Semiparametric Reduced-Form Estimation of Tuition Subsidies

By HIDEHIKO ICHIMURA AND CHRISTOPHER TABER*

The goal of this paper is to use a semiparametric reduced-form model to estimate the effects of various tuition subsidies. This approach expands on the tuition-subsidy example in Ichimura and Taber (2000) in a number of dimensions.

It has become common practice in the empirical literature to refer to any nonstructural empirical analysis as "reduced form." This is not the traditional sense of the phrase. A classic reduced-form analysis (see e.g., Jacob Marschak, 1953) first specifies a structural model and then derives the reduced-form parameters in terms of the structural parameters. While many recent studies have asserted to taking a reduced-form approach, the structural model from which the reduced-form model should correspond is rarely specified. We explicitly specify a structural model and then use the implied reduced-form structure to estimate the effect of tuition subsidy policies. Specifying the underlying model has the advantage of being explicit about the assumptions that justify the analysis. This avoids Mark Rosenzweig and Kenneth Wolpin's (2000 pp. 853–62) criticism of work on natural "natural experiments" that often leaves these conditions implicit. Our structural model is based on the model studied by Michael Keane and Wolpin (2001). It is highly nonlinear and allows for more unobserved heterogeneity than the typical simultaneous-equations framework that most previous work has used in reduced-form estimation. Using the specified structural model, we examine the assumptions discussed in Ichimura and Taber (2000) to justify reduced-form estimation of the policy effects.

* Ichimura: Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom, and the Centre for Microdata Methods and Practice; Taber: Department of Economics and the Institute for Policy Research, Northwestern University, 2003 Sheridan Road, Evanston, IL 60208. We thank Mark Rosenzweig for comments.

I. Model

The model we pose is strongly based on Keane and Wolpin (2001). The decision problem in the model begins at age 14 when students have completed eighth grade. We let each period represent one year so that $t = 0$ when the individual is 14. During that year the student chooses whether to attend college or not and how much to work. We assume that schooling is completed by age 25. Let s_{it} be an indicator of whether individual i attends school in year t, ℓ_{it} the number of weeks worked in a year, and h_{it} the number of hours worked per week. Individuals have preferences for consumption, work, and schooling so that the utility function is $u_{it}(c_{it}, s_{it}, S_{it}, \ell_{it}, h_{it}, \mathbf{Z}_i^u)$, where S_{it} is cumulative years of school (i.e., $S_{it} = \sum_{\tau=0}^{t-1} s_{i\tau}$) and $\mathbf{Z}_i^u$ is an exogenous vector. For example, Keane and Wolpin (2001) let marriage and parental coresidence statuses be elements of $\mathbf{Z}_i^u$.

We assume that hourly wages take a standard Mincer form so that

$$(1) \qquad \log(w_{it}) = \beta_0 + \mathbf{Z}_i^{w'}\boldsymbol{\beta}_1 + \beta_2 L_{it} + \beta_3 L_{it}^2 + \alpha_i(S_{it}) + v_{it}$$

where $\mathbf{Z}_i^w$ is a set of observable non-time-varying variables that influence wages, v_{it} is an unobservable variable, and L_{it} is accumulated weeks of experience as of the beginning of the year. That is $L_{it} = \sum_{\tau=0}^{t-1} \ell_{i\tau}$. We take equation (1) as a structural relationship.

A key feature of this model is that we allow the schooling effect function $\alpha_i(\cdot)$ to vary arbitrarily across individuals in the model.

Accumulation of assets (K_{it}) by students is complicated by two features of the model. First, students are borrowing-constrained in two different senses. There is a limitation to the amount of debt that they can accumulate, which may be individual-specific. Let this amount be $-\bar{K}_i$. They also borrow at rate R_i^{b} but lend at rate R^{ℓ}. Second, they may receive transfers from their parents.

These transfers are individual-specific and potentially depend on their schooling choices, college costs, their asset levels, and their wage rate. This yields the following budget constraint:

$$(2)\quad K_{i,t+1} = K_{it}[R_i^{\mathrm{b}}1(K_{it} \le 0) + R^{\ell}1(K_{it} > 0)] - c_{it} + \ell_{it}h_{it}w_{it} + P_{it}(s_{it}, S_{it}, \tau_{\mathrm{ic}} - \pi_i) - (\tau_{\mathrm{ic}} - \pi_i)s_{it}1(4 \le S_{it} < 8)$$

$$(3)\quad K_{i,t+1} \ge -\bar{K}_i$$

where τ_{ic} is the monetary cost of attending college and π_i is the value of a tuition subsidy offered to individual i if the individual attends college. The notation $1(\cdot)$ denotes the indicator function, taking the value 1 if its argument is true and 0 otherwise. The function $P_{it}(s_{it}, S_{it}, \tau_{ic} - \pi_i)$ is the transfer from parents to students. We assume that the student takes this transfer function as given. Note that we allow this transfer to depend directly on the net costs of college $(\tau_{ic} - \pi_i)$, while Keane and Wolpin (2001) do not. A key feature of the model is that the tuition subsidies we consider operate by changing the net cost of college only. For example, we do not allow for general-equilibrium effects of the type described by James Heckman et al. (1998). This is valid if the simulations we consider relate to small programs that would only affect a small fraction of the labor market.

The state variables in this model are assets, schooling, and experience (K_{it}, S_{it}, L_{it}). The control variables are consumption, school enrollment, and weeks per year and hours per week of labor supplied $(c_{it}, s_{it}, \ell_{it}, h_{it})$. Let $E_{it}(\cdot)$ represent the expectation of agent i with respect to information he has at age t. Thus one can write this model using Bellman's equation as

$$V_{it}(K_{it}, S_{it}, L_{it}) = \max_{c_{it}, s_{it}, \ell_{it}, h_{it}} \{u_{it}(c_{it}, s_{it}, S_{it}, \ell_{it}, h_{it}) + \delta E_{it}[V_{i,t+1}(K_{i,t+1}, S_{i,t+1}, L_{i,t+1})]\}$$

subject to (2) and (3).

Individuals vary in the population because of heterogeneity in preferences (u_{it}), information (E_{it}), wages (w_{it}), parental transfers (P_{it}), and borrowing rates and asset level $(R_i^{\mathrm{b}}, K_{it})$ and college costs (τ_{ic}).

Loosely following Keane and Wolpin (2001), we assume that

$$u_{it}(c_{it}, s_{it}, S_{it}, \ell_{it}, h_{it}) = \frac{c_{it}^{\gamma}}{\gamma} + \lambda_{is_1}s_{it}1(S_{it} < 4) + \lambda_{is_2}s_{it}1(4 \le S_{it} < 8) + \lambda_{is_3}s_{it}1(8 \le S_{it}) - \lambda_{i\ell}\frac{\ell_{it}^{\alpha}}{\alpha} - \lambda_{ih}\frac{h_{it}^{\rho}}{\rho}$$

$$P_{it}(s_{it}, S_{it}, \tau_{\mathrm{ic}} - \pi_i) = \lambda_{ip_1} + \lambda_{ip_2}s_{it}1(4 \le S_{it} < 8) \times (t_0 + \lambda_{ip_3}[\tau_{\mathrm{ic}} - \pi_i]).$$

Given this setup, we can solve for the optimal profile of schooling, consumption, savings, and leisure.

There are a number of sources of heterogeneity in the model. We model them in the following manner:

$$\lambda_{is_j} = \mathbf{Z}_i^{s'}\boldsymbol{\gamma}^s + \theta_i^{s_j} + v_{it}^{s_j} \qquad \text{for } j = 1, 2, 3$$

$$\lambda_{ip_j} = \mathbf{Z}_i^{p'}\boldsymbol{\gamma}^p + \theta_i^{p_j} + v_{it}^{p_j} \qquad \text{for } j = 1, 2, 3$$

$$\lambda_{ij} = \mathbf{Z}_i^{j'}\boldsymbol{\gamma}^j + \theta_i^{j} + v_{it}^{j} \qquad \text{for } j = \ell, h$$

$$\tau_{\mathrm{ic}} = \mathbf{Z}_i^{\tau'}\boldsymbol{\gamma}^{\tau} + \theta_i^{\tau} + v_{it}^{\tau}$$

$$\alpha_i(S_{it}) = \theta_{i0}^{w} + \theta_{i1}^{w}S_{it} + \theta_{i2}^{w}S_{it}1\ (4 \le S_{it}) + \theta_{i2}^{w}S_{it}1\ (8 \le S_{it})$$

where we assume that the error terms (terms with θ or v) are all independent of the observable variables (terms with $\mathbf{Z}$). Let $\mathbf{Z}_i$ comprise all of the components of $(\mathbf{Z}_i^w, \mathbf{Z}_i^s, \mathbf{Z}_i^{\ell}, \mathbf{Z}_i^h, \mathbf{Z}_i^p, \mathbf{Z}_i^{\tau})$.

II. Differences from the Keane-Wolpin Model

The first main difference between our model and that of Keane and Wolpin (2001) is that our model does not require specifying the distribution of unobserved variables (i.e., the v_{it} and θ_i terms) and the agent's process of learning about the value of these unobservables. Keane and Wolpin estimate the model by maximum likelihood, which calls for specifying the joint distribution. Consistency of the estimator depends on the correct specification of this distribution. Taber (2000) considers semiparametric identification of a similar class of dynamic-programming schooling models and shows that the conditions are very strong and cannot be met with the data at hand.

The second main difference is that we will exploit the variation in tuition under the assumption that (i) the variable is independent from the unobserved variable conditional on observed exogenous variables of the model, $\mathbf{Z}_i$, and (ii) the variable is excluded from the outcome variables to estimate the impact of tuition-subsidy policy.

If we could consistently estimate all the structural parameters, then we would be able to evaluate any policy that can be addressed within the model. We are not able to carry out this more general exercise. However, under the maintained assumptions (i) and (ii) and based on some aspect of the described model, we will be able to estimate consistently the tuition-subsidy policy effect without making distributional assumptions about the error terms.

III. Estimation

We do not need to estimate the full structural parameters in order to evaluate the tuition-subsidy policy. Instead, we will estimate a semiparametric reduced-form version of this model.

This is essentially a six-index model. That is, conditional on $\mathbf{Z}_i$ and age (A_{it}), for any potential outcome Y_{it} we can write

$$E(Y_{it}|\mathbf{Z}_i, A_{it}, \pi_i) = F_Y(\mathbf{Z}_i^{\tau'}\boldsymbol{\gamma}^\tau - \pi_i, \mathbf{Z}_i^{\ell'}\boldsymbol{\gamma}^\ell, \mathbf{Z}_i^{h'}\boldsymbol{\gamma}^h, \mathbf{Z}_i^{p'}\boldsymbol{\gamma}^p, \mathbf{Z}_i^{s'}\boldsymbol{\gamma}^s, \mathbf{Z}_i^{w'}\boldsymbol{\beta}_1, A_{it}).$$

We will examine the following five outcome variables: college attendance, years of schooling, weeks worked per year, hours worked per week, and log wage. The data are generated in the absence of a tuition subsidy ($\pi_i = 0$). Our goal is to simulate the effect of a tuition subsidy, summarized by π_i', which may vary across individuals. Our goal is thus to estimate $E(Y_{it}|\mathbf{Z}_i, A_{it}, \pi_i')$.

We define the net tuition index $\mathbf{Z}_i^{\tau'}\boldsymbol{\gamma}^\tau - \pi_i$ to be

$$\text{tuition}_{17i} + \gamma^\tau \text{anycol}_{17i} - \pi_i$$

where tuition_{i17} is the average tuition paid at two-year colleges in the state in which individual i lives at age 17, and anycol_{17i} is a dummy variable indicating whether there was a college in the county in which the individual lived at age 17. David Card (1995) was the first to use college in county as an instrumental variable for schooling. We follow Stephen Cameron and Taber (2001) by assuming that this represents a financial cost of college as one can live at home and attend a college if it is close. The structural model imposes two important restrictions that are being used here. The first is that these two variables in the cost-of-college index do not appear in any other indexes. The second is that the tuition subsidy enters the index as specified above.

With an additional regularity condition, one can identify γ^τ and other indexes when one normalizes the unknown functions' first partial derivatives suitably. We do not need to impose any other exclusion restriction as it is not needed. Thus, all variables in different indexes except for the tuition index are the same:

$$\mathbf{Z}_i^s = \mathbf{Z}_i^\ell = \mathbf{Z}_i^h = \mathbf{Z}_i^p = \mathbf{Z}_i^w = \mathbf{X}_i.$$

For clarity we will write variables in the tuition index as $\mathbf{T}_i$. Thus $\mathbf{Z}_i^\tau = \mathbf{T}_i$.

One strategy would be to estimate F_Y and the indexes of the model up to normalizations as discussed in the appendix to Ichimura and Taber (2002). This involves high-dimensional nonlinear optimization. Notice that the scale of coefficients are not identified. Furthermore, unless some exclusion restrictions are satisfied among indexes, there are further identification problems as discussed by Ichimura and Lung-Fei

Lee (1991). For our purpose, the crucial index is $\mathbf{T}_i'\boldsymbol{\gamma}^\tau$, and we assume that tuition variables affect decision only through this index. Under this assumption, the coefficients of this index are identified, and thus the tuition-policy effect is identified.

Let the model be

$$E(\mathbf{Y}|\mathbf{X}, \mathbf{T}) = f(\mathbf{XB}, \mathbf{T}\boldsymbol{\theta})$$

where $\mathbf{Y}$ is a $r \times 1$ vector, $\mathbf{X}$ is a $1 \times k_1$ regressor vector, $\mathbf{B}$ is a $k_1 \times r$ matrix ($k_1 > r$), $\mathbf{T}$ is a $1 \times k_2$ regressor vector, and $\boldsymbol{\theta}$ is a $k_2 \times 1$ vector ($k_2 > 1$); and where $\mathbf{f} = (f^1, \ldots, f^r)'$ and f^j ($j = 1, \ldots, r$) are continuously differentiable functions from R^{r+1} into R.

This representation of the model is a classic semiparametric reduced form of the original structural model. The goal of this paper is to estimate the effect of tuition subsidies. We assume that there is no common regressor among $\mathbf{X}$ and $\mathbf{T}$ and that there is no a priori zero restriction in the $\mathbf{B}$ matrix. We denote the partial derivative of f^j with respect to the pth argument by f^j_p.

In our application, $\mathbf{T}_i'\boldsymbol{\gamma}^\tau$ is $\mathbf{T}\boldsymbol{\theta}$, the remaining five indexes correspond to $\mathbf{XB}$, and $r = 5$. The five outcomes are: whether one goes to college or not, years of education, weeks worked, hours worked, and log wage.

We show the following: Let $\delta_{jp} = 1$ if $j = p$ and $\delta_{jp} = 0$ otherwise, and denote the matrix with the j, pth element being $f^j_p(0)$ by $\mathbf{A}$.

THEOREM 1: *Let* ($\mathbf{X}$, $\mathbf{T}$) *be continuously distributed with its support being* $R^{k_1+k_2}$. *For any* $\mathbf{B}$, $\boldsymbol{\theta}$, *and* $\mathbf{f}$, *with* θ_1 (*the first element of* θ) *not zero, there exists a unique* $\mathbf{B}^*$, $\boldsymbol{\theta}^*$ *with* $\theta_1^* = 1$, *and* $\tilde{\mathbf{f}}$ *such that for* $j = 1, \ldots, r$ *and* $p = 1, \ldots, r$,

$$\tilde{f}^j_p(0) = \delta_{jp}$$

and $\mathbf{f}(\mathbf{XB}, \mathbf{T}\boldsymbol{\theta}) = \tilde{\mathbf{f}}(\mathbf{XB}^*, \mathbf{T}\boldsymbol{\theta}^*)$ *almost everywhere if matrix* $\mathbf{A}$ *is nonsingular and* $\mathbf{B}$ *has full column rank.*

A proof is given in Ichimura and Taber (2002). This normalization is chosen so that it can be imposed easily when we use polynomial expansion to approximate the unknown function $\mathbf{f}$.

In principle, this model can be estimated by nonlinear least squares, but directly implementing it will involve a high-dimensional nonlinear optimization problem. Hence, we take an alternative approach.

Note that the nonlinearity mainly arises because we impose the model restriction that the coefficients in the "squared and cross terms" are related to the linear terms. If we relax this restriction, the dimension of nonlinearity reduces to 1.

The implication of this identification result is that the coefficients on the linear terms of $\mathbf{X}_i$ can be interpreted as the relevant index coefficients under the normalization specified in the theorem.

This observation can be used to construct a specification test of the index restriction, and also the estimation result can be used as the first step to impose the restriction implied by the index model.

When we take this approach and do not impose the cross-index restriction between the linear term and quadratic terms except for the tuition index, then the approach is equivalent to specifying a more flexible form: for any outcome variable Y_{it},

$$E(Y_{it}|\mathbf{Z}_i, A_{it}, \pi_i) = G_Y(\mathbf{T}_i'\boldsymbol{\gamma}^\tau - \pi_i, \mathbf{X}_i, A_{it}).$$

We can identify/estimate the effect of a tuition subsidy of level π' on outcome Y_{it} in the following manner:

(i) using any particular outcome, identify/estimate $\boldsymbol{\gamma}^\tau$ and G_Y for that outcome;
(ii) identify/estimate G_Y corresponding to all the rest of outcomes;
(iii) simulate $E(Y_{it}|\mathbf{Z}_i, A_{it}, \pi_i') = G_Y(\mathbf{T}_i'\boldsymbol{\gamma}^\tau - \pi_i', \mathbf{X}_i, A_{it})$.

The basic procedure for the first-stage estimation of the index is similar to semiparametric least squares (Ichimura, 1993). Let $\hat{G}_Y$ represent fit of a type of nonparametric regression of Y onto ($\mathbf{T}'\boldsymbol{\gamma}^\tau$, $\mathbf{X}$, $\mathbf{A}$). Note that the $\boldsymbol{\gamma}^\tau$ parameter is overidentified as there are multiple outcomes. Simple models suggest that an easy way to estimate the effects of tuition on outcomes without sacrificing efficiency too much may be to use college attendance (D_i). Thus, we estimate the following:

$$\hat{\gamma}^{\tau} = \arg\min \sum_{i=1}^{N} [D_i - \hat{G}_D(\text{tuition}_{17i} + \gamma^{\tau}\text{anycol}_{17i}, \mathbf{X}_i)]^2.$$

Age is dropped, as D_i is constructed for individuals over 25 and under our assumption individuals do not attend college after this.

In estimation we use a second-order polynomial (including all interactions) as our flexible functional form for G_Y. There are of course other possibilities. Note that given γ^{τ}, the use of a quadratic function for $\hat{G}_D$ implies that the rest of the parameters can be explicitly computed. Stage 2 of the estimation is also performed using second-order polynomial approximations.

A complication arises in interpreting the results. Wages at a given age depend on schooling both through its effect on experience and its direct effect. It is standard in the empirical literature to summarize the effect of schooling on wages holding experience constant. If we do not condition on experience, the effect of tuition subsidy on wages will depend on the level of experience and, thus, the age. However, given the structure above, this is not the case if we condition on experience. Since experience is endogenous (from the perspective of the model), conditioning on it is not trivial. However, given the form of the wage equation (1) above, there is a simple solution. Note that schooling only enters the model through $\alpha_i(S_i)$, which is fixed once individuals have entered the labor market. Thus, one can deal with the endogeneity of S_i using a fixed-effects strategy. There is another potential problem that labor supply, and thus L_{it}, will be correlated with v_{it}. We thus estimate β_2 and β_3 using fixed effects and instrumental variables with age and age-squared as instruments for actual experience and experience-squared. We can then report the effect of tuition subsidies on wages holding experience constant by using $Y_{it} = \log(w_{it}) - \hat{\beta}_2 L_{it} - \hat{\beta}_3 L_{it}^2$, where L_{it} is the level of experience without the tuition subsidy, as our outcome variable. We then proceed with this outcome variable in exactly the same way that we do with the other outcomes.

TABLE 1—SIMULATED EFFECTS OF TUITION SUBSIDY ON VARIOUS OUTCOMES—TUITION ONLY

Outcome	\$1,000 subsidy[a]	Means-tests subsidy[b]	Phased-out subsidy[c]
College attendance	0.0453	0.0604	0.0854
	(0.0106, 0.088)	(0.004, 0.109)	(0.009, 0.149)
Years of schooling	0.0447	0.1880	0.2286
	(−0.152, 0.148)	(−0.102, 0.343)	(−0.162, 0.457)
Weeks worked per year	0.2134	−0.0676	0.0288
	(−0.505, 0.960)	(−1.299, 1.363)	(−1.558, 1.906)
Hours per week	0.8260	1.4346	1.9093
	(0.066, 1.156)	(0.508, 2.514)	(0.771, 3.312)
Log wages	0.0176	0.0055	0.0139
	(−0.016, 0.055)	(−0.040, 0.067)	(−0.050, 0.092)
Log wages[d]	0.0133	0.0024	0.0091
	(−0.019, 0.050)	(−0.032, 0.056)	(−0.039, 0.083)
Implicit return to schooling[e]:	0.2984	0.0126	0.0398
Fraction eligible (percent):	100	31	69

Notes: Results are estimated from the National Longitudinal Survey of Youth (NLSY). The cost-of-schooling index is tuition only. The analysis considers men only. Bootstrapped 95-percent confidence limits are reported in parentheses.

[a] A \$1,000 tuition subsidy is given to anyone attending college.

[b] A \$1,000 tuition subsidy is given to individuals attending college whose family income is less than \$20,000. The number reported is for the eligible population (family income <\$20,000).

[c] The tuition subsidy is \$2,000 for students whose family income is less than \$20,000. The subsidy fades linearly to zero at \$40,000. The numbers reported are for the eligible population (family income <\$20,000).

[d] This row reports the effect of the policy on wages holding experience constant.

[e] The implicit return to schooling is simply the effect on log wages (conditional on experience) divided by the effect on years of schooling.

IV. Empirical Results

We present the results using the methodology described above. We consider three different types of tuition subsidies. The first is a straight \$1,000 tuition subsidy. That is, every individual would receive a subsidy of \$1,000 if he attends college. These results are presented in the first column of Tables 1 and 2. The second is means-tested and targeted toward low-income families. Only students whose parents income are less than \$20,000 are eligible for the \$1,000 subsidy. These results are presented in column (ii) (of both tables), and only averages for the eligible population (i.e., students from families who earn less than \$20,000) are presented. The third subsidy is more complicated. Students from families who earn less than \$20,000 are eligible for a \$2,000 subsidy. The subsidy then phases out at a rate of \$100 per \$1,000 of family income until it disappears at family income level \$40,000. We present these results in col-

TABLE 2—SIMULATED EFFECTS OF TUITION SUBSIDY ON VARIOUS OUTCOMES: TUITION AND WHETHER COLLEGE IS IN-COUNTY

Outcome	$1,000 subsidy	Means-tests subsidy	Phased-out subsidy
College attendance	0.0091	0.0029	0.0069
	(−0.022, 0.056)	(−0.012, 0.076)	(−0.021, 0.110)
Years of schooling	0.0254	0.0080	0.0194
	(−0.084, 0.143)	(−0.079, 0.288)	(−0.101, 0.364)
Weeks worked per year	0.0175	0.0331	0.0424
	(−0.142, 0.619)	(−0.261, 0.765)	(−0.336, 1.114)
Hours per week	−0.0836	−0.1243	−0.1712
	(−0.241, 0.433)	(−0.309, 0.672)	(−0.408, 0.922)
Log wages	0.0021	0.0011	0.0019
	(−0.004, 0.034)	(−0.007, 0.034)	(−0.006, 0.050)
Log wages[a]	0.0025	−0.0002	0.0008
	(−0.003, 0.033)	(−0.010, 0.023)	(−0.010, 0.041)
Implicit return to schooling:	0.0993	−0.0243	0.0399
Fraction eligible (percent):	100	31	69

Notes: Results are estimated from the National Longitudinal Survey of Youth (NLSY). The cost-of-schooling index is based on tuition and whether there is a college in the county of residence. The analysis considers men only. Bootstrapped 95-percent confidence limits are reported in parentheses. See Table 1 for explanation of columns.

[a] This row reports the effect of policy on wages holding experience constant.

umn (iii) (of both tables) only for students from families that earn less than $40,000.

We consider two different specifications. In the first, the only variable included in the cost of schooling index is tuition itself. A description of the basic data used can be found in Ichimura and Taber (2000, 2002). The first set of results are presented in Table 1. We examine several different outcomes. The effect of the tuition subsidy on college attendance is well within the range in the literature (see Larrt Leslie and Paul Brinkman, 1988). The effect is somewhat larger when targeted at the lower-income population, but one can see from the confidence intervals that the results are not very precise. Unfortunately, the confidence intervals are large on the other outcomes (particularly years of schooling and wages), making it hard to draw strong conclusions. To give some manner of judging the magnitude of the wage effect, we present the "implicit return to schooling effect" in the next-to-last row. This is simply the wage effect (controlling for experience) divided by the increase in years of schooling. This is the average return to a year of schooling averaged over the additional years of schooling that result from the subsidy. Given the confidence intervals of both the numerator and denominator, the effects are not well pinned down.

To help the precision we expanded the index of cost of college to include the presence of a college in one's county, following Cameron and Taber (2001). This variable tends to be a much stronger predictor than tuition itself, so we expected this to help the precision of the estimates substantially. This had the effect of substantially lowering the main effect of tuition on college attendance. The confidence intervals tend to be somewhat smaller in this table, but still very wide. We suspect that one reason why inclusion of college in county did not help that much is that it is a dummy variable and works off quite different variation in the data than tuition itself (and the tuition subsidy). There are a number of ways to try to improve this procedure, which we leave for future research.

V. Conclusions

In this paper we perform semiparametric reduced-form analysis to simulate the effects of several tuition-subsidy programs on a variety of outcomes. This is classical reduced-form estimation, in that we present a full structural model but only estimate a reduced form of the model. Presentation of the full model is necessary to justify the use of the reduced form for policy simulation. This is an example of the type of direct estimation of policy effects described more generally in Ichimura and Taber (2000). The basic idea is to specify the minimal level of conditions necessary for identification of the policy effect and then use those conditions to simulate the effect. When one has the limited objective of simulating a particular policy this will often have an advantage over estimation of the full structural model in that it requires weaker functional form requirements. It may also be computationally easier to estimate the reduced-form model than the full structural model. That is certainly the case in this example. On the other hand, we do require tuition variation which, under the maintained model, can be regarded as equivalent to the tuition-subsidy variation.

Using this approach we simulate several different tuition subsidies. As in other studies, we find that tuition subsidies have a substantial effect on college attendance. There is some weak evidence that this effect is likely to be larger when the subsidy is geared toward students from low-income families. The confidence intervals are wide, making it hard to make strong statements about the effect of tuition subsidies on other outcomes.

REFERENCES

Cameron, Stephen and Taber, Christopher. "Estimating Borrowing Constraints Using the Returns to Schooling." Unpublished manuscript, Northwestern University, 2001.

Card, David. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky, eds., *Aspects of labor market behavior: Essays in honour of John Vanderkamp*. Toronto: University of Toronto Press, 1995, pp. 201–22.

Heckman, James; Lochner, Lance and Taber, Christopher. "General Equilibrium Treatment Effects." *American Economic Review*, May 1998 (*Papers and Proceedings*), *88*(2), pp. 381–86.

Ichimura, Hidehiko. "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models." *Journal of Econometrics*, July 1993, *58*(1–2), pp. 71–120.

Ichimura, Hidehiko and Lee, Lung-Fei. "Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation," in W. A. Barnett, J. Powell, and G. Tauchen, eds., *Nonparametric and semiparametric methods in economics and statistics*. Cambridge, U.K.: Cambridge University Press, 1991, pp. 3–49.

Ichimura, Hidehiko and Taber, Christopher. "Direct Estimation of Policy Impacts." National Bureau of Economic Research (Cambridge, MA) Technical Working Paper No. 254, 2000.

______ . "Semiparametric Reduced Form Estimation of Tuition Subsidies." Unpublished manuscript, University College London, United Kingdom, 2002.

Keane, Michael P. and Wolpin, Kenneth I. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review*, November 2001, *42*(4), pp. 1051–1103.

Leslie, Larrt and Brinkman, Paul. *The economic value of higher education*. New York: Macmillan, 1988.

Marschak, Jacob. "Economic Measurements for Policy and Prediction," in W. Hood and T. Koopmans, eds., *Studies in econometric method*. New York: Wiley, 1953, pp. 1–26.

Rosenzweig, Mark R. and Wolpin, Kenneth I. "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature*, December 2000, *38*(4), pp. 827–74.

Taber, Christopher. "Semiparametric Identification and Heterogeneities in Discrete Choice Dynamic Programming Models." *Journal of Econometrics*, June 2000, *96*(2), pp. 201–29.