Additional Material for "Estimating the Technology of Cognitive and Noncognitive Skill Formation" (Cuttings from the Web Appendix)

Flavio Cunha
The University of Pennsylvania

James Heckman
The University of Chicago,
Geary Institute, University College Dublin,
Yale University, and
American Bar Foundation

Susanne Schennach
The University of Chicago

First draft, March 2006
This draft, January 10, 2010

Contents

A More Detailed Discussion of the Identification of the Technology

From the conditional joint distribution $F\left(Z_{2,k,t,1},...,Z_{2,k,t,M_{2,k,t}}\middle|\theta_t,\theta_P,y_t\right)$ we need to identify the distribution of π_t, F_{π_t}, as well as the policy function $g\left(\theta_t,\theta_P,\pi_t,y_t\right)$. We start with the case in which π_t is i.i.d. and independent of (θ_t,θ_P,y_t) and then relax it. Note that we can use the arguments in section 3.1 to identify the joint distribution of $F\left(\theta_t,\theta_P,I_{k,t},y_t\right)$. As shown in Matzkin (2003), if (1) $g_{k,t}$ is monotone in π_t and (2) we know a point $\left(\bar{\theta}_t,\bar{\theta}_P,\bar{y}_t\right)$ such that $g_t\left(\bar{\theta}_t,\bar{\theta}_P,\pi_t,\bar{y}_t\right)=\pi_t$, then we can identify the distribution F_π and the function $g_{k,t}$ nonparametrically. [1] To see why, note that we can compute at each point (θ_t,θ_P,y_t) the following probability:

$$\Pr\left(I_{k,t}\le q|\,\theta_t,\theta_P,y_t\right)=\Pr\left(g_{k,t}\left(\theta_t,\theta_P,\pi_t,y_t\right)\le q|\,\theta_t,\theta_P,y_t\right),$$

Consequently, computing that probability at $\left(\bar{\theta}_t,\bar{\theta}_P,\bar{y}_t\right)$ yields the probability distribution function for π_t:

$$\Pr\left(I_{k,t}\le q|\,\bar{\theta}_t,\bar{\theta}_P,\bar{y}_t\right)=\Pr\left(\pi_t\le q|\,\bar{\theta}_t,\bar{\theta}_P,\bar{y}_t\right)=\Pr\left(\pi_t\le q\right)=F_{\pi_t}\left(q\right)$$

for all $q\in\mathbb{R}$. For any other point (θ_t,θ_P,y_t), let $\tau\left(\theta_t,\theta_P,y_t\right)$ be such that $\tau\left(\theta_t,\theta_P,y_t\right)=\Pr\left(I_{k,t}\le q|\,\theta_t,\theta_P,y_t\right)$, which can be estimated from the data. But then,

$$\Pr\left(I_{k,t}\le q|\,\theta_t,\theta_P,y_t\right)=\Pr\left(\pi_t\le g_{k,t}^{-1}\left(\theta_t,\theta_P,y_t,q\right)\middle|\,\theta_t,\theta_P,y_t\right)=F_\pi\left(g_{k,t}^{-1}\left(\theta_t,\theta_P,y_t,q\right)\right).$$

But since F_π is known, we also know $g_{k,t}^{-1}$ because $g_{k,t}^{-1}\left(\theta_t,\theta_P,y_t,q\right)=F_\pi^{-1}\left(\tau\left(\theta_t,\theta_P,y_t\right)\right)$. Knowledge of $g_{k,t}^{-1}$ yields knowledge of $g_{k,t}$.Next, we proceed to a second stage in which we replace the identified policy function $g_{k,t}$ in the production function:

$$\theta_{k,t+1}=f_{s,k}\left(\theta_t,g_{k,t}\left(\theta_t,\theta_P,\pi_t,y_t\right),\theta_P,\pi_t,\nu_{k,t}\right).$$

Given that π_t is independent from (θ_t,θ_P,y_t), we can easily construct the joint distribution of the state variables because $F\left(\theta_t,\theta_P,\pi_t,y_t\right)=F\left(\theta_t,\theta_P,y_t\right)F\left(\pi_t\right)$. If (1) $f_{s,k}$ is monotone in $\nu_{k,t}$ and (2) there exists a known point $\left(\tilde{\theta}_t,\tilde{\theta}_P,\tilde{\pi}_t,\tilde{y}_t\right)$, such that

$$f_{s,k}\left(\tilde{\theta}_t,g_t\left(\tilde{\theta}_t,\tilde{\theta}_P,\tilde{\pi}_t,\tilde{y}_t\right),\tilde{\theta}_P,\tilde{\pi}_t,\nu_{k,t}\right)=\nu_{k,t},$$

[1] Although these may seem strong assumptions, note that parametric models in general impose similar normalizations implicitly. For example, if $g_t\left(\theta_t,\theta_P,I_t,\pi_t\right)$ is linear and separable in its arguments, then it is clearly monotonic in π_t and $g\left(0,0,0,\pi_t\right)=\pi_t$.

then by applying the analysis in Matzkin (2003), we can recover the technology function $f_{s,k}$ as well as the distribution of the error term $\nu_{k,t}$.

Allowing for Serial Correlation in Shocks Suppose that π_t follows an AR(1) process:

$$\pi_{t+1} = m(\pi_t) + \nu_{\pi,t}. \tag{A.1}$$

where the innovation $\nu_{\pi,t}$ is independent. Note that iid is not needed. Sufficient conditions are: (i) the process $\{\pi_t\}_{t=1}^{T}$ is independent of the vector $\left\{\theta_1, \theta_P, \{y_t\}_{t=1}^{T}\right\}$, and (ii) the policy function for investments can be written as:

$$I_{k,t} = g_{k,t}(\theta_t, \theta_P, y_t) + \pi_t.$$

Open Question (Cunha): Can we weaken additive separability?

Under these conditions, it is possible to identify the function $g_{k,1}(\theta_t, \theta_P, y_t)$ and the distribution F_{π_1} by following the argument in the last section. The independence assumption allows us to recover the joint distribution of $(\pi_1, \theta_1, \theta_P, y_1)$. Consequently, we can identify the technology $f_{1,k}(\theta_1, g_{k,1}(\theta_1, \theta_P, y_1) + \pi_1, \theta_P, \pi_1, \nu_{k,1})$ for the first stage of development as well as the distribution of $\nu_{k,1}$, F_{ν_k} for $k = C, N$. Next, we need to identify the evolution of π_1. Without loss of generality, assume that the factor loading $\alpha_{2,k,t,1} = 1$ for all k and t. Then, for $t = 2$ and $j = 1, ..., M_{2,k,t}$

$$Z_{2,k,2,j} = \alpha_{2,k,2,j} g_{k,2}(\theta_2, \theta_P, y_2) + \alpha_{2,k,2,j}\pi_2 + \varepsilon_{2,k,2,j}.$$

Identification is complicated by the fact that under serial correlation in π_2, θ_2 is correlated with π_2. To identify $g_{k,2}$ we can use θ_1 as an instrument for θ_2 while noting that θ_P and y_2 can serve as instruments for themselves. Clearly, we can identify the joint distribution of (π_1, π_2) from the residualized measurement equations:

$$Z_{2,k,t,j} - \alpha_{2,k,t,j} g_{k,t}(\theta_t, \theta_P, y_t) = \pi_t + \ \varepsilon_{2,k,t,j} \text{ for } t = 1, 2.$$

We recover the function m from the fact that $m(\pi_t) = \int \pi_{t+1} dF_{\pi_{t+1}|\pi_t}(\pi_{t+1}|\, \pi_t)$.This, in turn, allows us to recover the distribution of $\nu_{\pi,1}$, $F_{\nu_{m,1}}$,because for any q, $\Pr(\nu_{\pi,1} \leq q) = \Pr(\pi_2 - m(\pi_1) \leq q)$. Under the assumption that $\nu_{\pi,t}$ is i.i.d., this yields the distribution function for $\nu_{\pi,t}$ for all t.

Finally, note that we can identify the joint distribution of $(\pi_2, \theta_2, \theta_P, y_2)$. To see why, consider the probability of the event:

$$\begin{aligned}
&\Pr\left(\pi_2 \leq q_\pi, \theta_2 \leq q_\theta \middle| \theta_1, \theta_P, y_2, y_1, \pi_1\right) \\
&\quad = \Pr\left(m\left(\pi_1\right) + \nu_{\pi,1} \leq q_\pi, f_1\left(\theta_1, g_1, \theta_P, \pi_1, \nu_1\right) \leq q_\theta \middle| \theta_1, \theta_P, y_2, y_1, \pi_1\right) \qquad \text{(A.2)} \\
&\quad = \Pr\left(\nu_{\pi,1} \leq q_\pi - m\left(\pi_1\right), \leq f_1^{-1}\left(\theta_1, g_1, \theta_P, \pi_1, q_\theta\right) \middle| \theta_1, \theta_P, y_2, y_1, \pi_1\right) \\
&\quad = F_{\nu_\pi}\left(q_\pi - m\left(\pi_1\right)\right) F_\nu\left(f_1^{-1}\left(\theta_1, g_1, \theta_P, \pi_1, q_\theta\right)\right).
\end{aligned}$$

Note that we secured identification of the functions $f_{1,k}$, $g_{k,1}$, and the joint distribution of $(\pi_2, \theta_2, \theta_P, y_2)$. To proceed, assume that at every period t there is a new technology function to be identified, so that each period t is a new development stage s. Assume that both the technology and the policy functions $f_{s,k}$ and $g_{k,t}$ as well as the joint $(\pi_{t+1}, \theta_{t+1}, \theta_P, y_{t+1})$ have already been identified. We will show that it is also possible to identify the technology and policy functions $f_{s+1,k}$ and $g_{k,t+1}$ as well as the joint distribution of $(\pi_{t+2}, \theta_{t+2}, \theta_P, y_{t+2})$.

$$Z_{2,k,t+1,j} = \alpha_{2,k,t+1,j} g_{k,t+1}\left(\theta_{t+1}, \theta_P, y_{t+1}\right) + \alpha_{2,k,t+1,j}\pi_{t+1} + \varepsilon_{2,k,t+1,j}.$$

Because we know the joint distribution of $(\pi_{t+1}, \theta_{t+1}, \theta_P, y_{t+1})$, we can clearly identify the function $g_{k,t+1}(\theta_{t+1}, \theta_P, y_{t+1})$. We then substitute the policy function into the production function and obtain identification of $f_{s+1,k}$. If $t > T-2$, there is no need to proceed. If $t \leq T-2$, we use the assumption that the function m is stage (and time) invariant and repeat the steps in (A.2) to recover the joint distribution of $(\pi_{t+2}, \theta_{t+2}, \theta_P, y_{t+2})$.

B Equalizing Outcomes

The criterion adopted for the first problem assumes that the goal of society is to get the schooling of every child to a twelfth grade level. The required investments measure the power of initial endowments in determining inequality and the compensation through investment that is required to eliminate their influence. Let $e(\theta_{1,h})$ be the minimum cost of attaining 12 years of schooling for a child with endowment $\theta_{1,h}$. Assuming no discounting, the problem is formally defined by

$$e\left(\theta_{1,h}\right) = \min\left[I_{1,h} + I_{2,h}\right]$$

subject to a schooling constraint:

$$S\left(\theta_{C,3,h}, \theta_{N,3,h}, \pi_h\right) = 12,$$

where S maps end of childhood capabilities and other relevant factors (π_h) into schooling attainment, subject to the technology of capability formation constraint

$$\theta_{k,t+1,h} = f_{k,t}\left(\theta_{C,t,h}, \theta_{N,t,h}, \theta_{C,P,h}, \theta_{N,P,h}, I_{t,h}, \pi_h\right) \text{ for } k \in \{C, N\} \text{ and } t \in \{1, 2\},$$

and the initial endowments of the child and her parents. We have estimated all of the ingredient functions.[2]

Figures 2 (for child endowments) and 3 (for parental endowments) plot the percentage increase in investment over that required for a child with mean parental and personal endowments to attain high school graduation.[3] The shading in the graphs represents different values of investments. The lightly shaded areas of the graph correspond to higher values. Eighty percent more investment is required for children with the most disadvantaged personal endowments (Figure 2). The corresponding figure for children with the most disadvantaged parental endowments is 95% (Figure 3). The negative percentages for children with high initial endowments is a measure of their advantage. From the analysis of Moon (2009), investments *received* as a function of a child's endowments are typically in reverse order from what are required. Children born with advantageous endowments typically receive more parental investment than children from less advantaged environments.

[2]See Web Appendix 8 for the estimates of the schooling equation.

[3]In graphing the investments as a function of the displayed endowments, we set the values of other endowments at mean values.

C Econometric Issues Addressed in this Paper

To build intuition for our econometric procedure, in this subsection, we show how the technology can be identified if (i) both skills and investments are measured without any error and (ii) there exists a natural metric in which to measure skills. For simplicity, consider the situation in which there is only one skill, "human capital", θ_t. We decompose the error term η_t into two independent components, π_t and ν_t. The component π_t is known by parents at the time they make investment decisions. The component ν_t is realized at the end of period t, after investment decisions are made, and is not known by the parents. We assume that $E\left(\nu_t | \theta_t, I_t, \theta_P, \pi_t\right) = 0$. To focus on essential points, we rewrite technology (2.1) in the main text as:

$$\theta_{t+1} = f\left(\theta_t, I_t, \theta_P, \pi_t\right) + \nu_t. \tag{C.1}$$

We restate the problem of the parent in a more complete fashion. Because there is only one skill, it is without loss of generality that we assume $Q = \theta_{T+1}$. Let $R\left(\theta_{T+1}\right)$ denote the net present value of the child's lifetime income. At each period of childhood t, parental income is y_t. It is observed at the beginning of period t by the parents, but is stochastic from the point of view of the parent at periods $\tau < t$. For simplicity, assume that parents cannot borrow or save, so the budget constraint in period t is

$$c_t + I_t = y_t. \tag{C.2}$$

where c_t is parental consumption. The state variables are denoted by $\Omega_t = \left(\theta_t, \theta_P, y_t, \pi_t\right)$. Only $\left(\theta_t, \theta_P, y_t\right)$ are observed by the econometrician. Given the state variables, the parent chooses household consumption c_t and investments I_t in skill of the child. In the final period of childhood $t = T$, the problem solved by the parents is to maximize value:

$$V_T\left(\Omega_T\right) = \max \left\{ u\left(c_T\right) + \left(\frac{1}{1+\delta}\right) E\left[R\left(\theta_{T+1}\right) | \, \Omega_T \right] \right\} \tag{C.3}$$

subject to (C.1) and (C.2). The expectation in (C.3) is with respect to shock ν_T that is realized after investment I_T is made. For periods $t = 1, ..., T-1$, the problem of the parent is:

$$V_t\left(\Omega_t\right) = \max_{c_t, I_t} \left\{ u\left(c_t\right) + \beta E\left[V_{t+1}\left(\theta_{t+1}, \theta_P, y_{t+1}, \pi_{t+1}\right) | \, \Omega_t \right] \right\} \tag{C.4}$$

subject to (C.1) and (C.2). The expectation in (C.4) is computed with respect to income y_{t+1} and the shocks ν_t. In each period t, the solution of the problem is a policy function for investments,

$$I_t = g_t\left(\theta_t, \theta_P, \pi_t, y_t\right). \tag{C.5}$$

Note that even if we assume that π_t is i.i.d. and independent from the other state variables, the policy function (C.5) shows that investment will be correlated with the unobserved π_t. To overcome this type of endogeneity problem in estimating the production function, the recent literature follows Olley and Pakes (1996). Their approach proceeds in two stages. They assume that (C.5) is invertible in π_t for all θ_t, θ_p, and y_t. In the first stage of their procedure, they estimate an object that is the combination of the nonparametrically specified inverse policy function $g_t^{-1}(\theta_t, \theta_P, I_t, y_t)$ that links π_t to observed quantities and a production function that is linear and separable in its arguments.

In the second stage, the parameters associated with the production function are identified under assumptions about the dynamics of π_t. They assume that $\pi_t = \pi_{t-1} + \zeta_t$, where ζ_t is an innovation that is independent of π_{t-1} and everything else. The first-stage approach in Olley and Pakes (1996) substitutes in (C.1) for π_t using the inverse policy function (C.5) and estimates the nonparametric function κ_t which is defined as:

$$\theta_{t+1} = f\left(\theta_t, I_t, \theta_P, g_t^{-1}(\theta_t, \theta_P, I_t, y_t)\right) + \nu_t = \kappa_t(\theta_t, \theta_P, I_t, y_t) + \nu_t.$$

The function κ_t is clearly identified because all of its arguments are observed and exogenous with respect to ν_t. If f is monotonic in π_t, from the definition of κ_{t-1} we obtain:

$$\pi_{t-1} = f^{-1}(\theta_{t-1}, I_{t-1}, \theta_P, \kappa_{t-1}).$$

Using the law of motion for π_t, we conclude that:

$$\pi_t = f^{-1}(\theta_{t-1}, I_{t-1}, \theta_P, \kappa_{t-1}) + \zeta_t. \tag{C.6}$$

If we replace (C.6) in (C.1), we obtain the second stage equation:

$$\theta_{t+1} = f\left(\theta_t, I_t, \theta_P, f^{-1}(\theta_{t-1}, I_{t-1}, \theta_P, \kappa_{t-1}) + \zeta_t\right).$$

There are two problems in obtaining a consistent estimator of f from the second-stage equation. First, if f is not additive and separable in π_t, then it is necessary to make some form of normalization to identify the production function. This arises because the functions $f(\theta_t, I_t, \theta_P, \pi_t)$ and $\tilde{f}(\theta_t, I_t, \theta_P, a + b\pi_t)$ are observationally equivalent.[4]

Second, if I_t is correlated with π_t, then I_t is in general correlated with ζ_t. The exception

[4] To see why, suppose that $\kappa_{t-1} = f(\theta_{t-1}, I_{t-1}, \theta_P, a + b\pi_{t-1})$. Then, we can invert f with respect to π_{t-1} and obtain $\pi_{t-1} = \frac{f(\theta_{t-1}, I_{t-1}, \theta_P, \kappa_{t-1}) - a}{b}$. But then $f(\theta_t, I_t, \theta_P, a + b\pi_t) = f(\theta_t, I_t, \theta_P, f(\theta_{t-1}, I_{t-1}, \theta_P, \kappa_{t-1}) + b\zeta_t)$. Since we don't know b and we don't observe ζ, we can't tell $b\zeta_t$ and ζ_t apart.

is the case in which investments I_t are chosen before the realization of ζ_t, in which case the entire correlation between I_t and π_t is due to π_{t-1}. If this restriction on the timing of the arrival of information does not hold, then it is necessary to find instruments for investments. The simple model previously sketched suggests using income, or unanticipated innovations in income, as instruments for investments. Note that in Olley and Pakes (1996) this problem does not arise because it is assumed that the dynamic input (i.e., capital) is chosen before ζ_t is known.[5]

To handle the endogeneity of investments, we can use a nonparametric instrumental quantile regression to identify the dependence of

$$f\left(\theta_t, I_t, \theta_P, f^{-1}\left(\theta_{t-1}, I_{t-1}, \theta_P, \kappa_{t-1}\right) + q_\tau\right)$$

on $\theta_t, I_t, \theta_P, \theta_{t-1}, I_{t-1}, \kappa_{t-1}$, and τ, where q_τ is the t^{th}-quantile of ζ_t. Unfortunately, this argument does not recover the function q_τ (Chernozhukov, Imbens, and Newey, 2007). To recover this, consider the subset of values where $\theta_t = \theta_{t-1} = \theta$ and $I_t = I_{t-1} = I$. This special set of values generates an equation:

$$f\left(\theta, I, \theta_P, f^{-1}\left(\theta, I, \theta_P, \kappa\right) + q_\tau\right) = \kappa$$

which we can use to identify the τ such that $q_\tau = 0$. Next, for a fixed value of θ_t, I_t, and θ_P, consider a change in both κ_t and τ that leaves the function $f\left(\theta_t, I_t, \theta_P, f^{-1}\left(\theta_{t-1}, I_{t-1}, \theta_P, \kappa_{t-1}\right) + q_\tau\right)$ unchanged at $q_\tau = 0:$

$$0 \equiv df\left(\theta_t, I_t, \theta_P, f^{-1}\left(\theta_{t-1}, I_{t-1}, \theta_P, \kappa_{t-1}\right) + q_\tau\right)\Big|_{q_\tau=0} = d\kappa_t + \frac{\partial f\left(\theta_t, I_t, \theta_P, \pi_t\right)}{\partial \pi_t} \left.\frac{\partial q_\tau}{\partial \tau}\right|_{q_\tau=0} d\tau,$$

and we obtain:

$$\frac{d\kappa_t}{d\tau} = -\frac{\partial f\left(\theta_t, I_t, \theta_P, \pi_t\right)}{\partial \pi_t} \left.\frac{\partial q_\tau}{\partial \tau}\right|_{q_\tau=0}$$

which is the derivative of f with respect to π_t up to a constant term $b = \left.\frac{\partial q_\tau}{\partial \tau}\right|_{q_\tau=0}$. As a consequence, we know how f depends on π_t up to an additive constant, a. The argument above can be repeated for all values of θ_t, I_t, and θ_P to recover how f depends on all of its arguments. To pin down the constants a and b, we can proceed by normalizing $Var\left(\zeta_t\right) = 1$ and the location of $f\left(\theta_t, I_t, \theta_P, .\right)$.

Although the argument above extends Olley and Pakes (1996) to a nonparametric production function setting, it still relies on observing inputs and investments without measure-

[5] In their setup, the nondynamic input (i.e., labor) is correlated with ζ_t, but its coefficient is consistently estimated in the first-stage equation.

ment error. Such assumption does not hold when one has as inputs skills and observations of parental investments in skills.

References

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007, July). Instrumental variable estimation of nonseparable models. *Journal of Econometrics 139*(1), 4–14.

Matzkin, R. L. (2003, September). Nonparametric estimation of nonadditive random functions. *Econometrica 71*(5), 1339–1375.

Moon, S. H. (2009). Multi-dimensional human skill formation with multi-dimensional parental investment. Unpublished manuscript, University of Chicago, Department of Economics.

Olley, G. S. and A. Pakes (1996, November). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.