j \end{pmatrix} \forall\, k \neq l \in \{1, N\},$$

$$\Omega_l^j = \begin{pmatrix} 0 & \delta_6\pi_j & \delta_5\pi_j \end{pmatrix} \forall\, l \in \{1, N\},$$

and

$$\Omega^j = \pi_j.$$

μ_{j1} and Ω_1 reflect the mean and variance of the unobserved state variables at the end of period one. We will use these matrices as initial conditions for the Kalman filter.

As noted above, the state space for school j contains $3 \times N_j + 1$ unobserved variables. The vector of state variables is ordered such that the skill deviations and parental beliefs for students 1 through N come first, followed by the true average skill in school j. The mean of the unobserved vector of state variables follows the same structure, and the covariance matrix captures the relationships between the unobserved state variables. The off-diagonal elements of Ω are driven by the fact that a_{j1}, the true school average in period 1, impacts the beliefs of all parents. Thus, the unobservable driving the school average, ϵ_{j1}, will enter all of these state variables generating a correlation in parental beliefs within school j.

C.2 State Space Transitions

In the previous section, we derive the joint density of the unobserved state variables at the end of period $t = 1$. The joint density is given below,

$$\theta_{j1} \sim N\left(\mu_{j1}, \Omega_1\right).$$

In this section we show how these state variables transition between the end of any period $t-1$ and the end of period t. Between these two points, the state space is altered in two ways. First, child skill evolves according to a skill transition equation. Second, parents receive new skill signals that will alter their beliefs, and hence the state space. In the first step we will show that

$$\check{\theta}_{jt} \sim N\left(\mu_{jt} + \Upsilon_t \theta_{jt-1}, \check{\Omega}_t\right),$$

which is the vector of the state variables before parents receive period t information about their children. In the second step we will show that

$$\theta_{jt} \sim N\left(\Xi_t \check{\theta}_{jt}, \Omega_t\right).$$

Below we derive the updated density of the unobserved state space accounting for these transitions where we show how to derive the conditional means and variances reported above.

C.2.1 Skill Evolution

To describe how the unobserved states evolve, we rely on the skill transition equations that determine skill dynamics. We repeat these equations below, first for the student, followed by the school average:

$$a_{ijt} = \mu_{0t} + \beta_{jt} x_j + \beta_{it} (x_{ij} - x_j) + \rho_{it} a_{ijt-1} + \rho_{jt} a_{jt-1} + \epsilon_{jt} + \epsilon_{ijt}$$

and

$$a_{jt} = E_j (a_{ijt}) = \mu_{0t} + \beta_{jt} x_j + (\rho_{it} + \rho_{jt}) a_{jt-1} + \epsilon_{jt}.$$

The transition function for local skill deviations is simply the difference between the two equations above and is given by,

$$u_{ijt} = \beta_{it} (x_{ij} - x_j) + \rho_{it} u_{ijt-1} + \epsilon_{ijt}.$$

The equations for a_{jt} and u_{ijt} describe precisely the link between these initial time t state variables and the previous period state variables. However, we still need to illustrate what these transitions imply for the evolution of parental beliefs.

Parents are assumed to know how skills evolve, and hence will alter their beliefs according to the technology. Denoting parental expectations by $E_i(\cdot)$, parent predictions are given by

$$\begin{aligned} E_i(u_{ijt}) = \hat{u}_{ijt} &= \mu_i (x) + \rho_{it} \hat{u}''_{ijt-1} \\ E_i(a_{jt}) = \hat{a}_{jt} &= \mu_j (x) + (\rho_{it} + \rho_{jt}) \hat{a}''_{jt-1}. \end{aligned}$$

The above discussion outlines how the key state variables transition according to the skill transition functions. Below we aggregate these equations and show the general evolution for the full set of $3 \times N_j + 1$ state variables:

$$\underbrace{\begin{pmatrix} u_{1jt} \\ \hat{u}_{1jt} \\ \hat{a}_{j_1 t} \\ \vdots \\ u_{Njt} \\ \hat{u}_{Njt} \\ \hat{a}_{j_N t} \\ a_{jt} \end{pmatrix}}_{\breve{\theta}_{jt}} = \underbrace{\begin{pmatrix} \mu_1(x) \\ \mu_1(x) \\ \mu_j(x) \\ \vdots \\ \mu_N(x) \\ \mu_N(x) \\ \mu_j(x) \\ \mu_j(x) \end{pmatrix}}_{\mu_{jt}} + \underbrace{\begin{pmatrix} \Upsilon_{1t} & \mathbf{0} & \cdots & \mathbf{0} & 0 \\ \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & 0 \\ \vdots & \mathbf{0} & & & \\ \mathbf{0} & \cdots & \mathbf{0} & \Upsilon_{1t} & 0 \\ 0 & \cdots & 0 & 0 & \Upsilon_{2t} \end{pmatrix}}_{\Upsilon_t} \underbrace{\begin{pmatrix} u_{1jt-1} \\ \hat{u}''_{1jt-1} \\ \hat{a}''_{j_i t-1} \\ \vdots \\ u_{Njt-1} \\ \hat{u}''_{Njt-1} \\ \hat{a}''_{j_N t-1} \\ a_{jt-1} \end{pmatrix}}_{\theta_{jt-1}} + \begin{pmatrix} \epsilon_{1jt} \\ 0 \\ 0 \\ \vdots \\ \epsilon_{Njt} \\ 0 \\ 0 \\ \epsilon_{jt} \end{pmatrix}$$

where $\mathbf{0}$ is a 3x3 matrix of zeros,

$$\Upsilon_{1t} = \begin{pmatrix} \rho_{it} & 0 & 0 \\ 0 & \rho_{it} & 0 \\ 0 & 0 & \rho_{jt} + \rho_{it} \end{pmatrix},$$

and

$$\Upsilon_{2t} = \rho_{it} + \rho_{jt}.$$

Υ_t is a $(3 \times N_j + 1)$ by $(3 \times N_j + 1)$ matrix with an upper left block diagonal component that is $(3 \times N_j)$ by $(3 \times N_j)$. This matrix, along with the mean shift captured by μ_{jt}, describe the deterministic components of the skill transitions. The forecast step of the Kalman filter will utilize these two matrices. The new mean of the state variables will simply be the mean from the prior period transformed according to the equations above. The new variance of the unobserved state variables will depend on the variance of the prior state vector, plus the uncertainty induced by the skill shocks. The variance of the innovations to the state variables is given below,

$$var \begin{pmatrix} u_{1jt} \\ \hat{u}_{1jt} \\ \hat{a}_{j_1 t} \\ \vdots \\ u_{Njt} & |past \\ \hat{u}_{Njt} \\ \hat{a}_{j_N t} \\ a_{jt} \end{pmatrix} = \underbrace{\begin{pmatrix} \check{\Lambda}_{1t} & \mathbf{0} & \cdots & 0 \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \check{\Lambda}_{1t} & \\ 0 & \cdots & 0 & \check{\Lambda}_{2t} \end{pmatrix}}_{\check{\Lambda}_t}$$

where

$$\check{\Lambda}_{1t} = \begin{pmatrix} \sigma_{it} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\check{\Lambda}_{2t} = \sigma_{jt}.$$

The variance of $\check{\theta}_{jt}$, represented by $\check{\Omega}_t$, is then given by $\Upsilon_t \Omega_{t-1} \Upsilon_t' + \check{\Lambda}_t$. The above formulas fully characterize the joint density of the unobserved state variables post skill evolution. Next we describe how this system evolves in response to time t skill signals.

C.2.2 Time t Skill Signals

Within period t, parents receive school and student skill signals. Similar to period one, we can express the posterior of parental beliefs in the following manner,

$$\begin{pmatrix} \hat{a}''_{jt} \\ \hat{u}''_{ijt} \end{pmatrix} = \begin{pmatrix} \delta_{1t}\hat{a}_{jt} + \delta_{3t}\hat{u}_{ijt} + \delta_{5t}a_{jt} + \delta_{7t}u_{ijt} + \delta_{9t}v^J_{ijt} + \delta_{11t}v^I_{ijt} \\ \delta_{2t}\hat{a}_{jt} + \delta_{4t}\hat{u}_{ijt} + \delta_{6t}a_{jt} + \delta_{8t}u_{ijt} + \delta_{10t}v^J_{ijt} + \delta_{12t}v^I_{ijt} \end{pmatrix}$$

where again the δ's are functions of the underlying signal parameters. Appendix B.1 provides the detailed steps to arrive at the above equation.

The joint density of the unobserved state variables at the end of time t can now be expressed as a function of the joint density of the unobserved state variables following skill evolution. The equation below describes this relationship,

$$\underbrace{\begin{pmatrix} u_{1jt} \\ \hat{u}''_{1jt} \\ \hat{a}''_{j_1t} \\ \vdots \\ u_{Njt} \\ \hat{u}''_{Njt} \\ \hat{a}''_{j_Nt} \\ a_{jt} \end{pmatrix}}_{\theta_t} = \underbrace{\begin{pmatrix} \Xi_{1t} & \mathbf{0} & \cdots & 0 \\ \mathbf{0} & \ddots & \mathbf{0} & 0 \\ \mathbf{0} & \cdots & \Xi_{1t} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\Xi_t} \underbrace{\begin{pmatrix} u_{1jt} \\ \hat{u}_{1jt} \\ \hat{a}_{j_1t} \\ \vdots \\ u_{Njt} \\ \hat{u}_{Njt} \\ \hat{a}_{j_Nt} \\ a_{jt} \end{pmatrix}}_{\breve{\theta}_t} + \begin{pmatrix} 0 \\ \delta_{10t}v^J_{1jt} + \delta_{12t}v^I_{1jt} \\ \delta_{9t}v^J_{1jt} + \delta_{11t}v^I_{1jt} \\ \vdots \\ 0 \\ \delta_{10t}v^J_{Njt} + \delta_{12t}v^I_{Njt} \\ \delta_{9t}v^J_{Njt} + \delta_{11t}v^I_{Njt} \\ 0 \end{pmatrix}$$

where

$$\Xi_{1t} = \begin{pmatrix} 1 & 0 & 0 \\ \delta_{8t} & \delta_{4t} & \delta_{2t} \\ \delta_{7t} & \delta_{3t} & \delta_{1t} \end{pmatrix}.$$

$\mathbf{\Xi_t}$ is a $(3 \times N_j + 1)$ by $(3 \times N_j + 1)$ matrix with an upper left block diagonal component that is $(3 \times N_j)$ by $(3 \times N_j)$. The estimation algorithm will use this matrix in the forecasting step of the Kalman filter. The mean of the state variables at the end of period t will simply be the mean at the start of period transformed according to $\mathbf{\Xi_t}$. The variance will depend on the prior variance of the state vector, plus the uncertainty induced by the additional signals. The variance of the innovations to the state variables is given below,

$$var \begin{pmatrix} u_{1jt} \\ \hat{u}''_{1jt} \\ \hat{a}''_{j_1t} \\ \vdots \\ u_{Njt} \\ \hat{u}''_{Njt} \\ \hat{a}''_{j_Nt} \\ a_{jt} \end{pmatrix} | past = \underbrace{\begin{pmatrix} \Lambda_{1t} & \mathbf{0} & \cdots & 0 \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \Lambda_{1t} & \\ 0 & \cdots & & 0 \end{pmatrix}}_{\Lambda_t}$$

where

$$\Lambda_{1t} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \delta_{10t}^2 + \delta_{12t}^2 & \delta_{9t}\delta_{10t} + \delta_{11t}\delta_{12t} \\ 0 & \delta_{9t}\delta_{10t} + \delta_{11t}\delta_{12t} & \delta_{9t}^2 + \delta_{11t}^2 \end{pmatrix}$$

The variance of θ_{jt}, represented by Ω_t, is then given by $\Xi_t \check{\Omega}_t \Xi_t' + \Lambda_t$. The joint density of the unobserved state variables at the end of period t is fully characterized by the above equations.

C.3 Measurements, Kalman Filter, and Likelihood Construction

The previous sections describe the joint density and transition of the unobserved variables describing child skill and parental beliefs. The vector of state variables starts with

$$\theta_{j1} \sim N\left(\mu_{j1}, \Omega_1\right),$$

and then evolves over time according to

$$\theta_{jt} \sim N\left(\Xi_t \mu_{jt} + \Xi_t \Upsilon_t \theta_{t-1}, \Xi_t\left(\Upsilon_t \Omega_{t-1} \Upsilon_t' + \check{\Lambda}_t\right)\Xi_t' + \Lambda_t\right)$$

where we have combined the two steps described in the previous section. Recall that the vector of state variables for school j at the end of period t is defined as:

$$\theta_{jt} = \left(a_{1jt}, \hat{u}_{1jt}'', \hat{a}_{j_1t}'', \ldots, a_{Njt}, \hat{u}_{Njt}'', \hat{a}_{j_Nt}'', a_{jt}\right)'.$$

While we don't observe θ_{jt} directly, each period four measures are available for every child in school j:

$$h_{ijt} = \left(m_{ijt}, r_{ijt}, \mathbb{1}(\hat{u}_{ijt}'' + \hat{a}_{jt}'' > k_{1ij}), \mathbb{1}(\hat{u}_{ijt}'' > k_{2ij})\right)$$

where m_{ijt} and r_{ijt} are math and reading scores, and the indicator functions reflect parental global beliefs and local beliefs respectively. Define $k_{1ijt} = k_{1t} - \beta_j^G x_{ij} - \beta_j^G x_j$ and $k_{2ijt} = k_{2t} - \beta_j^L x_{ij} - \beta_j^L x_j$ following Equation (16).

We estimate our model using a likelihood function approach. In order to decompose the likelihood function, we use an algorithm similar to a Kalman filter. The key difference in our approach is that some of our measures are non-linear, and therefore, we need to extend the calculation of the linear projection to accommodate this feature. One approximation that we continue to make is that the conditional expectation of the state space is normally distributed and therefore completely described by the mean and variance of the state variables. While this is not exactly true when some of the measures are non-linear, we find that this approximation does not introduce biases in our estimation.

For the derivation of our non-linear extension of the Kalman filter, we first introduce some additional notation. Define the linear projection $\hat{\theta}_{jt+1|t} = E\left(\theta_{jt+1}|h_j^t\right)$ where h_j^t is the collection of all measurements in school j up to time t. The mean squared error is $P_{t+1|t} = E\left[\left(\theta_{jt+1} - \hat{\theta}_{jt+1|t}\right)\left(\theta_{jt+1} - \hat{\theta}_{jt+1|t}\right)'\right]$. Finally, the initial condition for the unobserved state variables in school j is determined by

$$
\begin{aligned}
\hat{\theta}_{j1|0} &= E\left(\theta_{j1}\right) = \mu_{j1} \\
P_{1|0} &= E\left[\left(\theta_{j1} - \hat{\theta}_{j1|0}\right)\left(\theta_{j1} - \hat{\theta}_{j1|0}\right)'\right] = \Omega_1
\end{aligned}
$$

Note that the initial conditions are defined according to the distribution of the state space at the end of period one, as outlined in Section C.2.1.

Assume that the linear projection and mean squared error for period $t-1$, $\hat{\theta}_{jt|t-1}$ and $P_{t|t-1}$ are known. We are going to show how to derive next period's linear projection and mean squared error, $\hat{\theta}_{jt+1|t}$ and $P_{t+1|t}$. This, jointly with the initial conditions, gives us a recursive method to describe in all periods the linear projection of the state variables and its mean squared error. We first describe the updating step that produces $\hat{\theta}_{jt|t}$ and $P_{t|t}$ from $\hat{\theta}_{jt|t-1}$ and $P_{t|t-1}$. Then we describe the forecasting step that produces $\hat{\theta}_{jt+1|t}$ and $P_{t+1|t}$ from $\hat{\theta}_{jt|t}$ and $P_{t|t}$.

For ease of exposition and computation, we split the updating step into two smaller steps. First, we update the linear projections to include the period t continuous measures. Second, we update the linear projections to include the period t discrete measures.

C.3.1 Updating with Continuous Measures

We denote the measurement system for the continuous test score measures in school j by,

$$h_{jt}^C = \mu_t^C + H_t^C \theta_{jt} + w_t = \begin{pmatrix} 0 \\ \mu_{1t}^R \\ \vdots \\ 0 \\ \mu_{1t}^R \end{pmatrix} + \begin{pmatrix} \kappa & 0 & \cdots & 0 \\ 0 & \kappa & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \kappa \end{pmatrix} \begin{pmatrix} u_{1jt} \\ \hat{u}_{1jt}'' \\ \hat{a}_{j_1 t}'' \\ \vdots \\ u_{Njt} \\ \hat{u}_{Njt}'' \\ \hat{a}_{j_N t}'' \\ a_{jt} \end{pmatrix} + \begin{pmatrix} e_{1jt}^M \\ e_{1jt}^R \\ \vdots \\ e_{Njt}^M \\ e_{Njt}^R \end{pmatrix}$$

where

$$\kappa = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_t^R & 0 & 0 \end{pmatrix}$$

and

$$E\left(w_t w_t'\right) = R_t = I_{N_j} \otimes \begin{pmatrix} \sigma_t^m & 0 \\ 0 & \sigma_t^r \end{pmatrix}$$

To set the scale of the skill unobservable, we normalize the loading on the math score to one and assume the math score is centered on the truth. The reading measure of skill can have a different mean (μ_{1t}^R) and loading (λ_t^R). The noise in the math and reading measures have variances equal to σ_t^m and σ_t^r respectively.

Our forecast of the time t continuous measures using information available up to time $t-1$ is

$$\hat{h}_{jt|t-1}^C = E\left(h_{jt}^C | h_j^{t-1}\right) = \mu_t^C + H_t^C \hat{\theta}_{jt|t-1}.$$

The error of this forecast is

$$E\left[\left(h_{jt}^C - \hat{h}_{jt|t-1}^C\right)\left(h_{jt}^C - \hat{h}_{jt|t-1}^C\right)'\right] = H_t^{C\prime} P_{t|t-1} H_t^C + R_t.$$

To update the state of the system incorporating these continuous measures, we use linear projection formulas:

$$\hat{\theta}^C_{jt|t} = \hat{\theta}_{jt|t-1} + P_{t|t-1}H^C_t \left(H^{C\prime}_t P_{t|t-1} H^C_t + R_t\right)^{-1} \left(h^C_{jt} - \hat{h}^C_{jt|t-1}\right)$$

where $\hat{\theta}^C_{jt|t} = E\left(\theta_{jt}|h^{t-1}_j, h^C_{jt}\right)$. The new mean square error is given by

$$P^C_{t|t} = P_{t|t-1} - P_{t|t-1}H^C_t \left(H^{C\prime}_t P_{t|t-1} H^C_t + R_t\right)^{-1} H^{C\prime}_t P_{t|t-1}$$

where $P^C_{t|t} = E\left[\left(\theta_{jt} - \hat{\theta}^C_{jt|t}\right)\left(\theta_{jt} - \hat{\theta}^C_{jt|t}\right)'\right]$.

C.3.2 Updating with Discrete Measures

In this section we illustrate how to update the linear projection of our state variables using the information contained in the discrete belief measures, $h^D_{ijt} = \left(\mathbb{1}(\hat{u}''_{ijt} + \hat{a}''_{jt} > k_{1ijt}), \mathbb{1}(\hat{u}''_{ijt} > k_{2ijt})\right)$.

The state of the system after incorporating the continuous measures is fully characterized by $\hat{\theta}^C_{jt|t}$ and $P^C_{t|t}$. We only describe how the updating works for local beliefs $h^{D_L}_{ijt} = \mathbb{1}(\hat{u}''_{ijt} > k_{2ijt})$. Updating using global beliefs is done in a similar fashion, making sure that local beliefs are now part of the conditioning set.

It is useful to first introduce the true vector of continuous local beliefs for school j in period t:

$$b^L_{jt} = H^{D_L}\theta_{jt} = \begin{pmatrix} b_{1jt} \\ b_{2jt} \\ \vdots \\ b_{Njt} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & & & & & & \ddots & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} u_{1jt} \\ \hat{u}''_{1jt} \\ \hat{a}''_{j_1 t} \\ \vdots \\ u_{Njt} \\ \hat{u}''_{Njt} \\ \hat{a}''_{j_N t} \\ a_{jt} \end{pmatrix}$$

While these continuous beliefs are unobserved, we can describe their joint distribution conditional on the contemporaneous continuous measures and all past measures. Approximating the conditional expectations with a normal distribution implies that

$$b_{jt}^L \sim N\left(H^{D_L}\hat{\theta}_{jt|t}^C, H^{D_L}P_{t|t}^C H^{D_L\prime}\right)$$

With a slight abuse of notation, the forecasted vector of binary local belief measures takes the following form,

$$\hat{h}_{jt|t}^{D_L} = \Pr\left(h_{jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right) = \begin{pmatrix} \Pr(b_{1jt}^L > k_{21jt}|h_j^{t-1}, h_{jt}^C) \\ \vdots \\ \Pr(b_{N_jjt}^L > k_{2Njt}|h_j^{t-1}, h_{jt}^C) \end{pmatrix}$$

where

$$\Pr\left(b_{ijt}^L > k_{2ijt}|h_j^{t-1}, h_{jt}^C\right) = 1 - \Phi\left(\frac{k_{2ijt} - E(\hat{u}_{ijt}''|\hat{\theta}_{jt|t}^C, P_{t|t}^C)}{\sqrt{Var(\hat{u}_{ijt}''|\hat{\theta}_{jt|t}^C, P_{t|t}^C)}}\right).$$

The standard formulae for updating the linear projection are:

$$\begin{aligned} \hat{\theta}_{jt|t}^{C,D^L} &= \hat{\theta}_{jt|t}^C + K_t\left(h_{jt}^{D_L} - \hat{h}_{jt|t}^{D_L}\right) \\ P_{t|t}^{C,D^L} &= P_{t|t}^C - K_t var\left(h_{jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right)K_t' \end{aligned}$$

where

$$K_t = cov\left(\theta_{jt}, h_{jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right) \times var\left(h_{jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right)^{-1}$$

The challenge with discrete measures is constructing both elements of the Kalman gain, K_t. We first describe how to construct the variance matrix for the local beliefs, and then show how to construct the covariance between the local beliefs and the unobserved state.

Each element of $var\left(h_{jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right)$ is either $var\left(h_{ijt}^{D_L}|h_j^{t-1}, h_{jt}^C\right)$ or $cov\left(h_{ijt}^{D_L}, h_{i'jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right)$. Since $h_{ijt}^{D_L}$ is binary, we can express the variance as

$$var\left(h_{ijt}^{D_L}|h_j^{t-1}, h_{jt}^C\right) = \Pr\left(h_{ijt}^{D_L} = 1|h_j^{t-1}, h_{jt}^C\right) \times \left(1 - \Pr\left(h_{ijt}^{D_L} = 1|h_j^{t-1}, h_{jt}^C\right)\right).$$

Exploiting again the binary nature of $h_{ijt}^{D_L}$, we can write the covariance between any two local belief measures as

$$cov\left(h_{ijt}^{D_L}, h_{i'jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right) = \Pr(h_{ijt}^{D_L} = 1, h_{i'jt}^{D_L} = 1|h_j^{t-1}, h_{jt}^C) - \Pr\left(h_{ijt}^{D_L}|h_j^{t-1}, h_{jt}^C\right) \times \Pr\left(h_{i'jt}^{D_L} = 1|h_j^{t-1}, h_{jt}^C\right)$$

We already presented the formula to calculate $\Pr\left(h_{ijt}^{D_L}|h_j^{t-1}, h_{jt}^C\right)$. Calculating the joint distribution for two beliefs is similar in that the relevant parameters for the bivariate normal are contained in $\hat{\theta}_{jt|t}^C$ and $P_{t|t}^C$.

The next step us to characterize $cov\left(\theta_{jt}, h_{jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right)$. Consider the covariance between the state vector and one measure of parental local beliefs, $cov\left(\theta_{jt}, h_{ijt}^{D_L}|h_j^{t-1}, h_{jt}^C\right)$. We can express this covariance in the following manner

$$cov\left(\theta_{jt}, h_{ijt}^{D_L}|h_j^{t-1}, h_{jt}^C\right) = E\left(\theta_{jt}|h_j^{t-1}, h_{jt}^C, h_{ijt}^{D_L} = 1\right)\Pr\left(h_{ijt}^{D_L} = 1|h_j^{t-1}, h_{jt}^C\right) - \hat{\theta}_{jt|t}^C \Pr\left(h_{ijt}^{D_L} = 1|h_j^{t-1}, h_{jt}^C\right)$$

The first term on the right hand side of the above equation is the only term that we have not already defined. We can write this conditional expectation in the following manner

$$E\left(\theta_{jt}|h_j^{t-1}, h_{jt}^C, h_{ijt}^{D_L} = 1\right) = E\left(\theta_{jt}|h_j^{t-1}, h_{jt}^C, \hat{u}_{ijt}'' > k_{2ijt}\right).$$

While θ_{jt} contains $3 \times N_j + 1$ elements, we can calculate the above conditional expectation in a pairwise fashion. For the mth element of the θ_{jt} vector, define the joint distribution of $(\theta_{jt}^{(m)}, \hat{u}_{ijt}'')$ according to

$$\left(\theta_{jt}^{(m)}, \hat{u}_{ijt}''\right) \sim N\left(\left(\begin{array}{c} \mu_{\theta^{(m)}} \\ \mu_{\hat{u}} \end{array}\right), \left(\begin{array}{cc} \Sigma_{\theta^{(m)}} & \Sigma_{\theta\hat{u}} \\ \Sigma_{\theta\hat{u}} & \Sigma_{\hat{u}} \end{array}\right)\right),$$

which is just a 2 dimensional extract from $\theta_{jt} \sim N\left(\hat{\theta}_{jt|t}^C, P_{t|t}^C\right)$. The distribution of $\theta_{jt}^{(m)}$ conditional on $\hat{u}_{ijt}''$ can then be expressed as,

$$\left(\theta_{jt}^{(m)}|\hat{u}_{ijt}''\right) \sim N\left(\mu_{\theta^{(m)}} + \frac{\Sigma_{\theta\hat{u}}}{\Sigma_{\hat{u}}}\left(\hat{u}_{ijt}'' - \mu_{\hat{u}}\right), \Sigma_{\theta^{(m)}} - \frac{\Sigma_{\theta\hat{u}}^2}{\Sigma_{\hat{u}}}\right).$$

Using this conditional distribution, we can calculate $E\left(\theta_{jt}^{(m)}|h_j^{t-1}, h_{jt}^C, \hat{u}_{ijt}'' > k_{2ijt}\right)$ accordinging to

$$
\begin{aligned}
E\left(\theta_{jt}^{(m)}|\hat{u}_{ijt}''>k_{2ijt}\right) &= \frac{\int\int_{k_{2ijt}}\theta_{jt}^{(m)}dF\left(\theta_{jt}^{(m)},\hat{u}_{ijt}''\right)}{\int\int_{k_{2ijt}}dF\left(\theta_{jt}^{(m)},\hat{u}_{ijt}''\right)} \\
&= \frac{\int_{k_{2ijt}}\left(\int\theta_{jt}^{(m)}f\left(\theta_{jt}^{(m)}|\hat{u}_{ijt}''\right)d\theta_{jt}^{(m)}\right)f\left(\hat{u}_{ijt}''\right)d\hat{u}_{ijt}''}{1-\Phi\left(\frac{k_{2ijt}-\mu_{\hat{u}}}{\sqrt{\Sigma_{\hat{u}}}}\right)} \\
&= \frac{1}{1-\Phi\left(\frac{k_{2ijt}-\mu_{\hat{u}}}{\sqrt{\Sigma_{\hat{u}}}}\right)}\int_{k_{2ijt}}\left(\mu_{\theta^{(m)}}+\frac{\Sigma_{\theta\hat{u}}}{\Sigma_{\hat{u}}}\left(\hat{u}_{ijt}''-\mu_{\hat{u}}\right)\right)\frac{1}{\sqrt{2\pi}\Sigma_{\hat{u}}}\phi\left(\frac{\hat{u}_{ijt}''-\mu_{\hat{u}}}{\sqrt{\Sigma_{\hat{u}}}}\right)d\hat{u}_{ijt}'' \\
&= \left(\mu_{\theta^{(m)}}-\frac{\Sigma_{\theta\hat{u}}}{\Sigma_{\hat{u}}}\mu_{\hat{u}}\right)+\frac{\Sigma_{\theta\hat{u}}}{\Sigma_{\hat{u}}}\left(\mu_{\hat{u}}+\sqrt{\Sigma_{\hat{u}}}\frac{\phi\left(\frac{k_{2ijt}-\mu_{\hat{u}}}{\sqrt{\Sigma_{\hat{u}}}}\right)}{1-\Phi\left(\frac{k_{2ijt}-\mu_{\hat{u}}}{\sqrt{\Sigma_{\hat{u}}}}\right)}\right) \\
&= \mu_{\theta^{(m)}}+\frac{\Sigma_{\theta\hat{u}}}{\sqrt{\Sigma_{\hat{u}}}}\frac{\phi\left(\frac{k_{2ijt}-\mu_{\hat{u}}}{\sqrt{\Sigma_{\hat{u}}}}\right)}{1-\Phi\left(\frac{k_{2ijt}-\mu_{\hat{u}}}{\sqrt{\Sigma_{\hat{u}}}}\right)}
\end{aligned}
$$

The above equation provides a closed-form solution to the conditional expectation where all the terms in the equation are contained in either $\hat{\theta}_{jt|t}^{C}$ or $P_{t|t}^{C}$. Doing the above for all elements of θ_{jt} and $h_{jt}^{D_L}$ allows us to completely characterize $cov\left(\theta_{jt},h_{jt}^{D_L}|h_j^{t-1},h_{jt}^{C}\right)$.

The previous two paragraphs show how to calculate $var\left(h_{jt}^{D_L}|h_j^{t-1},h_{jt}^{C}\right)$ and $cov\left(\theta_{jt},h_{jt}^{D_L}|h_j^{t-1},h_{jt}^{C}\right)$. We are now in a position to calculate the Kalman gain associated with the local belief measures and construct $\hat{\theta}_{jt|t}^{C,D^L}$ and $P_{t|t}^{C,D^L}$, which represent the updated mean and precision of the system after incorporating $h_{jt}^{D_L}$. The next step is to incorporate $h_{jt}^{D_G}$, our measures of parental global beliefs. All of the steps are analogous to the above. At the end of that process, we arrive at $\hat{\theta}_{jt|t}$ and $P_{t|t}$.

C.3.3 Forecasting

At this point, we have updated the linear projection (and conditional expectation given our approximation) $\hat{\theta}_{jt|t}$ and its mean squared error $P_{t|t}$. The next step is to forecast the conditional distribution of the state variables in the next period: $\hat{\theta}_{jt+1|t}$ and $P_{t+1|t}$. Two things happen between the end of period t and the end of period $t+1$: (1) students accumulate skills according to the transition equation and (2) parents receive time $t+1$ signals regarding school and child skill. At the beginning of this section we already showed that this means that

$$\theta_{jt} \sim N\left(\Xi_t \mu_{jt} + \Xi_t \Upsilon_t \theta_{t-1}, \Xi_t \left(\Upsilon_t \Omega_{t-1} \Upsilon_t' + \check{\Lambda}_t\right) \Xi_t' + \Lambda_t\right)$$

This implies that

$$\begin{aligned}
\hat{\theta}_{jt+1|t} &= \Xi_{t+1}\mu_{jt+1} + \Xi_{t+1}\Upsilon_{t+1}\hat{\theta}_{jt|t} \\
P_{jt+1|t} &= \Xi_{t+1}\Upsilon_{t+1}P_{jt|t}\Upsilon_{t+1}'\Xi_{t+1}' + \Xi_{t+1}\check{\Lambda}_{t+1}\Xi_{t+1}' + \Lambda_{t+1}
\end{aligned}$$

where all objects are defined and derived in sections C.2.1 and C.2.2. With $\hat{\theta}_{jt+1|t}$ and $P_{t+1|t}$, we can calculate the likelihood of the period $t+1$ measures, incorporate them into the system, and proceed in a recursive manner.

C.3.4 Likelihood

We are now in a position to calculate all the components of the likelihood contribution stemming from school j. The equation below describes the contribution of school j:

$$\begin{aligned}
L_j &= \prod_{t=1}^{T} p_{t|t-1}\left(h_{jt}|h_j^{t-1}\right) \\
&= \prod_{t=1}^{T} p_{t|t-1}\left(h_{jt}^C|h_j^{t-1}\right) \times p_{t|t-1}\left(h_{jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right) \times p_{t|t-1}\left(h_{jt}^{D_G}|h_j^{t-1}, h_{jt}^C, h_{jt}^{D_L}\right).
\end{aligned}$$

The first element above describes the likelihood contribution of the test scores of all students of school j and it is easy to derive given that we showed that

$$h_{jt}^C \sim N\left(\mu_t^C + H_t^C \hat{\theta}_{jt|t-1}, H_t^{C\prime} P_{t|t-1} H_t^C + R_t\right),$$

which can be calculated using the multivariate normal distribution.

The second and third elements are more difficult to calculate because they involve the computation of multidimensional integrals given that parental beliefs in school j are correlated with each other. Below we discuss the calculation of $p_{t|t-1}\left(h_{jt}^{D_L}|h_j^{t-1}, h_{jt}^C\right)$. $p_{t|t-1}\left(h_{jt}^{D_G}|h_j^{t-1}, h_{jt}^C, h_{jt}^{D_L}\right)$ is calculated in analogous fashion.

In section C.3.2 we show that local beliefs are approximately normally distributed:

$$b_{jt}^{L} \sim N\left(H^{D_L}\hat{\theta}_{jt|t}^{C}, H^{D_L}P_{t|t}^{C}H^{D_L\prime}\right)$$

Employing the above distribution we can calculate the likelihood contribution associated with the discrete local belief measures. If we sort the data such that the first I^L parents believe their child is above average, then the likelihood contribution can be expressed as

$$p_{t|t-1}\left(h_{jt}^{D_L}|h_j^{t-1}, h_{jt}^{C}\right) = \int_{k_{21jt}} ... \int_{k_{2I^L jt}} \int^{k_{2I^{L}+1jt}} ... \int^{k_{2N_j jt}} d\Phi\left(\cdot|h_j^{t-1}, h_{jt}^{C}\right),$$

where k_{2ijt} are the cutoffs for the local belief discrete measures adjusted for the observable characteristics of student i. We compute this multidimensional integral numerically using an algorithm proposed in Genz (1992).

Finally, the joint likelihood across all schools is simply the product of the L_j, given below:

$$L = \prod_{j=1}^{J} L_j.$$

We maximize the log-likelihood using the Nelder-Mead algorithm in order not to rely on gradients and Hessians, although we have experimented also using a quasi-Newton method (Davidon-Fletcher-Powell) with a numerical gradient.

C.3.5 Heterogeneous Signals

As discussed in Section 4.1.2, we allow for observed heterogeneity in the signal parameters related to x_{ij} and x_j. In particular, we assume that $\alpha_{ct} = \check{\alpha}_{ct}\exp\left(\alpha_{xj}x_j + \alpha_{xi}x_{ij}\right)$ for $c \in \{j, i\}$. We model observed heterogeneity in D_{ct} for $c \in \{j, i\}$ and d_{jt} in a similar fashion. This means that the rate at which households learn and the size of the distortion vary across schools and across children in the same school. As a result the parental updating procedure outlined in Appendix B.1 will be individual in nature.

If we were to index by i and j the signal loadings and signal distortions and follow the steps in Section B.1, it is straightforward to show that the posterior for parental beliefs in period one can be expressed as:

$$\left(\begin{array}{c}\hat{a}''_{j1}\\ \hat{u}''_{ij1}\end{array}\right)=\left(\begin{array}{c}\delta_{1ij}\mu_j\left(x\right)+\delta_{3ij}\mu_i\left(x\right)\\ \delta_{2ij}\mu_j\left(x\right)+\delta_{4ij}\mu_i\left(x\right)\end{array}\right)+\left(\begin{array}{c}\delta_{5ij}\\ \delta_{6ij}\end{array}\right)a_{j1}+\left(\begin{array}{c}\delta_{7ij}\\ \delta_{8ij}\end{array}\right)u_{ij1}+\left(\begin{array}{c}\delta_{9ij}v^I_{ij1}+\delta_{11ij}v^J_{ij1}\\ \delta_{10ij}v^I_{ij1}+\delta_{12ij}v^J_{ij1}\end{array}\right)$$

where the δ's continue to be functions of the underlying parameters, but are now individual and school specific. This does not materially alter any of the previous derivations. The only difference would be that all of the δ's would be i and j specific. Additionally, the off diagonal elements of the covariance matrices for the signal updates will now contain terms such as $\delta_{6ij}\delta_{6i'j}$ instead of δ_6^2. Further details available upon request.