

On Research Partnerships and Economic Science

Flávio Cunha, Rice University · October 2026

The study described here is reported in "Estimating, Bounding, and Decomposing the Effects of a Parent-Directed School Readiness Program," with Qinyou Hu, Andrea Salvati, Kenneth I. Wolpin, and Rui Zeng.

Economists have long recognized the value of partnerships with institutions that deliver social programs. These partnerships create opportunities to evaluate interventions, estimate causal effects, and produce evidence that helps institutions make better decisions. A school district may want to know whether a program improves children's readiness for school. A workforce organization may want to understand whether its training programs improve employment and earnings. Researchers can help answer these questions.

But there is a second reason to develop research partnerships, one that receives considerably less attention.

Research partnerships can help advance economics as a science. Working with institutions that implement social programs creates opportunities to investigate economic behavior, distinguish competing explanations for observed outcomes, and test predictions of economic models that would otherwise remain beyond the reach of empirical evidence.

This possibility becomes especially important when we combine randomized experiments with structural economic models. Experiments can identify causal effects under relatively weak assumptions, and structural models can explain the mechanisms generating those effects and predict what would happen under policies that have not been implemented. But the ability to produce counterfactual predictions raises a fundamental question: how do we know whether those predictions are reliable?

I want to explore that question using an example from my own research. Over several years, my colleagues and I worked with the Alief Independent School District in Houston to evaluate JumpStart, a program designed to prepare young children for preschool. The partnership allowed us to conduct a randomized experiment, investigate the mechanisms through which the program affected children's development, and examine how much of our structural model's predictions were supported by the experimental evidence rather than imposed by its assumptions. The experience also clarified a broader opportunity: by designing experiments with structural questions in mind, researchers and their institutional partners can generate evidence that makes economic theories more vulnerable to empirical contradiction. That is an important contribution to scientific progress.

1. From Evaluating a Program to Understanding Its Mechanisms

JumpStart was designed and implemented by the Alief Independent School District to help parents prepare their three-year-old children for entry into preschool. The district built the curriculum by asking its own pre-K teachers what entering children should know, and it delivered the program

with its own staff. Parents met with a family liaison at their local elementary school three times a month over seven months, attending one session each, and received materials to support learning activities at home. The program cost roughly \$450 per family.

That last fact matters more than it appears to. Most of what we know about parent-directed programs comes from interventions designed by researchers and delivered under unusually favorable conditions, or from externally designed protocols transplanted into a new setting with their content already validated elsewhere. JumpStart is something different. It is what a scaled program looks like when the scaling institution is the school district itself.

Together with Qinyou Hu, Andrea Salvati, Kenneth Wolpin, and Rui Zeng, I conducted a randomized evaluation across three annual cohorts, drawing families from the catchment areas of all 24 district elementary schools. In total, 1,216 families entered lotteries that determined who would be offered access.

The experiment produced encouraging results, though they need a comparison to read properly. Children in the control group gained more than 18 percentage points on the district's curriculum-aligned assessment over the program period, simply by growing older. The offer of JumpStart added 7.6 percentage points to that gain, and because 78 percent of offered families completed the program, the effect on families who finished it was 9.8 percentage points. We also found increases in parental investment at home, most visibly in how often parents read to the child and in the number of children's books in the house. On a nationally normed school-readiness assessment, whose content largely falls outside the JumpStart curriculum, the effect was about 2 percentage points and not robust to corrections for attrition. The program moved what it taught, and detectably less beyond it.

That was the beginning of the investigation rather than the end of it, because knowing the program worked told us very little about why. Was it effective because parents increased their investments in their children's development? Did participation change the productivity of those investments? Did children benefit from contact with the family liaisons themselves? Did the program change aspects of parental behavior our measurements failed to capture? A further question emerged from the experiment itself: not every family offered JumpStart completed it, and we wanted to know what would happen if families who did not complete the program were induced to do so. Would their children gain as much as the children of families who completed it voluntarily?

The randomized comparison between treatment and control groups cannot answer these questions. To investigate them, we needed an economic model.

2. The Promise and Challenge of Structural Economics

Structural economic models explain observed behavior using assumptions about preferences, constraints, information, and the technologies through which decisions produce outcomes. Their appeal is that they allow economists to investigate counterfactual situations that have not been observed. A model of parental investment, for example, can describe how parents decide how much to invest in their children, how those investments contribute to skill formation, and how

participation in an intervention changes these decisions. Once estimated, the model can predict outcomes under alternative circumstances.

But structural modeling presents an epistemological challenge.

A model may fit the observed data well and still generate incorrect predictions for counterfactual situations that the data do not identify.

Consider the influential approach to validating structural models associated with Todd and Wolpin. Their work shows how a model estimated without using the experimental impacts of an intervention can be evaluated by asking whether it successfully predicts those impacts. This is an important test of specification, and a model that cannot reproduce experimentally identified effects gives us reason to question its assumptions. But successful prediction does not establish that the model will accurately predict the effects of every alternative policy.

The reason is straightforward. Randomized experiments identify particular causal effects, while structural models typically generate a much larger collection of counterfactual predictions. In JumpStart, random assignment identifies the effect of offering the program, and with the additional assumption that offers matter only through completion, it identifies the average effect of completion among families who would complete when offered. It does not identify the effect of completion among families who would decline.

That counterfactual is an extrapolation.

Two structural models could reproduce the treatment and control outcomes observed in the experiment, including its estimated causal effects, while generating substantially different predictions about the benefits of completion for families who would otherwise decline. Both might pass a predictive validation exercise based on the observed experimental outcomes, and yet they could recommend different policies for encouraging participation. This distinction is central to understanding what we learn from structural models: predicting experimentally identified outcomes provides evidence about the model's performance in those settings, not about the reliability of its predictions in settings that remain unidentified.

The question, then, is whether we can develop additional ways to evaluate the extrapolations themselves.

3. Using Bounds to Discipline Structural Extrapolation

In the JumpStart paper, we approached this problem by combining structural estimation with partial identification. We first developed and estimated a model of parental investment, skill formation, and program completion, allowing parents to differ both in their willingness to complete the intervention and in the benefits their children would receive. The model generates estimates of the average treatment effect, the effect on families who complete JumpStart, and the effect on families who would decline it. The last of these is the interesting one, because the experiment does not identify it.

Rather than stopping with the structural estimates, we asked a different question: what could we learn about these treatment effects without imposing all the assumptions of the structural model?

Partial identification provides an answer. Instead of requiring enough assumptions to pin down a single parameter value, these methods characterize the set of values consistent with the observed data and a specified collection of restrictions. In our application, random assignment, the assumption that an offer matters only through completion, and monotone treatment response together imply that the average effect of completing JumpStart lies between 7.1 and 14.0 percentage points. Monotone treatment response rules out the possibility that completing the program lowers a child's score. It is an assumption rather than a consequence of randomization, and while its testable implication is not rejected by our experiment, the experiment cannot establish that it holds for every child.

Our structural estimates of the average treatment effect fall within these bounds. We then used additional information about program completion and outcomes to construct more restrictive bounds, which required further assumptions about the relationship between observed characteristics and potential outcomes. These exercises let us examine which features of the marginal treatment effect function were restricted by empirical information and which depended primarily on the structural model's functional form.

This changes the nature of the specification exercise. Rather than asking only whether the structural model reproduces experimentally identified outcomes, we ask whether its predictions about unidentified counterfactuals are consistent with what can be learned under weaker restrictions.

There is an important asymmetry in how the results should be read. If a structural prediction lies outside a valid identified set, then the structural model and the assumptions underlying that set cannot all be correct. If the prediction lies inside the set, we have not established that the structural model is correctly specified, because many different models, including misspecified ones, may generate predictions consistent with the bounds.

Compatibility is not validation. But incompatibility is informative, and the exercise creates an opportunity for the structural model to fail a test that speaks directly to the counterfactual parameter we want to estimate.

This matters because conventional statistical uncertainty does not capture all the uncertainty involved in structural extrapolation. A structural estimate can have a small standard error while depending heavily on assumptions about behavior that the data cannot verify. By examining what weaker restrictions imply for the same counterfactual, we make the role of those assumptions visible.

Our paper does not introduce new methods for constructing bounds or estimating structural models, since both have well established foundations. The methodological contribution lies in combining them so that the sources of empirical information supporting structural predictions

become apparent. The objective is not to eliminate assumptions, which would be impossible. It is to understand what the assumptions are doing.

4. Understanding What Drives the Model's Results

We followed a similar approach when investigating the mechanisms through which JumpStart affected children's outcomes. Our structural model attributes roughly one quarter of the program's effect to the parental investments we measured. The remaining three quarters reach the child without passing through those measured investments.

This decomposition requires assumptions, because parental investment is not randomly assigned. Parents choose how much to invest, and their decisions may respond to characteristics that also influence children's development. Investment is also measured imperfectly. A regression relating children's knowledge to measured parental investment can therefore produce misleading estimates of the return to investment and, in turn, misleading conclusions about the mechanisms of the intervention. A naive mediation regression in our data attributes only 8.7 percent of the impact to measured investment, and correcting for measurement error alone brings that figure close to one quarter.

We addressed both problems within the structural model. But we also wanted to know whether the decomposition was an artifact of the model's parametric assumptions, so we conducted an instrumental variables mediation analysis that dispenses with the model's selection and distributional structure while maintaining the exclusion restriction needed to identify the return to investment. The results were remarkably similar: after correcting for measurement error and endogeneity, the mediation analysis also attributed about one quarter of the effect to measured parental investment.

We then asked how the results would change if the exclusion restriction were violated. That exercise was informative in a specific way. The evidence against complete mediation proved considerably more robust than the conclusion that the direct channel is larger than the investment channel. The direct effect remains significant under compensating violations as large as 78 percent of the estimated return to investment, and full mediation requires a violation more than twice that size.

The distinction matters, because it tells us not only what the model predicts but which of its conclusions are sensitive to the assumptions required for identification. It also helps interpret the substantive finding. An effect that does not operate through measured parental investment is not necessarily an effect that bypasses parents. Our investment measures capture activities such as reading and access to learning materials, and they say much less about the quality of parental interactions. The program may have helped parents teach their children more effectively without producing large changes in the quantities we measured. Our data cannot distinguish that possibility from other unmeasured channels, but recognizing the limitation is itself informative, because it identifies what future research would need to measure.

This is one of the advantages of combining structural modeling with experimental evidence, bounds, and sensitivity analysis. We learn not only what the model implies, but where its conclusions are supported by evidence, where they remain sensitive to assumptions, and what additional information would improve our understanding.

5. Designing Experiments to Make Economic Theories More Falsifiable

The JumpStart experience suggests a more interesting possibility. What if we designed randomized experiments from the beginning to evaluate the extrapolations generated by structural economic models?

Consider the participation decision again. In JumpStart we randomized access to the program, and families who received an offer then decided whether to complete it. Because that decision was endogenous, the experiment provided limited information about the benefits that families who declined would have received had they completed.

Suppose instead we designed an experiment with two sources of random variation. We would randomize access as before, but among families offered access we would also randomize encouragement intended to increase completion. Under appropriate assumptions, including that encouragement affects children's outcomes only through completion and that it discourages no one who would otherwise complete, this design would identify additional treatment effects. Variation in encouragement would identify effects for families whose completion decisions respond to it. Now imagine varying the intensity of encouragement so that different experimental groups face different probabilities of completion. Under a suitable selection model, those differences identify average treatment effects over different intervals of families' unobserved resistance to participation, and with sufficiently informative variation the experiment could place increasingly restrictive bounds on the marginal treatment effect function.

Why does this matter? Suppose two economic models fit the original experimental evidence equally well. Both predict similar benefits for families who voluntarily complete JumpStart, but one predicts large benefits for families reluctant to participate while the other predicts that these families would gain much less. The original randomized offer might not distinguish them. An experiment that generates exogenous variation in completion would provide evidence at precisely the participation margins where the models disagree, and the two models would then face a more demanding empirical test.

The additional randomization must of course be informative. Encouragement may affect outcomes through channels other than completion, violating the exclusion restriction. Weak encouragement may change participation too little to be useful. Dividing a fixed sample into more experimental groups reduces statistical precision. These are not arguments against designing richer experiments. They are reasons to think carefully about which experimental variation would be most informative for the economic questions we want to answer.

The larger point is that experiments can be designed not merely to estimate the causal effect of an intervention, but to distinguish competing economic theories.

We can use experimental design to increase the empirical falsifiability of structural models.

This changes how we think about the relationship between economic theory and experimentation. Instead of developing a model, estimating it, and then asking whether it reproduces outcomes in an existing experiment, we can begin by asking which predictions distinguish competing models, and then investigate what experimental variation would let us evaluate those predictions. The research design becomes part of the process of testing economic theory.

6. Why Research Partnerships Matter for Scientific Progress

None of this can be accomplished by economists working in isolation. A randomized experiment requires an institution willing and able to implement it. A more demanding design may require that institution to vary recruitment procedures, encouragement strategies, program components, or the timing and intensity of services. It may require collecting information the institution would not otherwise collect, and modifying operational procedures in ways that impose costs on staff and participants. These decisions cannot be made by researchers alone. The institution must understand why the research questions matter and how answering them can contribute to its mission.

This is one reason research partnerships should be understood as more than arrangements for obtaining data or recruiting participants. They create the institutional conditions under which particular forms of scientific inquiry become possible.

The partnership with Alief made it possible to evaluate a program developed and operated by a school district, follow families over time, measure parental investments and children's knowledge, and study participation decisions. Those data allowed us to estimate a structural model; the experimental design identified some treatment effects directly; and the combination of experimental and observational variation allowed us to construct bounds and examine which conclusions were driven by additional modeling assumptions. The original partnership therefore created opportunities to investigate questions that extended well beyond whether the program worked. A more ambitious experimental design would have created still more.

This has implications for how economists approach research partnerships. The choice of partner affects which scientific questions can be answered, as does the willingness of researchers to engage with the institution's operational constraints and the willingness of the institution to incorporate research into its decisions.

There is also a responsibility here. Social programs exist to serve people, not to provide economists with opportunities to test their theories, and research must respect the institution's mission and the interests of those it serves. But scientific inquiry and institutional improvement need not be competing objectives. An experiment designed to understand why families do not complete a program can help the institution improve participation, and the same experiment provides the variation needed to investigate selection into treatment and heterogeneous treatment effects. A study of the mechanisms through which an intervention works can help the institution decide which program components are most valuable, and it can help economists distinguish

competing theories of behavior and skill formation. The strongest partnerships create opportunities for both kinds of learning.

Their scientific value also accumulates. A sustained partnership develops shared knowledge, measurement capacity, institutional trust, and an understanding of which experimental designs are operationally feasible. Those capabilities make it possible to investigate questions that could not have been answered at the beginning of the relationship. This is a reason to think of research partnerships as investments in scientific infrastructure, whose value extends beyond the findings of any single evaluation.

7. A Broader View of Scientific Discovery

Economics has developed increasingly sophisticated tools for studying behavior and evaluating public policies. Randomized experiments provide credible identification of causal effects. Structural models organize our understanding of the behavioral mechanisms that produce those effects and allow us to consider counterfactual policies. Partial identification clarifies what can be learned under weaker assumptions.

These tools are sometimes presented as competing approaches. I believe they are considerably more powerful used together. Experiments tell us what happened when an intervention was implemented. Structural models help us understand why it happened and what might happen under alternative circumstances. Bounds help us distinguish what the evidence restricts from what the model assumes. Sensitivity analysis reveals how conclusions change when identifying assumptions are relaxed. And carefully designed experiments can generate additional information that challenges the predictions of structural models.

The objective is not to demonstrate that our preferred model is correct. The objective is to create opportunities for it to be wrong. There is a fundamental difference between building a model that successfully reproduces observed outcomes and designing research that exposes the model to demanding tests. Both activities have value, but only the second is essential to scientific progress.

Our experience with JumpStart illustrates how a research partnership makes this possible. The school district wanted to know whether its program improved children's readiness for school, and the randomized experiment helped answer that question. But the same partnership also made it possible to investigate parental investment, endogenous participation, skill formation, and the credibility of structural extrapolation, and it suggested how future experiments could be designed to distinguish competing theories more effectively.

That, to me, is the broader promise of research partnerships.

Social programs can be more than objects of economic research. When researchers and institutions collaborate to generate informative experimental variation, social programs can become instruments for advancing economic science itself.

They can help us discover not only which interventions work, but which economic explanations survive careful confrontation with evidence. And they can help us identify what we still do not know, which assumptions remain untested, and which experiments we should design next.