A Partial Identification Approach to Identifying the Determinants of Human Capital Accumulation: An Application to Teachers

Nirav Mehta

University of Western Ontario *

August 10, 2020

Abstract

This paper views teacher quality through the human capital perspective. Teacher quality exhibits substantial growth over teachers' careers, but why it improves is not well understood. I use a human capital production function nesting On-the-Job-Training (OJT) and Learning-by-Doing (LBD) and experimental variation from Glewwe et al. (2010), an output-based teacher incentive pay experiment in Kenya, to discern the presence and relative importance of these forces. The identified set for the OJT and LBD components has a simple, closed-form, solution. The results provide evidence of an LBD component, as well as an informative upper bound on the OJT component.

*`nirav.mehta@uwo.ca` I thank Enrique Martin Luccioni and Tian Liu for research assistance, and Magne Mogstad, Camillie Terrier, and Seth Richards-Shubik for their comments.

1 Introduction

Teacher quality, typically measured by a teacher's value added, is an important determinant of student achievement and economic growth (Rivkin et al., 2005; Hanushek, 2011). While the research and policy focus has typically been on cross-sectional variation in teacher quality, researchers have also documented substantial growth in quality over teachers' careers (Harris and Sass, 2011; Wiswall, 2013; Papay and Kraft, 2015). For example, using data from North Carolina, Wiswall (2013) finds that the average growth in teacher quality over 35 years of experience is equal to almost one standard deviation of the cross-sectional teacher quality among novice teachers.

While the growth in quality over teachers' careers is clearly important, little is known about why it occurs. Economists have developed main two theories to explain how workers accumulate human capital, which is then used to produce output: *On-the-Job Training* (OJT) and *Learning-by-Doing* (LBD). In the "pure OJT" model, workers (in the current application, this will be teachers) allocate work hours away from production (here, classroom teaching) to invest in their human capital (here, teachers' professional development) (Becker, 1964; Ben-Porath, 1967).[1] Notably, the pure OJT model implies a tradeoff between current and future production because of the multitask problem (Hölmstrom and Milgrom, 1991). No such tradeoff is present under the "pure LBD" model, wherein workers accumulate human capital via the process of producing output (i.e., classroom teaching) (Weiss, 1972; Killingsworth, 1982; Shaw, 1989).[2]

Knowing which of these forces are present in the technology governing human capital accumulation may be important when trying to design effective education policy. For example, if there is only an OJT component to teacher human capital accumulation then an incentive pay scheme that increased current achievement could reduce long-run teacher quality, by diverting teacher energies from human capital investments. In contrast, there is no such tradeoff in LBD. In principle, guidance could be obtained by examining the prevalence of these forces for similar occupations. However, it is notoriously difficult to separate the

[1]For an example of an OJT-intensive job, consider being a student in school.

[2]For an example of an LBD-intensive job, consider "chicken sexing", where workers gain an ability to determine the sex of baby chickens through their experience sorting chicks (McWilliams, 2018).

OJT and LBD mechanisms in general worker contexts; as Heckman et al. (2002) discuss, identification is hampered by equilibrium re-adjustments that occur in the private sector. This has left researchers and policymakers lacking even basic qualitative information about the presence of OJT or LBD forces for workers of any type, including teachers. Indeed, prior empirical research in labor economics has adopted "either-or" specifications, only allowing for either OJT or LBD as the force generating human capital. Information about even basic qualitative features of the human capital production function constitutes an important first step towards designing effective educational policy that targets teachers.

To fill this gap, this paper develops a simple, linear, framework that yields new findings about the roles played by OJT and LBD forces in the technology governing teacher human capital accumulation. The approach is a partial identification one, where the identified set is a locus of points representing the values of the OJT and LBD components of teacher human capital accumulation consistent with the data and the model; this can then be projected onto the marginal identified set for either component. I estimate the identified set using data from an experimental intervention, Glewwe et al. (2010), which studies a teacher performance pay scheme enacted in many sites across Kenya. In contrast to applications considered by the literature, labor prices are often fixed in the education sector because teachers are typically paid according to public salary schedules[3] and worker-specific output is observed, even if noisily so (i.e., teacher value added), making teacher quality trajectories a particularly attractive context for discriminating between these forces.

Two features of the intervention in Glewwe et al. (2010) are crucial to applying the novel approach developed in this paper. First is the balanced random assignment of an output-based pay scheme. Second, Glewwe et al. (2010) includes a follow-up measurement of the effects of the program. The intuition behind the identification result is simple. Suppose that achievement increased in the treatment group while the incentive scheme was in place (as was the case in Glewwe et al. 2010) and, for simplicity, that there is complete depreciation of student achievement. Experimental balance implies that the increase in achievement during the intervention must come from an increase in treatment-group teachers' time allocated to production. Then, in the absence of strong income effects, a positive treatment effect after

[3]See Podgursky and Springer (2011) for an extensive discussion of this point.

the intervention ended would indicate the presence of the LBD component.

This paper makes progress by way of several methodological innovations. First, it works with a human capital production function that nests both the OJT and LBD mechanisms, as opposed to the "either-or" specifications used in prior empirical work. This allows for standard hypothesis testing about important qualitative features, in particular, the presence of either force in the human capital production function. Second, it uses a partial identification approach that features a closed-form solution for the identified set and does not require sophisticated estimation techniques. The partial identification algorithm proceeds via an explicit characterization of the (very simple and intuitive) identified set: it is a locus characterized by a line, meaning the set-identified parameters depend only on a slope and an intercept parameter; adding a returns-to-scale-type assumption would yield point identification. This characterization sharply contrasts with the literature's lack of a proof showing how to separately identify OJT and LBD forces. Moreover, the primary input to the identified set is an experimentally estimated average treatment effect (ATE), which is known to have very appealing properties (notably that unobserved heterogeneity averages out when there is experimental balance) that are passed to the identified set via the linear framework. Thus, it is also easy to estimate the identified set and to do inference on it, which is rare in the partial identification literature (Imbens and Manski, 2004; Stoye, 2009; Tamer, 2010). Finally, the approach does not have to make assumptions regarding the optimality of input choices or outcomes observed in the data.

I find that the 90% confidence interval for the identified set is informative. Specifically, the lower bound on the LBD component is greater than zero and the upper bound on the OJT component is tighter than its uninformative level. One implication of these findings is that I reject the "pure OJT" specification, in which there is no LBD component, at the 10% significance level.

This paper provides a simple way to distinguish between the OJT and LBD forces of human capital accumulation, and uses very simple measurement (i.e., ATE estimates) to help inform researchers and policymakers about the presence and relative importance of these forces. Although a partial identification approach is typically viewed as imposing fewer assumptions than one affording point identification—here, for example, there is no assumption that observed decisions or outcomes are optimal—the approach taken in this paper assumes linear technologies for human capital and output (i.e., achievement). The linear specification for output is consistent with the overwhelming mass of specifications used by applied researchers (McCaffrey et al., 2003), but the linear human capital production function is less conventional in the human capital literature (see, e.g., Brown, 1976; Haley, 1976; Heckman, 1976). There are also substantive tradeoffs associated with adopting a linear human capital production function. First, this paper's approach is uninformative about heterogeneity in teacher quality trajectories. Closely related, bounds are sharp, but "wide"; they would likely be tighter if we made additional assumptions that allowed examination of treatment effects for different subgroups, such as younger teachers, who the literature has found to have the most substantial quality increases (Hanushek and Rivkin, 2006). Second, while this paper's approach does not impose optimality of the observed data, the linear framework actually makes it difficult to rationalize the observed behavior of teachers, which is unsatisfying.[4] The most important concern is that the linear technologies may represent a misspecification that could call into question this paper's substantive findings. However, the finding that the data evince the presence of an LBD component to human capital accumulation is arguably consistent with the conclusion of Glewwe et al. (2010), that the intervention seemed to increase teaching to the test, not underlying student knowledge.[5] Intuitively, it seems more natural that efficacy in teaching to the test would increase more under LBD than under OJT. Therefore, while the conclusion of Glewwe et al. (2010) might affect interpretation of the substantive findings of this paper, it does not diminish this paper's methodological contributions, which could be applied to other contexts.

The current paper reports on the technology of human capital production, which at best can answer ceteris paribus questions about the effects of input changes but cannot predict behavioral responses to policy interventions (Todd and Wolpin, 2003). This is similar in motivation and goals to the literature reporting on different aspects of education production

[4]Linear frameworks tend to produce corner solutions. However, interior solutions could potentially be achieved by adding in nonlinearities in the relative cost of investment for one component or, potentially, by interpreting the technology as log-linear (although this would impose homogeneity or force the researcher to use data moments other than the experimentally estimated average treatment effect).

[5]Their main evidence for this is that, in contrast to the results for the incentivized exam, student achievement for a non-incentivized exam that covered similar material did not significantly increase in the treatment group.

functions (e.g., the effect of class size on student outcomes). At the same time, this paper represents an important first step towards understanding why teacher quality grows over the course of teachers' careers, and more generally, why workers accumulate human capital. As is well understood, while the experimental variation used to estimate the parameters confers upon them appreciable internal validity, they cannot reasonably be extrapolated to other contexts without imposing additional structure.[6] A complementary structural approach that used behavioral assumptions to rationalize the observed input choices and outcomes would provide a more general quantitative understanding of the forces underlying the generation of human capital and also permit simulation of outcomes under counterfactual incentive schemes. Such an approach could also allow for alternative specifications for technologies and could explore cross-sectional heterogeneity in the growth of teacher quality (which cannot be done using this paper's approach, as it only uses first moments).

2 Data and Variables

Study Design Glewwe et al. (2010) implemented an output-based teacher incentive pay scheme in Kenya, which provided bonuses to teachers and headmasters at schools where students did well on standardized exams. The treatment group was exposed to the incentive scheme in 1998 and 1999, which I refer to as the "active-treatment" years. In addition to being observed during active-treatment years, outcomes for both the treatment and control groups were also observed for both a pre-treatment year (1997), and a post-active-treatment year (2000).[7] Table 1 summarizes the study design.

Variables I report only on the variables that are needed for the bounds developed in this paper; readers are referred to Glewwe et al. (2010) for additional details. I measure output using the standardized achievement test presented in the main results of Glewwe et al. (2010), which is an average of (seven) standardized subject-specific achievement tests,

[6] This concern is particularly salient in light of the aforementioned possibility of misspecification, due to the linear framework.

[7] Another potential candidate dataset is Muralidharan and Sundararaman (2011). Unfortunately, that study does not measure output the year after the intervention, which precludes application of the approach developed here.

Table 1: Design of Glewwe et al. (2010) Teacher Incentive Pay Experiment

	Output-based incentive pay?			
Year	1997	1998	1999	2000
Treatment	✗	✓	✓	✗
Control	✗	✗	✗	✗

Note: "✓" indicates that members of the group (e.g., "Treatment") received output-based incentive pay in that year and "✗" indicates that the group did not receive incentive pay that year.

where standardization was performed with respect to the means and standard deviations of the control group. As in Glewwe et al. (2010), I exclude one of the districts (Teso), which did not offer the achievement test in the first year (Glewwe et al., 2010, p. 210, footnote 3). The data also include measures of whether teachers were present at schools and in classrooms during site visits. Crucially for the algorithm used to compute the bounds, the experiment was balanced; of particular importance, the mean test scores in the pre-active-treatment year were balanced between the control and treatment groups (Glewwe et al., 2010, p. 212, Table 1).

3 Model

I now develop the model that provides the foundation for the partial identification of the components governing human capital accumulation. The technologies for output (here, achievement) and human capital production are both linear. This enables a simple constructive identification proof and calculation of bounds using only ATE estimates, which have remarkably good properties (e.g., unobserved heterogeneity averages out). Linear production functions are the norm for modeling academic achievement in the economics of education. However, linear human capital production functions and output production functions are not the norm in labor economics; this may be, in part, because they tend to produce corner solutions for input choices. To show how the specification formulated here relates to other specifications used in labor economics, Appendix A shows how a model featuring curvature in both technologies relates to the specification developed here.

Let the periods be indexed by $t = 0, 1, 2$, where $t = 0$ is the pre-treatment period, $t = 1$ is the "active-treatment" period, in which teachers in the treatment group are offered output-based incentives, and $t = 2$ is the post-active-treatment period, where teachers in either treatment or control group no longer are offered the incentives. As I discuss in Section 4.1, I pool the two active-treatment years (see Table 1) to map the empirical application to the three-period structure of the model. Note that I do not refer to $t = 1$ as the "treatment period" because, post active treatment, teachers in the treatment group may have been affected by their decisions made during active treatment.

The human capital of teacher i in period t, θ_{it}, follows the law of motion

$$\theta_{it} = \delta_\theta \theta_{it-1} + \delta_I I_{it-1} + \delta_h h_{it-1}, \tag{1}$$

where θ_{it-1} is teacher i's human capital last period, I_{it-1} is teacher i's on-the-job investment in human capital last period, and h_{it-1} is the hours teacher i spent on production last period. Output produced by teacher i teaching student j in period t, y_{ijt}, which is measured by performance on a standardized achievement test, follows a standard value-added specification (Hanushek, 1979):[8]

$$y_{ijt} = \beta_\theta \theta_{it} + \beta_h h_{it} + \beta_y y_{jt-1} + \epsilon_{ijt}, \tag{2}$$

where y_{jt-1} is student j's prior test achievement, $\beta_\theta \theta_{it} + \beta_h h_{it}$ is teacher i's value added to student j's achievement in period t, and ϵ_{ijt} is an IID productivity shock, realized after the choice of h_{it} has been made in period t.[9]

Hours spent producing output, h_{it}, and on-the-job investment, I_{it} are rival, as captured by the time constraint

$$h_{it} + I_{it} + l_{it} = H, \tag{3}$$

where l_{it} is teacher i's leisure time in period t. Although, as discussed below, there was no evidence of an average labor supply response to the incentive pay scheme, I explicitly model

[8] Recent research supports the view that controlling for prior achievement, as is done in a value-added model, does a reasonably good job of controlling for unobserved prior inputs (see, e.g., Kinsler, 2012; Chetty et al., 2014).

[9] Note that, while the above notation does not accommodate student characteristics, these are on average the same between the treatment and control groups and, therefore, average out in the conditional mean differences used throughout.

labor supply for the potential application of this approach to other contexts. Because a specific functional form for preferences is not required for the analysis, I do not make this function explicit.[10]

As in the literature, identification is difficult because none of $(\theta_{it}, I_{it}, h_{it})$ are directly observed or computable given the data. Note that the parameters, $(\delta_\theta, \delta_I, \delta_h, \beta_\theta, \beta_h, \beta_y)$, all relate to either the technology governing the law of motion for human capital or the technology describing the production function. I assume that $\delta_k \in D_k$, where $D_k = [0, \overline{\delta}]$, for $k = \theta, I, h$, and that $\beta_k \in B_k$, where $B_k = [0, \overline{\beta}]$, for $k = \theta, h, y$. The lower bound of zero for each parameter captures the natural assumption means that inputs are, ceteris paribus, not unproductive. The upper bound for each parameter (i.e., $\overline{\delta}$ or $\overline{\beta}$) is taken to be large; as will become apparent, this value does not affect this paper's main, qualitative, findings.[11] I define the "pure OJT" specification for the human capital production function as $\delta_I > 0$ and $\delta_h = 0$. Analogously, in the "pure LBD" specification, we have $\delta_h > 0$ and $\delta_I = 0$. Some scenarios representing different possibilities for true combinations of (δ_h, δ_I) are illustrated in Figure 1 (bounds for the *identified set* of (δ_h, δ_I) are derived in Section 3.1). For example, the point labeled "pure OJT", on the horizontal axis, features a positive OJT component, with no LBD component, in contrast to the point labeled "pure LBD", on the vertical axis. The interior point, labeled "OJT and LBD both present", represents the possibility that teachers accumulate human capital via both OJT and LBD components; in contrast, the "pure" versions of the OJT and LBD human capital production functions are mutually exclusive.

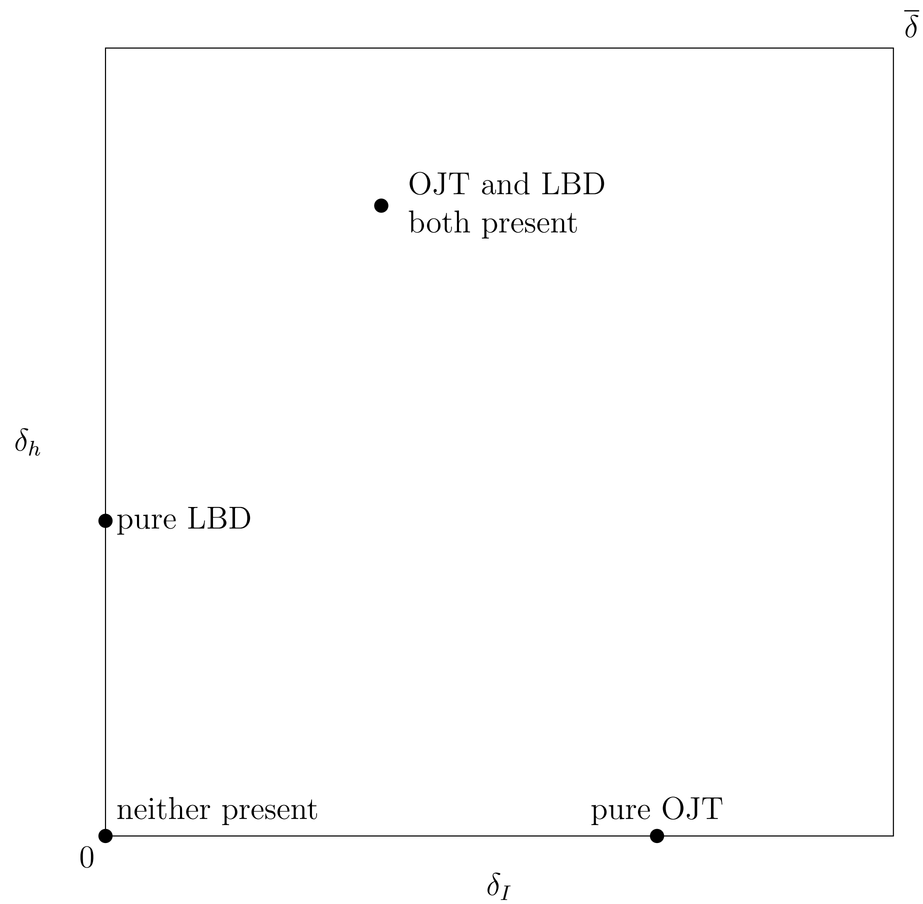
3.1 Derivation of Bounds

This section develops bounds for (δ_h, δ_I) that only depend on period-specific estimates of the ATE on student achievement and an estimate of the persistence of student knowledge (which could be estimated using the same dataset, or obtained from another source). The goals is to obtain sharp bounds on (δ_h, δ_I), i.e., to return the potentially set-identified values

[10] As discussed below, all that is required is the absence of strong income effects.

[11] The specification developed in Appendix A could potentially provide guidance about reasonable values of $\overline{\delta}$.

Figure 1: Examples of OJT and LBD Specifications



of (δ_h, δ_I) consistent with the assumptions and data. For example, the sharp bound on δ_I would increase the lower bound for δ_I as much as possible, which could reject the pure LBD specification. Again, identification is difficult because $(\theta_{it}, I_{it}, h_{it})$ are not directly observed or computable given the data.

Define the mean difference between the treatment and control groups for variable z in period t as $\Delta z_t := \overline{z_t^T} - \overline{z_t^C}$, where $\overline{z_t^T}$ and $\overline{z_t^C}$ respectively denote the treatment and control group means of z in t.

First, note that in any period the constant time endowment H implies that

$$\Delta h_t + \Delta I_t + \Delta l_t = 0. \tag{4}$$

For the pre-treatment period, $t = 0$, the mean difference in achievement between the treatment and control groups is

$$\Delta y_0 = \beta_\theta \underbrace{\Delta\theta_0}_{=0} + \beta_h \underbrace{\Delta h_0}_{=0} + \beta_y \underbrace{\Delta y_{-1}}_{=0} + \underbrace{\Delta\epsilon_0}_{=0} = 0, \tag{5}$$

i.e., a balanced experimental design implies there will be no average difference in the pre-treatment average scores, which is indeed consistent with Glewwe et al. (2010). I assume the experiment was implemented with high fidelity in the subsequent periods, which allows me to omit $\Delta\epsilon_t$ hereafter.

For the active-treatment period, $t = 1$, we have

$$\Delta y_1 = \beta_\theta \underbrace{\Delta\theta_1}_{=0} + \beta_h \Delta h_1 + \beta_y \underbrace{\Delta y_0}_{=0} = \beta_h \Delta h_1. \tag{6}$$

Equation (6) shows that the difference between treatment and control achievement in the active-treatment period can only come from the change in mean working hours Δh_1, as the fact that θ_{it} depends on lagged inputs and balance between the treatment and control groups implies that $\Delta\theta_0 = 0$, while, as shown in eq. (5), balance between the treatment and control groups implies that $\Delta y_0 = 0$.

The mean difference in achievement between the treatment and control groups for the post-active-treatment period, $t = 2$, is

$$\begin{aligned}
\Delta y_2 &= \beta_\theta \Delta\theta_2 + \beta_h \Delta h_2 + \beta_y \Delta y_1 \\
&= \beta_\theta \delta_I \Delta I_1 + \beta_\theta \delta_h \Delta h_1 + \beta_h \Delta h_2 + \beta_y \Delta y_1 \\
&= \beta_\theta \delta_I \underbrace{[-\Delta l_1 - \Delta h_1]}_{=\Delta I_1,\text{ from eq. (4)}} + [\beta_\theta \delta_h + \beta_h \beta_y] \Delta h_1 + \beta_h \Delta h_2 \\
&= -\beta_\theta \delta_I \Delta l_1 + \left[\frac{\beta_\theta \delta_h - \beta_\theta \delta_I + \beta_h \beta_y}{\beta_h}\right] \underbrace{\Delta y_1}_{\beta_h \Delta h_1,\text{ from eq. (6)}} + \beta_h \Delta h_2,
\end{aligned} \tag{7}$$

which uses $\Delta\theta_1 = 0$ to go from the first to the second line and eq. (6) to go from the second to the third line. Equation (7) shows that δ_h and δ_I both appear in the same equation, meaning that, in general, only a locus of (δ_h, δ_I) will be identified, and, further, that many other variables appear in the same equation: the parameters $(\beta_\theta, \beta_h, \beta_y)$ and, even after using eq. (6) to eliminate Δh_1, the quantities $(\Delta l_1, \Delta h_2)$.[12]

I proceed by first pinning down some parameters using "atheoretical information", meaning either assumptions utilizing a priori assumptions about bounds for the parameter space or quantities that can be estimated maintaining the assumed forms of the achievement production technology and the technology governing the evolution of teacher human capital. First note that β_θ can be normalized to 1 because it is attached to the latent variable θ_t, which can only ever be estimated from residual information (e.g., teacher-year fixed effects). A statistical technique (treatment-year fixed effects in the production function, coupled with the statistical assumption that the experiment was appropriately balanced) yields consistent estimates of β_y.[13] Without further assumptions or, perhaps, direct estimation, it is not clear what value β_h should take—the principle issue is that the scales of (h, I, l) are not pinned down by data—so I will analyze bounds for a given value of β_h. However, because $\beta_h \geq 0$ (by assumption) and because, as I argue just below, $\Delta h_1 \geq 0$ and, as I show below in Section 4.1, Δy_1 is significantly greater than zero (i.e., there is a positive effect of the intervention during the active-treatment period), it is reasonable to treat $\beta_h > 0$, based on eq. (6). As I show below, qualitative tests (e.g., those of the pure OJT or LBD specifications) are possible just knowing $\beta_h > 0$.

I next proceed by working through implications of some simplifying behavioral assumptions. If there do not exist strong income effects, wherein the level of θ_t or prior compensation strongly affect how teachers allocate their time, then the increase in output-based pay for the treatment group has two useful implications: (i) hours allocated to production will not decrease in the active-treatment period, i.e., $\Delta h_1 \geq 0$ and (ii) in the post-active-treatment period, even though there may be a mean difference in θ_2 between the treatment and control groups, h_2 will be similar in both the treatment and control groups, meaning it would be natural to treat $\Delta h_2 = 0$.[14]

[12] Note that δ_θ does not appear in eq. (7). Intuitively, this parameter measures how differences in teacher human capital emanating from differences in inputs from two periods ago affect production today; the balanced experimental design means these differences are all zero, causing δ_θ to drop out.

[13] This is because, for teacher-student pair ij in group g (i.e., treatment or control), we have $\mathrm{E}_{ij\in g}\left[y_{ijt}\right] = \mathrm{E}_{ij\in g}\left[\beta_\theta \theta_{ijt} + \beta_h h_{it}\right] + \beta_y \mathrm{E}_{ij\in g}\left[y_{ijt-1}\right] + \underbrace{\mathrm{E}_{ij\in g}\left[\epsilon_{ijt+1}\right]}_{=0}$ and $\mathrm{E}_{ij\in g}\left[\beta_\theta \theta_{ijt} + \beta_h h_{it}\right]$ can be measured by putting a parameter in front of a group-period indicator variable.

[14] Note that $I_t = 0$ if $\delta_I = 0$ (which would be the case under the pure LBD specification), regardless

Table 2: Atheoretical and Behavioral Pinnings-down

Parameter	Value	Notes
β_θ	1	Normalization of unobserved teacher human capital distribution
β_h	> 0	
Δh_1	≥ 0	Income effects not strong
Δh_2	≈ 0	Income effects not strong

The above discussion is summarized in Table 2. Using these values, eq. (7) becomes

$$\Delta y_2 = \left[\frac{\delta_h - \delta_I}{\beta_h} + \beta_y\right] \Delta y_1 - \delta_I \Delta l_1,$$

which we can rearrange to get the following expression for the LBD component δ_h as a function of the OJT component δ_I and remaining parameters:

$$\delta_h = \beta_h \left[\frac{\Delta y_2}{\Delta y_1} - \beta_y\right] + \delta_I \left[1 + \frac{\beta_h \Delta l_1}{\Delta y_1}\right], \tag{8}$$

i.e., eq. (8) characterizes the identified set, $d_h \times d_I$. We can find the marginal identified set for a parameter by projecting the locus characterized by eq. (8) onto the relevant axis, e.g., d_h can be obtained by projecting $d_h \times d_I$ onto the δ_h-axis. Note that, even with the pinnings-down summarized in Table 2, the relationship between (δ_h, δ_I) depends on $(\beta_h, \beta_y, \Delta y_1, \Delta y_2, \Delta l_1)$; Section 4.1 discusses how values for these variables were obtained.

We can understand the restrictions afforded by eq. (8) by considering some cases, where for the sake of illustration assume $\Delta y_1 > 0$ (as is shown in Section 4.1, this is consistent with the data) and that $\beta_h = 1$. To start, suppose that $\Delta l_1 = 0$ and suppose that $\frac{\Delta y_2}{\Delta y_1} - \beta_y = 0$; this corresponds to case (a) in Table 3. Although the composition of work hours is naturally treated as unobserved, total work hours (i.e., the sum of on-the-job investment and production) is commonly measured. Intuitively, the increase in test score coming via the positive effect of Δh_1 on Δy_1 entirely accounts for Δy_2. The first line of eq. (7) then implies that $\Delta \theta_2 = 0$, i.e., teacher human capital post-active-treatment is on average the

of treatment status, if the teacher valued leisure; this would imply that $\Delta I_t = 0$ under the pure LBD specification.

same in the treatment and control groups. Further, $\Delta l_1 = 0$ implies that $\Delta I_1 = -\Delta h_1$, meaning the only way to satisfy eq. (8) is for $\delta_I = \delta_h$. That is, any increase in δ_I can satisfy the condition by a concomitant increase in δ_h. Without further information on either of these parameters, we cannot shrink the identified set. This scenario corresponds to the dashed, 45-degree, line in Figure 2. Projecting the identified set onto each axis, we can see that the marginal identified set for either δ_I or δ_h (depicted by the dashed lines just outside that parameter's axis) has not shrunk at all, because any feasible value of, e.g., δ_I, can be rationalized by the same value for the coefficient on the other input to teacher human capital (in this example, δ_h). Such bounds are typically referred to as being "uninformative".

Next consider case (b), which differs from case (a) in that $\frac{\Delta y_2}{\Delta y_1} - \beta_y < 0$, i.e., teacher human capital is lower in the treatment group, post active treatment ($\Delta\theta_2 < 0$). Because we still have $\Delta I_1 = -\Delta h_1$, the fact that $\Delta\theta_2 < 0$ means that we can rule out very low δ_I—the OJT component—and very high δ_h—the LBD component. Intuitively, if the net effect of increasing time spent on production on teacher human capital is negative, knowing that $\Delta I_1 = -\Delta h_1$ implies that the OJT parameter must be larger than the LBD one. At the same time, the slope of eq. (8) is unaffected due to the one-to-one tradeoff between different time uses in the budget constraint and the fact that there is no change in leisure time implies there is a one-to-one tradeoff between δ_I and δ_h. This case is depicted by the dotted lines in Figure 2. Though not pictured here, note that increasing β_h, making time spent in production more productive, would represent a further downward shift of the (δ_h, δ_I) locus, further tightening the lower bound for δ_I and the upper bound for δ_h.

In cases (a) and (b), there was no average difference in leisure time between the treatment and control groups (i.e., $\Delta l_1 = 0$). Consider now case (c), where we start from case (b) but now assume that $\Delta l_1 < 0$. Here, we know that $\Delta I_1 > -\Delta h_1$, i.e., the absolute difference in investment is smaller than the absolute difference in hours spent in production. As shown in the dash-dotted lines in Figure 2, this rotates the locus eq. (8) downward from the intercept (which was already negative, as the starting point was case (b)), increasing the lower bound on δ_I and decreasing the upper bound on δ_h. Intuitively, all else equal, a smaller change in investment must be coupled with a relatively bigger technological effect of investment (δ_I) to rationalize the same data.

The cases discussed above are not exhaustive. For example, the signs of $\frac{\Delta y_2}{\Delta y_1} - \beta_y$ or Δl_1 could be opposite to those considered in cases (b) or (c), in which case the sharp bounds would be adjusted; an example of this is case (d), in the bottom row. Also, adding additional assumptions would create tighter sharp bounds on the identified set. For example, if we further assumed that $\delta_I + \delta_h \leq 1$ (a natural constraint if we viewed eq. (1) as a log-linearized Cobb-Douglas technology exhibiting non-increasing returns; see Appendix A for discussion), the identified set would be contained by a triangle, instead of by the rectangle pictured in Figure 2. Appendix B discusses how this additional information, which is embedded in an assumption instead of provided by additional data, would yield tighter bounds.

Finally, note that the levels of h, I, l are not identified without further information. If we assumed that H was, e.g., total time per year this would fix the scale of the time variables. But because they are not observed, one can only ever deduce $\delta_I I_t$ and $\delta_h h_t$. Information on the sign of changes in l_t can help discern whether changes in output manifest via changes in the input levels (i.e., choices) or the technological parameters (i.e., identification of eq. (1)).

Table 3: Examples of Marginal Identified Set Cases

				Marginal Identified Sets				
Case	Δl_1	β_h	$\frac{\Delta y_2}{\Delta y_1} - \beta_y$	$\min d_I$	$\max d_I$	$\min d_h$	$\max d_h$	Bounds Notes
(a)	0	1	0	$\min D_I$	$\max D_I$	$\min D_h$	$\max D_h$	Uninformative
(b)	0	1	< 0	$> \min D_I$	$\max D_I$	$\min D_h$	$< \max D_h$	Informative
(c)	< 0	1	< 0	$>> \min D_I$	$\max D_I$	$\min D_h$	$<< \max D_h$	Tighter than (b)
(d)	0	1	> 0	$\min D_I$	$< \max D_I$	$> \min D_h$	$\max D_h$	Informative

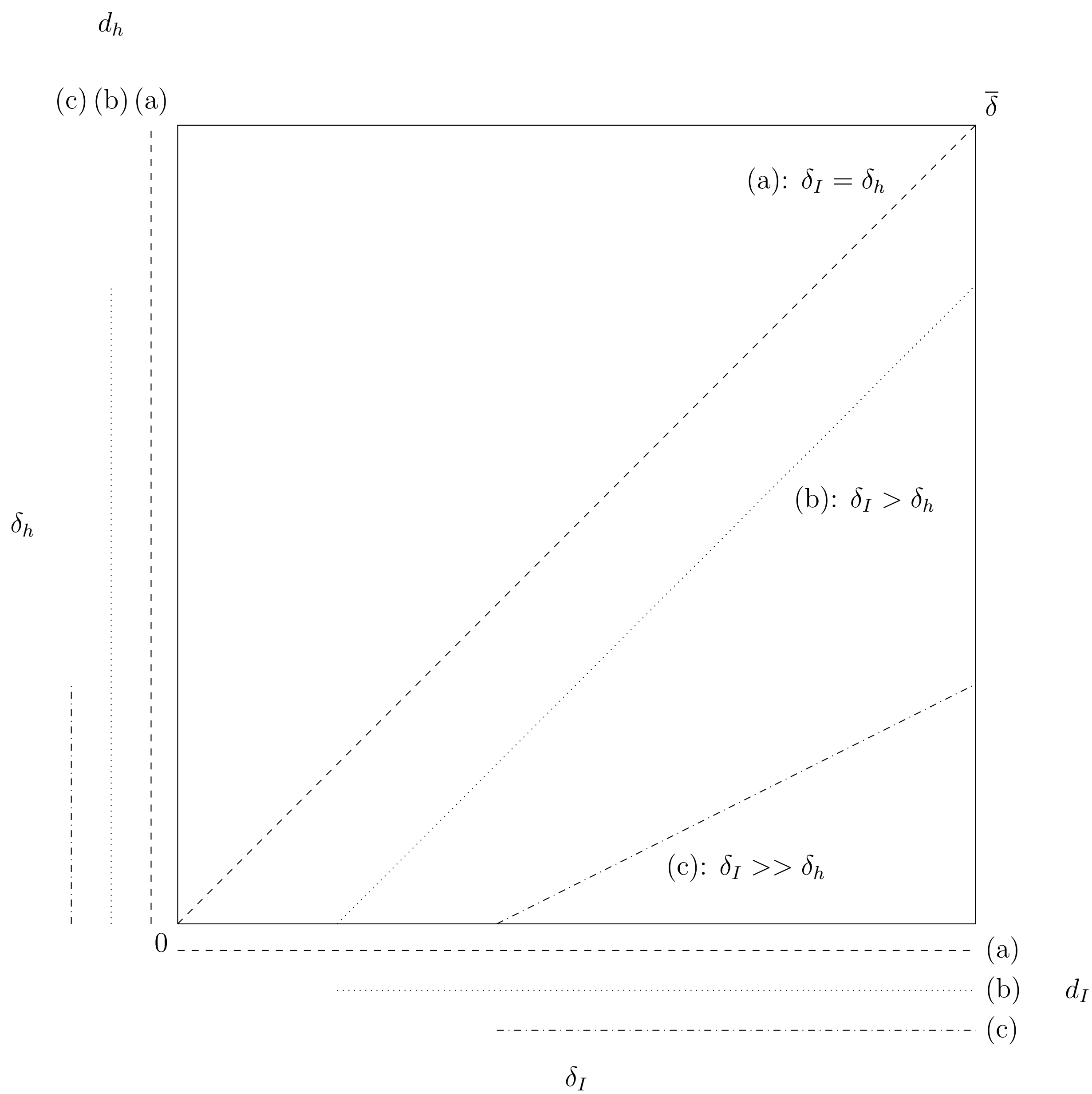
Note: $\Delta y_1 > 0$. "$>>$" means the bound is further from the uninformative bound of, e.g., $\min D_I$, than "$>$" (likewise for "$<<$", "$<$", and "$\max D_I$", respectively).

4 Empirical Results

4.1 Estimation

As indicated by eq. (8), the identified set for (δ_h, δ_I) depends on $(\beta_h, \beta_y, \Delta l_1, \Delta y_1, \Delta y_2)$. This section discusses how values of these variables are obtained, for use in the computation of

Figure 2: Illustration of Example Identified Set Cases



Note: Marginal identified sets for parameters are indicated by the lines outside their respective axes. Cases depicted in the figure are summarized in Table 3.

bounds presented in Section 4.2.

Table 4: Estimates of ATEs

	Dependent variable:	
	Test score y_{ijt}	
	Unpooled	Pooling 1998-99
	(1)	(2)
I(year = 1997)	0.007	0.007
	(0.008)	(0.008)
I(year = 1998)	0.001	
	(0.021)	
I(year = 1999)	0.001	
	(0.022)	
I(year = 1998 |year = 1999)		0.001
		(0.019)
I(year = 2000)	-0.001	-0.001
	(0.017)	(0.017)
I(year = 1998∗treated)	0.041^{**}	
	(0.018)	
I(year = 1999∗treated)	0.149^{***}	
	(0.020)	
I(year = 1998∗treated |year = 1999∗treated)		0.089^{***}
		(0.013)
I(year = 2000∗treated)	0.098^{***}	0.098^{***}
	(0.025)	(0.025)
Observations	26,537	26,537
R^2	0.006	0.004
F Statistic	21.103^{***}	23.185^{***}

Note: *p<0.1; **p<0.05; ***p<0.01. The (year) t subindex on the test score y_{ijt} refers to a study year, not a pooled period.

Table 4 presents the ATE estimates, where "I(·)" indicates that the condition contained therein is true. The estimates exploit the balanced experimental design assumption, i.e., they present raw treatment effects—differences in conditional means for treatment and control groups—for active-treatment and post-active-treatment periods.[15] Column (1) presents the unpooled estimates, i.e., separate estimates for each year of the study, 1997–2000. The top rows present coefficients on time dummies, which are not of particular note. The sixth row, which presents the coefficient on I(year = 1998∗treated), says that the average treatment effect in the first active-treatment year (1998) was positive and significant, while the seventh row, which presents the coefficient on I(year = 1999∗treated), suggests that the average treatment effect was larger in the second active-treatment year. Glewwe et al. (2010) conjecture that the difference in these point estimates may be attributable to teachers learning how the incentive scheme worked during the first active-treatment year, which makes it natural to pool the active-treatment years into one period. The bottom row of column (1) of Table 4 indicates that student achievement in the post-active-treatment year (2000) remained significantly higher in the treatment group. Through the lens of the model, this positive ATE could be due to persistence of higher achievement from the active-treatment period and (potentially) higher teacher human capital for the treatment group in the post-active-treatment period. Establishing the relative importance of these effects is the goal of the next section.

Table 5: Mapping from Study Years to Model Periods

Study Year	1997	1998	1999	2000
Model period	0	1		2

Column (2) of Table 4 presents the pooled ATE results, where both active-treatment years are pooled into one period, corresponding to $t = 1$ in the model. Table 5 summarizes how the pooled analysis combines study years into model periods. Based on the reasoning outlined just above, these results are used to apply the bounds developed in Section 3.1. The results in the penultimate row, for the coefficient on I(year = 1998∗treated |year = 1999∗treated), indicate that the average treatment effect, pooled over both active-treatment years, is positive and significantly different than zero (0.089). The bottom row indicates that student achievement in the post-active-treatment year (2000) remained significantly higher

[15] I include the treatment year as a regressor to control for secular trends.

in the treatment group, and is exactly the same as the unpooled estimate from column (1), as expected. Summarizing the achievement findings from the pooled analysis, we have

$$\widehat{\begin{bmatrix} \Delta y_1 \\ \Delta y_2 \end{bmatrix}} \sim \mathrm{N}\left(\begin{bmatrix} 0.0888 \\ 0.0981 \end{bmatrix}, \begin{bmatrix} 1.75e-04 & 0 \\ 0 & 6.26e-04 \end{bmatrix}\right), \tag{9}$$

where the off-diagonal elements of the variance-covariance matrix for the estimates are estimated to be numerically equal to zero.[16]

Table 6 presents an estimate of the persistence component β_y, estimated using the pooled data. I estimate the persistence component β_y to be 0.592, and statistically greater than zero. This estimate is similar to those in Andrabi et al. (2011); see that paper for a detailed study of estimation of the persistence of student achievement.

Table 6: Estimate of Persistence Component, β_y

	Dependent variable:
	Test score y_{ijt}
I(treated)	0.038^{***}
	(0.014)
$y_{j,t-1}$	0.592^{***}
	(0.009)
Constant	-0.052^{***}
	(0.009)
Observations	7,476
R^2	0.375
F Statistic	$2{,}237.128^{***}$

Note: *p<0.1; **p<0.05; ***p<0.01. These estimates are obtained from a regressions run on the pooled data for pooled periods 1 and 2. Specifically, the (period) t subindex on the test score y_{ijt} refers to a pooled period, the mapping of study years to which is described in Table 5. The reported treatment effect coefficient (on "I(treated)") pools the estimates from 1998/99 and 2000 (i.e., pooled periods one and two). Student j's lagged test score in a given pooled period, $y_{j,t-1}$, is the one from the previous pooled period.

[16]Specifically, they are $-4.37e-20$.

I now discuss how I obtain a value for Δl_1. The literature estimating OJT models of human capital accumulation typically treats total work hours (which are the sum of time spent in production and investment) as observed (see, e.g., Brown, 1976; Heckman, 1976). I follow this literature and use measures of total work hours, which are sufficient for the current paper because the change in total work hours ($\Delta I_t + \Delta h_t$) is the complement of the change in leisure (Δl_t). Glewwe et al. (2010) finds no evidence that teachers in the treatment group on average altered their school attendance or classroom presence (see Table 5, Panels A and B, columns (2) and (3), of Glewwe et al. 2010); while either would be a reasonable measure of total work hours, perhaps the more natural measure would be whether the teacher was observed attending school when the observer checked. Therefore, I argue that it is reasonable to treat $\Delta l_1 = 0$.[17]

Finally, I set $\beta_h = 1$. Although this affects the quantitative results, i.e., estimates of the bounds, it does not affect qualitative results, such as whether there are informative bounds or whether I can reject either the pure OJT or pure LBD specifications.

4.2 Estimates of Bounds

Based on the results from Section 4.1, eq. (8) can be rewritten as

$$\delta_h = \pi_{\text{icept}} + \underbrace{\pi_{\text{slope}}}_{=1} \delta_I, \tag{10}$$

where the intercept of the locus of permissible combinations of (δ_h, δ_I) is $\pi_{\text{icept}} := \frac{\Delta y_2}{\Delta y_1} - \beta_y$ and the slope of the locus, $\pi_{\text{slope}} := 1 + \frac{\beta_h \Delta l_1}{\Delta y_1}$, is equal to one because $\Delta l_1 = 0$, as in cases (a), (b), and (d) in Table 3.

Consider now different cases for $\pi_{\text{icept}} = [\delta_h - \delta_I]$. Recall from Section 3.1 that the bounds on (δ_h, δ_I) are uninformative when $\pi_{\text{icept}} = 0$; i.e., neither the pure OJT nor the pure LBD specifications can be rejected. When $\pi_{\text{icept}} < 0$, there is an informative lower bound on δ_I and an informative upper bound on δ_h; in this case the pure LBD specification (where $\delta_I = 0$), can be rejected. On the other hand, when $\pi_{\text{icept}} > 0$, there is an informative upper

[17] Strictly speaking, all that is necessary is that leisure time be a linear function of either of these variables.

bound on δ_I and an informative lower bound on δ_h; here, the pure OJT specification (where $\delta_h = 0$) would be rejected.

Of course, the true value of π_{icept} is unknown; however, it can be measured by $\widehat{\pi}_{\text{icept}} := \widehat{\frac{\Delta y_2}{\Delta y_1} - \beta_y}$. To simulate the distribution of $\widehat{\pi}_{\text{icept}}$, I use the bivariate normal distribution of $[\ \widehat{\Delta y}_1 \quad \widehat{\Delta y}_2\ ]'$, from eq. (9), and the normal marginal distribution of $\widehat{\beta}_y$, from Table 6.

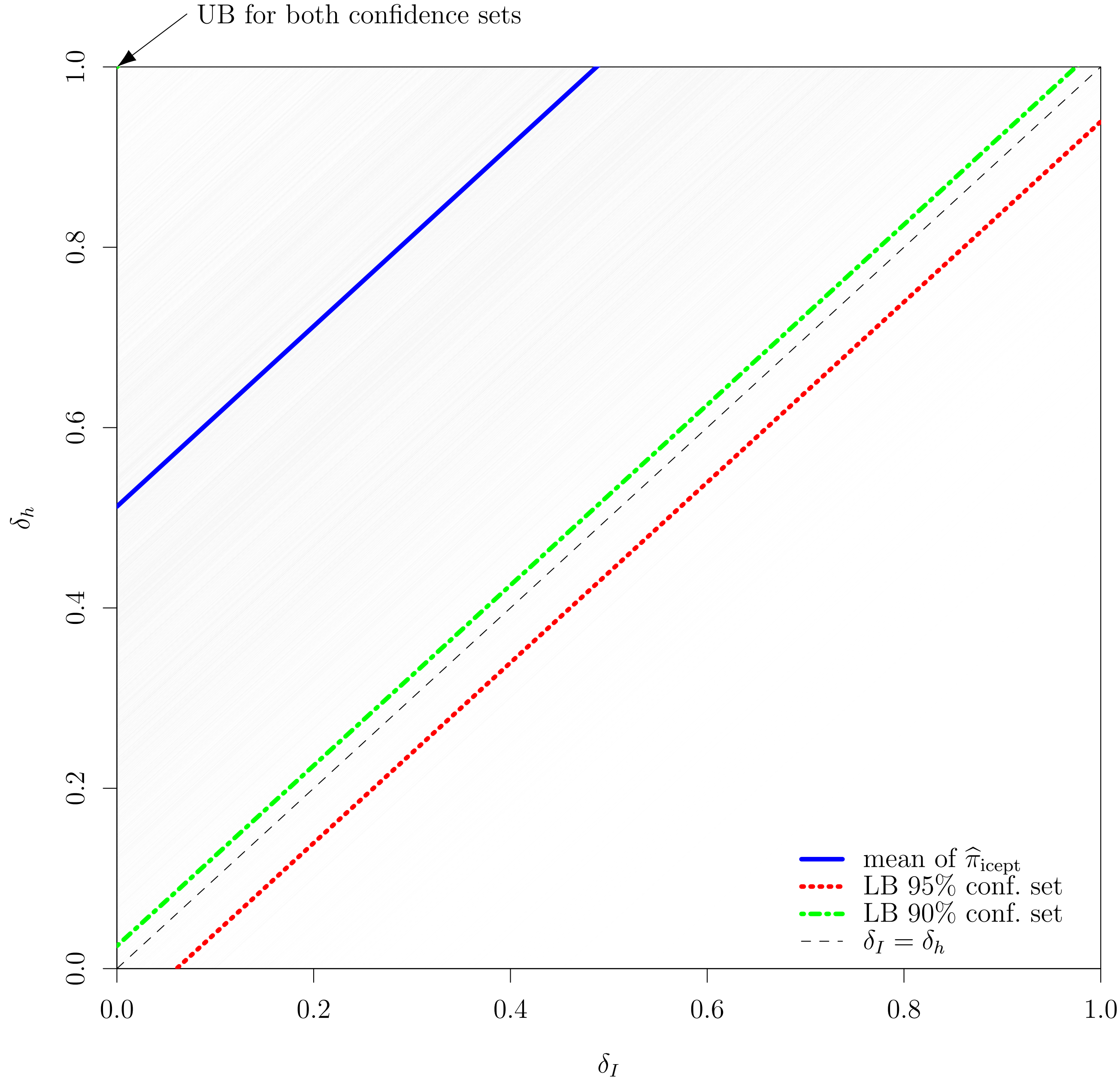
Results The results are illustrated in Figure 3, which sets $\overline{\delta} = 1$. Note that the assumption that $\beta_h = 1$ does not affect qualitative findings discussed here, so long as the upper bounds for δ_h and δ_I are the same. Table 7 presents the 95% and 90% confidence intervals for $[\delta_h - \delta_I]$. We can see that, although the point estimate of $\widehat{\frac{\Delta y_2}{\Delta y_1} - \beta_y}$ is positive, it is only significantly different from zero at the 8% level. Put differently, we cannot reject that the locus of (δ_h, δ_I) passes through the origin of Figure 3 at the conventional 5% significance level. This means we cannot reject either the pure LBD or pure OJT specification at the 5% significance level. At the 10% significance level, there is an informative upper bound on the OJT component (δ_I) and an informative lower bound on the LBD component (δ_h). Further, we can reject the pure OJT specification at the 10% significance level. Intuitively, the ATE in the post-active-treatment period is larger than would be accounted for by the positive ATE in the active-treatment period (caused by an increase in h_1, an input to contemporaneous student achievement) and persistence of this increased student achievement. That is, the OJT component, operating through $\Delta I_1 \leq 0$, is dominated by the LBD component, operating through $\Delta h_1 > 0$.

Table 7: Confidence Interval for $\pi_{\text{icept}} (= [\delta_h - \delta_I])$ and Significance Level for $H_0 : [\delta_h - \delta_I] = 0$

	95% Interval		90% Interval		
$\overline{\widehat{\pi}}_{\text{icept}}$	Lower	Upper	Lower	Upper	Sig. level for $H_0 : [\delta_h - \delta_I] = 0$
0.512	-0.061	1.0	0.025	1.0	0.082

Notes: The significance level is computed against a two-sided alternative.

Figure 3: Estimated Confidence Sets for π_{icept}



Note: Lower bounds (LB) for confidence sets for π_{icept} are those that are visible. Upper bounds (UB) are at the top-left corner of the figure and, as such, are not visible. The mean of $\widehat{\pi}_{\text{icept}}$ is 0.512.

5 Conclusion

I develop a framework nesting the OJT and LBD forces of human capital accumulation, and derive theoretical bounds for OJT and LBD components to teacher human capital accumulation. The developed bounds are sharp, and yield qualitative information about the presence and relative importance of the forces generating human capital. Although the bounds are "wide", it is important to note that the derived bounds only require information about "aggregate moments" (e.g., ATEs and the persistence of student achievement). Additional information, for example, by using microdata on younger teachers, would likely yield tighter bounds. Additional, behavioral, assumptions could tighten the bounds obtained here.[18]

A very promising, complementary, tack would be the structural econometric approach, which could provide a quantitative assessment of the importance of OJT and LBD forces in teachers' human capital accumulation. Such an approach would also be well-suited to rationalize the observed patterns in the data, and would yield other benefits, such as allowing for heterogeneity in teacher human capital accumulation trajectories and, thus, heterogeneity in the growth of teacher quality.[19]

[18]For example, note that the fact that $\Delta l_1 = 0$, if combined with the assumption that the observed data are the result of optimal behavior of teachers, implies that the LBD component must be nonzero because, according to the model, $\Delta h_1 > 0$ to match the increase in achievement in the data, implying that $\Delta I_1 = -\Delta h_1 < 0$, i.e., there were positive levels of I in the control group (because $I \geq 0$). For there to be a positive level of I in the control group it must be the case that $\delta_I > 0$ if there is a positive cost of investment (i.e., if teachers value their leisure). That is, $\Delta l_1 = 0$ coupled with $\Delta y_1 > 0$ tells us that, qualitatively, that $\delta_I > 0$.

[19]To provide another example, the simple bounds developed in this paper take as an input data from an intervention with one period of active treatment, requiring the data to be pooled between both active-treatment years to correspond to the model. A structural approach likely could naturally accommodate variation in effects by active-treatment year.

APPENDIX

A Relationship to a Log-Linear Specification

To fix ideas, this section illustrates how the linear technologies (1)-(2) relate to specifications that are closer to those used in prior literature (see, e.g., Blandin, 2018). The illustration considers one teacher, i, teaching one student, j. Consider the following production function for human capital, denoted here as $\tilde{\theta}_{it}$, which features curvature:

$$\tilde{\theta}_{it} = \tilde{\theta}_{it-1}^{\tilde{\delta}_\theta}[\tilde{\theta}_{it-1}\tilde{I}_{it-1}]^{\tilde{\delta}_I}[\tilde{\theta}_{it-1}\tilde{h}_{it-1}]^{\tilde{\delta}_h}, \tag{11}$$

where $\tilde{\theta}_{it-1}$ is teacher i's human capital last period (which may depreciate), $\tilde{I}_{it-1}$ is the share of teacher i's human capital spent on on-the-job investment in human capital last period, and $\tilde{h}_{it-1}$ is the share of teacher i's human capital spent on production last period. Consider the following output production function for teacher i's value added, $\tilde{y}_{ijt} - \beta_y \tilde{y}_{jt-1}$, which also features curvature:

$$\tilde{y}_{ijt} - \tilde{\beta}_y \tilde{y}_{jt-1} = [\tilde{\theta}_{it}\tilde{h}_{it}]\tilde{\epsilon}_{ijt}, \tag{12}$$

where $\tilde{y}_{jt-1}$ is student j's prior achievement, $\tilde{h}_{it}$ represents the share of teacher i's human capital allocated to production in period t, and $\tilde{\epsilon}_{ijt}$ is an IID productivity shock, realized after the choice of $\tilde{h}_{it}$ has been made in period t.

Denoting logs of variables with double dots (e.g., $log(\tilde{\theta}_{it}) = \ddot{\tilde{\theta}}_{it}$), we then have the log-linearized version of the human capital production function, eq. (11):

$$\tilde{\ddot{\theta}}_{it} = [\tilde{\delta}_\theta + \tilde{\delta}_I + \tilde{\delta}_h]\tilde{\ddot{\theta}}_{it-1} + \tilde{\delta}_I \tilde{\ddot{I}}_{it-1} + \tilde{\delta}_h \tilde{\ddot{h}}_{it-1}, \tag{13}$$

which is similar to eq. (1), except that the variables are all in logs, and that $\tilde{\ddot{h}}_{it}$ represents the share of human capital allocated to production in period t here, while h_{it} represents the total hours allocated to period t production in the specification developed in Section 3. However, given that none of $\tilde{\ddot{\theta}}, \tilde{\ddot{h}}, \tilde{\ddot{I}}$ (or θ, h, I, as the case may be) are directly measured and that qualitative changes in inputs are what drive the analysis, this does not substantially affect interpretation within this illustration. Similarly, we have the log-linearized version of the achievement-value-added production function, eq. (12):

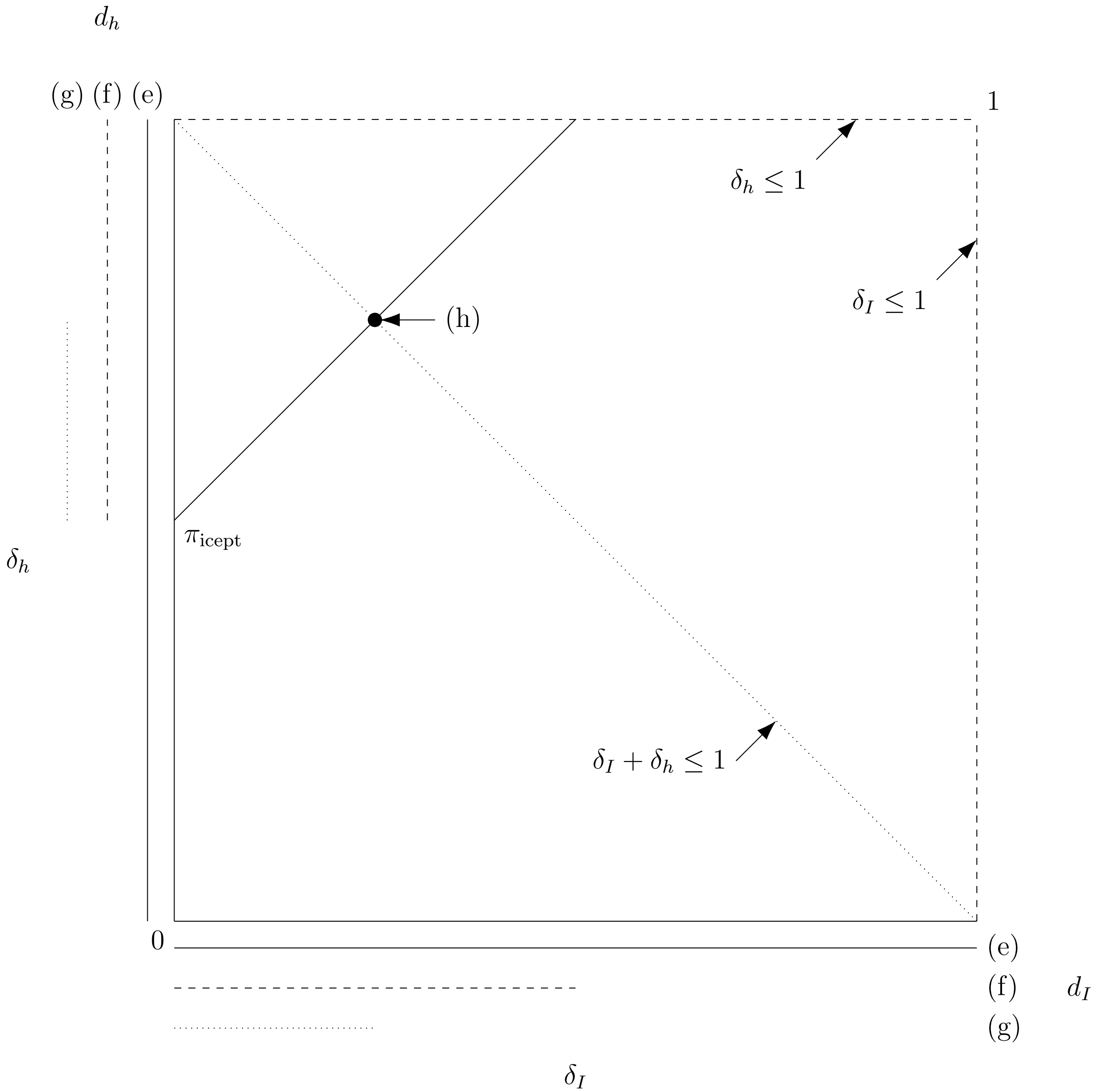
$$log(\tilde{y}_{ijt} - \tilde{\beta}_y \tilde{y}_{jt-1}) = \tilde{\ddot{\theta}}_{it} + \tilde{\ddot{h}}_{it} + \ddot{\tilde{\epsilon}}_{ijt}, \tag{14}$$

which is similar to eq. (2), except that the latter is in terms of test score and input levels, and that the latter contains two additional degrees of freedom, β_θ and β_h (although, as discussed earlier, β_θ is normalized in any case). Given the earlier argument that qualitative differences in inputs drive the analysis, and the known sensitivity of achievement tests to monotonic transformations (Bond and Lang, 2013), this also does not substantially affect interpretation within this illustration.

B Adding a Returns to Scale Assumption

Table 8 and Figure 4 show the additional information embedded in assumptions about the returns to scale for (δ_h, δ_I). Case (e) starts without infor-

Figure 4: Illustration of the Power of an Additional Assumption About the Returns to Scale



Note: Marginal identified sets for parameters are indicated by the lines outside their respective axes. Figure is drawn with $\pi_{\text{icept}}(= \Delta y_2/\Delta y_1 - \beta_y) > 0$, $\Delta l_1 = 0$, and $\overline{\delta} = 1$. Case (e) is the uninformative case, in which marginal identified sets are identical with and without the additional assumption about the returns to scale. Case (f) corresponds to case (d) in Figure 2, in which $(\delta_h, \delta_I) \in [0, 1] \times [0, 1]$. Case (g) further imposes the restriction that $\delta_h + \delta_I \leq 1$, where the uninformative identified set for (δ_h, δ_I) is the lower right triangle. Case (h) further imposes the restriction that $\delta_h + \delta_I = 1$, which leads to point identification, as pictured. Additional detail about the cases is in Table 8.

Table 8: Example of the Power of an Additional Assumption, about the Returns to Scale

			Marginal Identified Sets				
Case	$(\delta_h, \delta_I) \in$	$\frac{\Delta y_2}{\Delta y_1} - \beta_y$	$\min d_I$	$\max d_I$	$\min d_h$	$\max d_h$	Notes about Bounds
(e)	$\delta_h \leq 1,\ \delta_I \leq 1$	—	$\min D_I$	$\max D_I$	$\min D_h$	$\max D_h$	Uninformative
(f)	$\delta_h \leq 1,\ \delta_I \leq 1$	> 0	$\min D_I$	$< \max D_I$	$> \min D_h$	$\max D_h$	Informative
(g)	$\delta_h + \delta_I \leq 1$	> 0	$\min D_I$	$<< \max D_I$	$> \min D_h$	$< \max D_h$	Tighter than (f)
(h)	$\delta_h + \delta_I = 1$	> 0	Point identification achieved				

Note: $\Delta y_1 > 0$. ">>" means the bound is further from the uninformative bound of, e.g., $\min D_I$, than ">" (likewise for "<<", "<", and "$\max D_I$", respectively). $\Delta l_1 = 0$ and $\beta_h = 1$ in all cases.

mation about $\frac{\Delta y_2}{\Delta y_1} - \beta_y$, which corresponds to the uninformative case (solid line in Figure 4). Case (f) corresponds to case (d) in Table 3, in which $(\delta_h, \delta_I) \in [0,1] \times [0,1]$ but there is a measure of $\frac{\Delta y_2}{\Delta y_1} - \beta_y$ so there are informative bounds on the marginal identified sets (dashed lines in Figure 4). Case (g) further imposes the restriction that $\delta_h + \delta_I \leq 1$, where the uninformative identified set for (δ_h, δ_I) is the southwest right triangle. By looking at the dotted lines in Figure 4, corresponding to this case, we can see that the marginal identified sets are tighter than those in case (f). Intuitively, the tighter upper bound for δ_h, combined with the nonincreasing returns to scale assumed in case (g), yields a tighter upper bound for δ_I. Case (h) further imposes the restriction that $\delta_h + \delta_I = 1$, which affords point identification (interior solid point in Figure 4).

References

Andrabi, T., J. Das, A. I. Khwaja and T. Zajonc, "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics," *American Economic Journal: Applied Economics*, 3(3):29–54, 2011.

Becker, G. S., *Human Capital Theory*, Columbia University Press, New York, 1964.

Ben-Porath, Y., "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4):352–365, 1967.

Blandin, A., "Learning by Doing and Ben-Porath: Life-cycle Predictions and Policy Implications," *Journal of Economic Dynamics and Control*, 90:220–235, 2018.

Bond, T. N. and K. Lang, "The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results," *Review of Economics and Statistics*, 95(5):1468–1479, 2013.

Brown, C., "A Model of Optimal Human-Capital Accumulation and the Wages of Young High School Graduates," *Journal of Political Economy*, 84(2):299–316, 1976.

Chetty, R., J. N. Friedman and J. E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104(9):2593–2632, 2014.

Glewwe, P., N. Ilias and M. Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics*, 2(3):205–227, 2010.

Haley, W. J., “Estimation of the Earnings Profile from Optimal Human Capital Accumulation,” *Econometrica*, pp. 1223–1238, 1976.

Hanushek, E. A., “Conceptual and Empirical Issues in the Estimation of Educational Production Functions,” *Journal of Human Resources*, pp. 351–388, 1979.

—, “The Economic Value of Higher Teacher Quality,” *Economics of Education Review*, 30(3):466–479, 2011.

Hanushek, E. A. and S. G. Rivkin, “Teacher Quality,” *Handbook of the Economics of Education*, 2:1051–1078, 2006.

Harris, D. N. and T. R. Sass, “Teacher Training, Teacher Quality and Student Achievement,” *Journal of Public Economics*, 95(7):798–812, 2011.

Heckman, J., L. Lochner and R. Cossa, “Learning-by-Doing Vs. On-the-Job Training: Using Variation Induced by the EITC to Distinguish Between Models of Skill Formation,” *NBER Working Paper Series*, p. 9083, 2002.

Heckman, J. J., “A Life-Cycle Model of Earnings, Learning, and Consumption,” *Journal of Political Economy*, 84(4):S11–44, 1976.

Hölmstrom, B. and P. Milgrom, “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design,” *Journal of Law, Economics, and Organization*, 7(Special issue):24–52, 1991.

Imbens, G. W. and C. F. Manski, “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 72(6):1845–1857, 2004.

Killingsworth, M. R., "'Learning by Doing' and 'Investment in Training': A Synthesis of Two 'Rival' Models of the Life Cycle," *Review of Economic Studies*, 49(2):263–271, 1982.

Kinsler, J., "Assessing Rothstein's Critique of Teacher Value-Added Models," *Quantitative Economics*, 3(2):333–362, 2012.

McCaffrey, D. F., J. Lockwood, D. M. Koretz and L. S. Hamilton, *Evaluating Value-Added Models for Teacher Accountability. Monograph.*, ERIC, 2003.

McWilliams, J., "The Lucrative Art of Chicken Sexing," *Pacific Standard*, 2018.

Muralidharan, K. and V. Sundararaman, "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 119(1):39–77, 2011.

Papay, J. P. and M. A. Kraft, "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-term Career Improvement," *Journal of Public Economics*, 130:105–119, 2015.

Podgursky, M. and M. Springer, "Teacher Compensation Systems in the United States K-12 Public School System," *National Tax Journal*, 64(1):165–192, 2011.

Rivkin, S. G., E. A. Hanushek and J. F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, 73(2):417–458, 2005, ISSN 1468-0262.

Shaw, K. L., "Life-cycle Labor Supply with Human Capital Accumulation," *International Economic Review*, pp. 431–456, 1989.

Stoye, J., "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77(4):1299–1315, 2009.

Tamer, E., "Partial Identification in Econometrics," *Annual Review of Economics*, 2(1):167–195, 2010.

Todd, P. E. and K. I. Wolpin, "On the Specification and Estimation of the Production Function for Cognitive Achievement," *Economic Journal*, 113(485):F3–F33, 2003.

Weiss, Y., "On the Optimal Lifetime Pattern of Labour Supply," *Economic Journal*, 82(328):1293–1315, 1972.

Wiswall, M., "The Dynamics of Teacher Quality," *Journal of Public Economics*, 100:61–78, 2013.