3 The Limits of Inference without Theory

3.1 The Role of Theory in Econometric Specification

3.1.1 The Effect of Unemployment Compensation on the Duration of Unemployment

3.1.1.1 Background

In 1977, the *Industrial and Labor Relations Review* (*ILRR*) published a conference volume, "The Economics of Unemployment Insurance: A Symposium." The purpose of the volume was to gather together papers that would evaluate the extent to which the unemployment insurance (UI) system affects labor market behavior. Four of the papers were original empirical studies attempting to provide quantitative estimates of the effect of UI benefit generosity on the search behavior of the unemployed (Classen, 1977; Ehrenberg and Oaxaca, 1977; Holen, 1977; Burgess and Kingston, 1976), one of the papers surveyed the empirical literature on that question (Welch, 1977), and one provided a theoretical model for interpreting the estimates from the empirical literature (Mortensen, 1977).[1] According to the editor's introduction to the volume, the empirical contributions were founded on "recent job search theories," that is, on the (then new) sequential job search model of McCall (1970) and Mortensen (1970).

In the classification scheme I have adopted, these empirical papers would be considered quasi-structural in the sense that the parameters of the estimating equations, which are (approximations of) decision outcomes, are functions of the underlying parameters of the optimization problem. Since that symposium, there have been dozens of quasi-structural papers addressing the same question, either implicitly or explicitly based on the sequential search model (or its extensions). Over time the statistical procedures used in estimation have been improved. However, the basic specification of the relationship between search outcomes, the duration of unemployment in particular, and their determinants has not been altered. That commonality, of course, would be appropriate had the connection between the theory and the empirical specification been unassailable from the beginning. However, that was not the case then, and, unfortunately, it is not the case now.

Given the development of estimation methods for discrete choice dynamic programming (DCDP) models, it was natural that the sequential job search model would be structurally implemented. Structural estimation of the standard nonstationary partial equilibrium job search model was first considered by Wolpin (1987) and van den Berg (1990) and later extended beyond the standard model.[2] Examples include Stern (1989), who allowed for simultaneous search, that is, for the submission of multiple job applications in a period; Blau (1991), who dropped the assumption of wealth maximization, allowing for job offers to include not only a wage offer but also an hours offer; Ferrall (1997), who modeled search accounting for all of the rules of the Canadian UI system; Gemici (2007), who considered the joint husband and wife search-migration decision in an intra-household bargaining framework; and Paserman (2008), who adopted a behavioral approach and allowed for hyperbolic discounting. The standard search model has also been extended beyond the consideration of the single transition from unemployment to employment. Wolpin (1992) incorporated job-to-job transitions and both involuntary and voluntary transitions into unemployment.[3] Rendon (2006) allowed for a savings decision in a setting where agents can also transit, both through quits and layoffs, from employment to unemployment.[4] Both of these latter papers also allowed for wage growth with the accumulation of work experience, employer-specific (tenure) in the case of Rendon (2006) and both general and employer-specific in the case of Wolpin (1992).

The standard job search model assumes that *ex ante* identical workers may receive different wage offers or, analogously, that the same unemployed worker may receive different offers over time. Diamond (1971) showed that with the assumptions of the standard job search model, in a game in which firms are aware of worker search strategies, the wage offer distribution will be degenerate at the worker's reservation wage or outside option. This result led to the development of models in which wage dispersion could be rationalized as an equilibrium outcome, which in turn led to a structural empirical literature. The empirical literature has focused on two kinds of models, those based on a search-matching-bargaining approach to wage determination (Diamond and Maskin, 1979; Mortensen, 1982; Wolinsky, 1987) and those based on wage posting by firms that gain monopsony power through search frictions (Albrecht and Axell, 1984; Burdett and Mortensen, 1998). These models not only rationalized wage dispersion but also allowed for quantification and policy analyses in an equilibrium setting, for example, for changes in UI benefits or changes in the level of the minimum wage.

The empirical implementation of equilibrium search models has become a major strand of the structural job search literature. Embedded within those models are different variants of the standard partial equilibrium search model, and, in that sense, the development of the DCDP estimation approach was an important precursor. That literature is instructive in terms of empirical methods. The empirical literature on estimating equilibrium search models began with the implementation of the wage-posting model; Eckstein and Wolpin (1990) implemented the Albrecht and Axell (1984) model, followed by a number of papers based on the Burdett and Mortensen (1998) model. Without going into details about them individually, the hallmark of those papers was the development of theoretical extensions of the Burdett and Mortensen (1998) model, incorporating more realistic features of the labor market that went hand in hand with the empirical implementation. These papers include contributions by Kiefer and Neumann (1993), van den Berg and Ridder (1998), Bontemps, Robin, and van den Berg (1999, 2000), Bowlus, Kiefer, and Neumann (2001), and Postel-Vinay and Robin (2002). Papers based on search-matching-bargaining models include, among others, Eckstein and Wolpin (1995), Cahuc, Postel-Vinay, and Robin (2006), and Flinn (2006). The papers that have resulted from the equilibrium search enterprise in structural estimation demonstrate the best examples of combining theory, econometrics, and data to produce empirical results that are policy relevant and that have the potential to fundamentally alter our view of the working of the labor market.[5]

As I have noted, the conventional empirical specification in the quasi-structural literature, although putatively based on the standard search model, has not been faithful to the theory. To demonstrate the possible consequences of that failure, and to make the discussion self-contained, I begin with a brief review of the standard search model.[6]

3.1.1.2 *The Standard Discrete-Time Job Search Model*

In the discrete-time formulation, an unemployed individual receives a job offer in each period with probability q. Wage offers are drawn from the known cumulative distribution function, $F(w)$. An accepted job offer and its associated wage are permanent. While unemployed, an individual receives unemployment benefits, which, net of the cost of search, is denoted by b. The individual is assumed to maximize the present discounted value of net income. We consider both infinite and finite horizon models, which have somewhat different empirical implications, and implications for the identification of model parameters.

Infinite Horizon Model

The value of a wage offer of w, given a discount factor of $\delta(=\frac{1}{1+r})$, is

$$W(w) = w + \delta w + \delta^2 w + \ldots = \frac{w}{1-\delta}.$$

In any period, the value of continuing to search, V, either because an offer was rejected or because an offer was not received, consists of the current period payoff (unemployment benefits net of the cost of search), b, plus the discounted expected value of waiting another period. In that case, if an offer is received, with probability q, the individual chooses between the maximum of the value of working at a wage w, $W(w)$, and the value of continuing to search, V. If no offer is received, which occurs with probability $1 - q$, the individual must continue to search. Thus, the alternative-specific value function, the Bellman equation, for the search decision is

$$V = b + \delta\left[qE\max(W(w), V) + (1-q)V\right].$$

Rearranging yields

$$V(1-\delta) = b + \delta qE\max(W(w) - V, 0),$$

which has a unique solution for $V > 0$ as long as the cost of search is not so large as to make the right-hand side negative.[7] Defining w^*, the reservation wage, to be the wage offer that equates the value of search and the value of accepting the job, that is, $w^* = (1 - \delta)V$, a little further algebra yields the following implicit equation for the reservation wage (which must have a unique solution given that V does):

$$w^* = b + \frac{q}{r}\int_{w^*}^{\infty}(w - w^*)dF(w).^8 \tag{3.1}$$

Thus, the reservation wage is a function of $b, \frac{q}{r}$, and $F(w)$:

$$w^* = w^*(b, \tfrac{q}{r}, F). \tag{3.2}$$

The individual accepts any wage offer that exceeds the reservation wage and declines offers otherwise.

Although the reservation wage is a deterministic function, the length of an unemployment spell is stochastic because the timing and level of wage offers are probabilistic. Thus, measures of the outcomes of search, such as the duration of unemployment spells and the level of accepted wages, are probabilistic. In particular, the survivor function, the probability that the duration of unemployment is at least as large as some given length, is

$$\Pr(T_u \ge t_u) = [qF(w^*) + (1-q)]^{t_u} = [1 - q(1 - F(w^*))]^{t_u}. \tag{3.3}$$

The term inside the brackets in equation 3.3 is the probability of receiving an offer in a period and rejecting it (because it is below the reservation wage) plus the probability of not receiving an offer. The *cdf*, *pdf*, and hazard function are:

$$cdf : \Pr(T_u < t_u) = 1 - [1 - q(1 - F(w^*))]^{t_u},$$

$$pdf : \Pr(T_u = t_u) = [1 - q(1 - F(w^*))]^{t_u} q(1 - F(w^*))$$

$$.HazardFunction : \Pr(T_u = t_u \mid T_u \ge t_u) = \frac{\Pr(T_u = t_u)}{\Pr(T_u \ge t_u)} = q(1 - F(w^*)) = h.$$

The survivor function, the *cdf*, and the *pdf* can all be written as functions of the hazard rate, the exit rate from unemployment conditional on not having previously exited. As seen, the hazard rate is constant. Thus, in a homogeneous population, the infinite horizon search model implies the absence of duration dependence.

Given parameter values, mean duration is given by

$$E(t_u) = \sum_0^{\infty} t_u \Pr(T_u = t_u)$$
$$= q(1 - F(w^*))^{-1} = \frac{1}{h}.$$

Notice that mean duration is simply the reciprocal of the hazard rate.[9] Likewise, the mean of the accepted wage is

$$E(w \mid w \ge w^*) = \int_{w^*}^{\infty} \frac{w}{1 - F(w^*)} dF(w),$$

which clearly is larger than the mean of the wage offer distribution.[10]

In addition to implying a constant hazard rate, the infinite horizon model has predictions about the impact of changes in $b, \frac{q}{r}$, and $F(w)$ on the reservation wage, on the hazard rate, and on the moments of the accepted wage distribution. It thus is, in principle, possible to test the theory. Comparative static effects of the hazard rate (and thus mean duration) with respect to its arguments are given by (see Mortensen, 1986):

$$\frac{dh}{db} = -\lambda f(w^*)\left(\frac{1}{1+h/r}\right) < 0,$$

$$\frac{dh}{dq} = -f(w^*)\left(\frac{w^*-b}{1+h/r}\right) + (1-F(w^*)) \gtrless 0,$$

$$\frac{dh}{d\mu} = -qf(w^*)\left(\frac{h/r}{1+h/r}\right) > 0,$$

$$\frac{dh}{ds} = -qf(w^*)\left(\frac{(\lambda/r)\int_0^{w^*} F_s(w,s)dw}{1+h/r}\right) < 0.$$

An increase in the level of unemployment compensation benefits increases the reservation wage and reduces the unemployment hazard rate. An increase in the offer probability has an ambiguous effect: it increases the reservation wage, which reduces the hazard rate, but also directly increases the hazard rate through the higher offer probability.[11] Increasing the mean of the wage offer distribution, μ, increases the hazard rate; although an increase in μ increases the reservation wage, the increase is less than one for one. Finally, increasing the mean preserving spread of the distribution, s, reduces the hazard because an increase in the mass of the right tail of the wage offer distribution increases the payoff to search, thus increasing the reservation wage. An additional set of implications follow about the mean of the accepted wage; anything that increases the reservation wage also increases the mean accepted wage.

Quasi-Structural Approach to Estimation

As exemplified by the papers in the *ILRR* symposium, the early quasi-structural approach to estimating the job search model was regression based. The primary concern of that literature, as well as the later literature based on hazard modeling, was the estimation of the impact of unemployment benefits on the duration of unemployment and postunemployment wages.[12] As noted, the regression specification was motivated by the standard job search model. Classen (1977) provides a clear statement of the connection between the two. Classen (1977) writes equations representing the outcomes of the search process as

$$D = \alpha_0 + \alpha_1 WBA + \sum \alpha_i X_i + u_D,$$

$$Y = \beta_0 + \beta_1 WBA + \sum \beta_i X_i + u_Y,$$

where D is spell duration, Y is a measure of the postunemployment wage, *WBA* is the weekly UI benefit amount, and the X's are "proxies" for a worker's skill level, the cost of search, and the job offer rate. As Classen notes, the determinants of both spell duration and the postunemployment wage should be exactly the same, as they are joint outcomes derived from optimal search behavior. Among the proxy variables used in Classen's analysis are demographics, such as age, race, and sex, and a measure of the wage on the job held prior to beginning the unemployment spell. Although not included in the Classen study, other variables often included in this type of specification are education, marital status, number of dependents, and a measure of aggregate unemployment in the relevant labor market.

A test of the theory amounts to a test that benefits increase both expected duration and the mean accepted wage, that is, that α_1 and β_1 are both positive. Any further test of the theory would involve specifying how proxy variables are related to the structural parameters, q, the offer probability, and $F(w)$, the wage offer distribution.

Classen is clear as to the purpose of including the preunemployment wage, namely as a proxy, most directly perhaps for the mean of the wage offer distribution. However, although the inclusion of that variable or, as is also common, of the replacement rate, the ratio of the benefit level to the preunemployment wage, was and continues to be standard in the quasi- structural literature, the need for stating a rationale has been lost. Without an explicit rationale, the point that the inclusion of the preunemployment wage (or the replacement rate) cannot be justified by the standard search model (it does not appear in equation 3.2) and thus should not have any impact on search outcomes given q and $F(w)$, has also been lost.

Regardless of whether it is stated, the original rationale for its inclusion remains, namely to avoid omitted variable bias. For example, suppose that the preunemployment wage is meant to proxy the mean of the wage offer distribution, μ, which is assumed to vary in the sample. Now, UI benefits are usually tied to the preunemployment wage, at least up to some limit. Thus, if some determinants of μ are omitted, variation in the benefit level will reflect, in part, the fact that those with higher preunemployment wages have higher μ's. In that case, a positive correlation between UI benefit levels and the preunemployment wage will lead to a negative bias in the effect of UI benefits on the duration of unemployment (recall that the higher μ is, the greater will be the hazard rate).

Although omitted variable bias is well understood, the potential for bias introduced by using proxy variables is less well appreciated. The source of the problem is that, in the context of a search model, the preunemployment wage must have been the outcome of a search during a prior unemployment spell. To isolate the effect of using this proxy, assume that the duration of the prior unemployment spell was governed by the same behavioral process and fundamentals as the current spell. Such an assumption is consistent with a model in which there are exogenous layoffs and resulting unemployment spells are renewal processes (the stochastic properties of all unemployment spells are the same).[13] To simplify, suppose that only the benefit level and the mean of the wage offer distribution vary in the sample. Then, taking deviations from means (without renaming the variables, to conserve on notation), the duration equation for any individual i is

$$D_i = \pi_{11} b_i + \pi_{21} \mu_i + v_{1i}. \tag{3.4}$$

Assuming μ_i is unobserved and thus omitted from the regression, the bias in the ordinary least squares (OLS) estimator of π_1 is given by

$$E(\hat{\pi}_{11} - \pi_{11}) = \pi_{21} \frac{\sigma_\mu^2}{\sigma_b^2} \beta_{b,\mu},$$

where $\beta_{b,\mu}$ is the regression coefficient of b on μ. Thus, if b and μ are positively correlated, and $\pi_{21} < 0$ as theory suggests, the bias in the estimated effect of UI benefits on duration will be negative.

Now, because the preunemployment wage, Y_{-1}, is the result of a prior search, it will have the same arguments as equation 3.4, namely

$$\begin{aligned} Y_{i,-1} &= \pi_{12} b_i + \pi_{22} \mu_i + v_{2i} \\ &= \pi_{22} \mu_i + \omega_i, \end{aligned} \tag{3.5}$$

where $\omega_i = \pi_{12} b_i + v_{2i}$ and where $E(\omega b_i) \neq 0$ given the definition of ω_i. To derive a regression equation that includes b_i and $Y_{i,-1}$, solve for μ_i in equation 3.5 and substitute into equation 3.4, yielding

$$\begin{aligned} D_i &= \pi_{11} b_i + \pi_{21} \frac{(Y_{i,-1} - \omega_i)}{\pi_{22}} + v_{1i} \\ &= \pi_{11} b_i + \frac{\pi_{21}}{\pi_{22}} Y_{i,-1} + v'_{1i}, \end{aligned}$$

where $v'_{1i} = -\frac{\pi_{21}}{\pi_{22}} \omega_i + v_{1i}$. We are interested in whether the OLS estimator for π_{11} is biased, that is, whether $E(b_i v'_{1i} \mid Y_{i,-1}) = E(-\frac{\pi_{21}\pi_{12}}{\pi_{22}} b_i^2 \mid \pi_{12} b_i + \pi_{22}\mu_i + v_{2i}) = 0$. It is easiest to see whether this holds by explicitly deriving the bias expression. The bias is given by

$$E(\hat{\pi}_{11} - \pi_{11}) = \pi_{21}^2 \frac{\sigma_\omega^2 \sigma_\mu^2}{\sigma_b^2 \sigma_{Y_{-1}}^2 - (\sigma_{bY_{-1}})^2} (\pi_{21}^{-1} \beta_{b,\mu} - \pi_{22} \beta_{b,\omega}).$$

The bias is zero if either $\sigma_\omega^2 = \pi_{12}^2 \sigma_b^2 + \sigma_{v_2}^2 = 0$ or if $\pi_{21}^{-1} \beta_{b,\mu} - \pi_{22} \beta_{b,\omega} = 0$, either of which would be a fortuitous property of the sample.

What is the relationship between the biases that arise from omitting μ_i (omitted variable bias) and those from including $Y_{i,-1}$ (proxy variable bias)? It turns out that a sufficient condition for the bias from omitting the preunemployment wage to be smaller than from including it is that $\beta_{b,\omega} = \pi_{12} \sigma_b^2 = 0$, which only will hold if $\pi_{12} = 0$, a violation of the theory. In general, the biases cannot be ordered, and it is unclear which is the better strategy to follow to minimize the bias. The implicit (that is, without justification) assumption made by almost all researchers is that omitted variable bias is greater than proxy variable bias in this context. Moreover, if the benefit level varies with the preunemployment wage and we have good measures of the mean of the wage offer distribution, variation in the benefit level from this source is helpful in identifying the UI benefit effect. The variation in the preunemployment wage around the mean of the offer distribution that induces benefit variation is purely due to random draws from the wage offer distribution.

Researchers adopting the quasi-structural approach have universally included the preunemployment wage and thus assumed that omitted variable bias is greater than proxy variable bias. There is a larger point reflected by this choice. Structural work requires that all variables be explicitly accounted for in the model. A similar standard for quasi-structural work might have revealed the existence of the choice between omitted variable and proxy variable bias, a choice not explicitly acknowledged in the empirical literature.

As we noted, it is also usual to include some aggregate labor market statistic such as the local unemployment rate. The idea is that the aggregate statistic reflects labor market demand and so will affect the offer rate or the wage offer distribution. However, because the aggregate statistic is simply the aggregation of the search decisions over the unemployed population, it does not solely reflect demand conditions, and estimates of UI benefit effects suffer from proxy variable bias.

Finite Horizon Model

Spells of unemployment tend to be short (weeks or months) in relation to an individual's life span. A finite lifetime would not seem, therefore, to be a reason to explore the implications of a finite horizon search model. On the other hand, it is reasonable to assume that individuals generally will not be able to self-finance extended periods of unemployment and that external borrowing is limited. One can think, then, of the finite (search) horizon as corresponding to the maximal unemployment period that can be financed through internal and external funds. Once a job is accepted, however, it lasts forever; that is, the horizon is infinite subsequent to accepting a job. In addition to the previous notation, denote by T the end of the search horizon. To close the model, it is assumed that if the terminal period is reached without having accepted a job, the individual receives b forever.[14]

Without going into the details, the reservation wage path can be shown to satisfy the following difference equation:

$$\frac{w_t^*}{1-\delta} = b + \frac{\delta}{1-\delta} w_{t+1}^* + \frac{\delta}{1-\delta} q \int_{w_{t+1}^*}^{\infty} (w - w_{t+1}^*) dF(w) \text{ for } t < T, \tag{3.6}$$

$$w_T^* = b. \tag{3.7}$$

Notice that equation 3.6 reduces to the implicit reservation wage equation for the infinite horizon problem if $w_t^* = w_{t+1}^*$. Given a distributional assumption for wage offers, $F(w)$, the solution of the finite horizon reservation wage path can be obtained numerically by starting from period T and working backward.[15]

In the finite horizon case, the reservation wage is decreasing in the duration of the spell, $\frac{dw_t^*}{dt} < 0$, rather than being constant as in the infinite horizon case. In addition, the reservation wage is bounded from below by b (at T), and from above by the infinite horizon reservation wage (w^*). The hazard rate is thus increasing in t, $\frac{dh_t}{dt} = -\lambda f(w^*) \frac{dw_t^*}{dt} > 0$. As a result, the longer the spell duration, the greater the exit rate. The important property of the finite horizon reservation wage is that it depends on the time left until the horizon is reached. The reservation wages are equal under two different horizons not when they have the same amount of time elapsed since beginning the spell of unemployment (t) but when they have the same amount of time left until the horizon is reached ($T - t$).

The finite horizon model can be used to incorporate the fact that UI eligibility is exhausted after a fixed number of weeks (typically 26 weeks). Mortensen (1977) incorporates benefit exhaustion into a more general search model that also allows for endogenous search intensity as well as multiple spells resulting from layoffs. Benefit exhaustion, in a single spell search model, can be easily accommodated by combining the infinite and finite horizon models. The period subsequent to exhaustion can be treated as an infinite horizon search problem with zero UI benefits, and the preexhaustion period as a finite horizon problem with positive benefits (Engberg, 1990). The infinite horizon value function serves as the value function at the time of benefit exhaustion. Specifically, the solution to the model is given by replacing equation 3.7 with the infinite horizon reservation wage obtained from solving equation 3.1 for the case that UI benefits are zero, together with equation 3.6, where T is the period at which benefits are exhausted. The reservation wage declines until T and is constant thereafter.

Meyer (1990) incorporated benefit exhaustion into his quasi-structural specification of the duration hazard. A complication is that the length of eligibility changed for some individuals in the sample during their unemployment spell. Although Meyer loosely appeals to a search model with a fixed eligibility period, the relevant search model would be one that took into account the probability of a change in the eligibility period occurring at any future time.[16] Suppose, for example, that the normal eligibility period is T_1, but may be changed to $T_2 > T_1$ with probability $\alpha(Z)$, where Z are variables that reflect the state of the economy that would trigger a change in the eligibility period.[17] Assume that once the eligibility period is changed for an individual who had already started an unemployment spell, the eligibility horizon cannot revert back to T_1 for that individual. Given the assumption that the agents solve the infinite horizon model on reaching the end of eligibility, the value function is the same at both T_1 and T_2, given by $\frac{w^*}{r}$ where w^* is the solution to equation 3.1.

The reservation wage at any unemployment duration depends on which of the eligibility horizons prevails at the time. Any individual who has been unemployed longer than the current eligibility period has a reservation wage w^*. If the individual currently faces an eligibility horizon of T_2 and has been unemployed less than T_2, then the reservation wage is given by solving equation 3.6 recursively back from T_2 to the current period. If an individual currently has been unemployed less than T_1 and the eligibility horizon is T_1, then, with the value function of search at period t under the horizon T_i denoted as $V_t^{T_i}$, the value of search is

$$V_t^{T_1} = b + \delta\{(1-\omega(Z_t))[qE\max(W(w),V_{t+1}^{T_1}) + (1-q)V_{t+1}^{T_1}] + \omega(Z_t)[qE\max(W(w),V_{t+1}^{T_2}) + (1-q)V_{t+1}^{T_2}]\} \tag{3.8}$$

for $t < T_1$.[18] Rewritten in terms of reservation wages, this becomes

$$\frac{w_t^{*T_1}}{1-\delta} = b + \delta(1-\omega(Z_t))\frac{w_{t+1}^{*T_1}}{1-\delta} + \delta\omega(Z_t)\frac{w_{t+1}^{*T_2}}{1-\delta} + \frac{\delta}{1-\delta}q(1-\omega(Z_t))\int_{w_{t+1}^{*T_1}}^{\infty}(w - w_{t+1}^{*T_1})dF(w) + \frac{\delta}{1-\delta}q\omega(Z_t)\int_{w_{t+1}^{*T_2}}^{\infty}(w - w_{t+1}^{*T_2})dF(w) \tag{3.9}$$

for $t < T_1$. The reservation wage path for any t when the eligibility horizon is T_2 is

$$\frac{w_t^{*T_2}}{1-\delta} = b + \frac{\delta}{1-\delta}w_{t+1}^{*T_2} + \frac{\delta}{1-\delta}q\int_{w_{t+1}^{*T_2}}^{\infty}(w - w_{t+1}^{*T_2})dF(w) \text{ for } t < T_2. \tag{3.10}$$

Notice that the reservation wage under the T_1 horizon (equation 3.9) is bounded between the reservation wage that would apply if $\omega(Z_t) = 0$ for all t, that is, $T = T_1$, and the reservation wage that applies with the horizon $T = T_2$.

The reservation wage can follow two possible paths. If the eligibility period never changes, then the reservation path is declining from $t = 1$ to T_1 (according to equation 3.9) and then is constant from T_1 on. If the eligibility horizon changes from T_1 to T_2 at some $t = t_c < T_1$, then the reservation wage declines until $t_c - 1$ according to equation 3.9 and then jumps up at t_c to the reservation wage $w_{t_c}^{*T_2}$. From that point, the reservation wage declines according to equation 3.10 until T_2 and is then constant from T_2 on. The hazard rate then also can follow two paths.

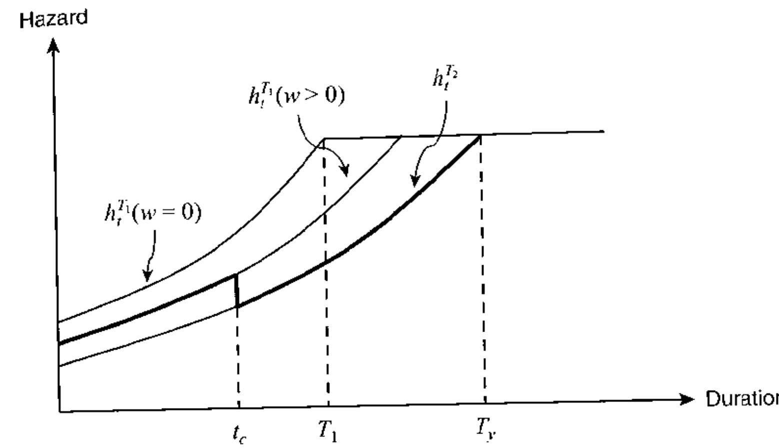

Figure 3.1
Hazard function.

In the first case, where the eligibility horizon does not change, the hazard rate is increasing until $t = T_1$ and then is constant. In the second case, the hazard rate increases until the change in the eligibility horizon, jumps down at the point at which the horizon changes, and then rises until the end of the (new) horizon, after which it is constant. The hazard rate thus has a spike at the point at which the change in the eligibility horizon occurs. Figure 3.1 illustrates the hazard rate when there is an extension of benefits.

The reservation wage at any t is thus a function of the following arguments:

$$w_t^* = w^*((T_1 - t)I(T_t = T_1, t \le T_1), (T_2 - t)I(T_t = T_2, t \le T_2), b, q, \delta, Z_t, F(w)), \tag{3.11}$$

where $I(T_t = T_k, t \le T_k)$ is 1 if the terminal period is T_k and the duration of the unemployment spell is less than T_k and 0 otherwise. The hazard function, $h_t = q(1 - F(w_t^*))$, has the same arguments.

Why is it important to pay attention to theory in analyzing Meyer's data? Theory is informative as to the selection of the arguments in the hazard function, both in terms of what should and what should not be included.[19] As in the case of the standard search model, as seen in equation 3.11, the preunemployment wage is not an argument of the hazard function. It is included by Meyer. Analyzing the consequence of including it as a proxy for an omitted variable can only be done in the context of a theory. Similarly, the above theory provides a direct rationale for and interpretation of aggregate variables (Z_t), such as the state unemployment rate in Meyer's specification, that are known to affect whether benefits are extended. As already discussed, the interpretation of aggregate variables as a demand-side factor is problematic.

Theory also provides guidance as to the form of the specification. As seen in equation 3.11, the reservation wage rate depends on the time left until benefit exhaustion ($T_k - t$). Uncertainty about the eligibility horizon implies that the reservation wage is not the same for horizon T_1 and T_2 given the same amount of time left before benefit exhaustion; the possibility of an extension exists when the horizon is T_1 but not when it is T_2.[20] The reservation wages are the same if changes in the horizon are completely unanticipated ($\omega = 0$), which is the implicit assumption made by Meyer. Notice that if the change in the horizon occurs at $t = t_c$, then the downward jump in the reservation wage is given by $w^*((T_2 - T_c)I(T_t = T_2, t_c \le T_1), 0, \cdot) - w^*(0, (T_2 - t_c)I(T_t = T_2), t_c \le T_1 < T_2), \cdot)$. In general, the downward jump will be greater the closer the period in which the change occurs is to the original horizon, that is, the closer that t_c is to T_1. Thus, the specification of the hazard function should capture this nonlinearity if the empirical model is to have a chance at generating the larger spikes observed in the data around typical values of T_1.[21]

It would be possible to structurally implement the search model described above. The structure would need to be modified in order to capture the fact that the hazard rate is U-shaped, for example, by adding unobserved heterogeneity and/or period effects in some of the structural parameters. One of Meyer's goals was to see whether the spikes in the hazard could be explained by covariates. Structural estimation of the search model parameters explicitly captures the inherent nonlinearities and spikes in the hazard function. Whether they would match the data is an open question.

Identification in the Standard Search Model

Identification is an important issue in structural empirical work. Constructive identification results are available in the search context. To simplify, I consider identification of the standard search model parameters assuming data on durations of unemployment and on accepted wages come from a homogeneous population.[22] To establish identification, it is useful to rewrite the reservation wage implicit equation (equation 3.1) as

$$\begin{aligned} w^* &= b + \frac{q}{r}\int_{w^*}^{\infty}(w - w^*)dF(x) \\ &= b + \frac{q}{r}[E(w \mid w \ge w^*)(1 - F(w^*)) - w^*(1 - F(w^*)) \\ &= b + \frac{h}{r}[E(w \mid w \ge w^*) - w^*]. \end{aligned} \tag{3.12}$$

From the accepted wage data, note that a consistent estimator of the reservation wage is the lowest observed wage: $\lim_{n\to\infty}\Pr(\mid w_{\min} - w^* \mid) = 0$.[23] Then, given an estimate of h from the duration data, and recognizing that from the accepted wage data, we can also obtain an estimate of $E(w \mid w \ge w^*)$, b can be identified if we take r as given. This result does not require a distributional assumption for wage offers.

One cannot, however, separate q and F without a distributional assumption. Although $F(w \mid w \ge w^*)$ is known, given an estimate of w^*, one can recover the wage offer distribution, $F(w)$, only if it is possible to recover the untruncated distribution from the truncated distribution. Obviously, that cannot be done without making a distributional assumption. Assuming that $F(w)$ is recoverable from the accepted wage distribution, then from $h = q(1 - F(w^*))$, one can recover the offer probability, q. Recoverability of the wage offer distribution is not always possible (Flinn and Heckman, 1982). Fortunately, most of the commonly used distributions for wage offer functions, for example, the lognormal distribution, are recoverable from the distribution of accepted wages. However, there is an important lesson to draw, namely that parametric assumptions do not always assure identification.

The analysis of identification in the finite horizon case is similar. The reservation wage at each period can be consistently estimated from the period-specific minimum observed wages. By analogy to equation 3.12, we can write the implicit reservation wage equation as

$$w_t^* = b(1-\delta) + \delta w_{t+1}^* + \delta h_{t+1}[E(w \mid w \ge w_{t+1}^*) - w_{t+1}^*].$$ [24]

This equation must hold exactly at each time t. As long as there are durations of unemployment that extend through more than two periods, that is, given $w_t^*, w_{t+1}^*, w_{t+2}^*, h_{t+1}, h_{t+2}, E(w \mid w \ge w_{t+1}^*)$ and $E(w \mid w \ge w_{t+2}^*)$, δ (or r) and b can be separately identified by solving the two difference equations for the two unknowns, δ and b. Recall that this separation was impossible in the infinite horizon case. Moreover, if we have more than three periods of data, the model is rejectable.

Identifiability of q and F, however, still requires recoverability of the wage offer distribution.

Likelihood Function

Consider the likelihood contribution of an individual solving an infinite horizon search model who has a completed unemployment spell of length d_i and accepted wage w_i^a:

$$L_i = \left[q\Pr(w < w^*) + (1-q)\right]^{d_i-1} \times q\Pr(w_i^a > w^*, w_i^a)$$
$$= \left[q\Pr(w < w^*) + (1-q)\right]^{d_i-1} \times q\Pr(w_i^a > w^* \mid w_i^a) f(w_i^a). \quad (3.13)$$

Recall that the reservation wage is a function of the model parameters b, δ, q, and $F(w)$. The first (bracketed) term in the likelihood is the probability that in each of the periods up to $d_i - 1$ the individual received an offer and rejected it or did not receive an offer. The second term is the probability that the individual received an offer of w_i^a in period d_i and accepted it. Individuals who have incomplete unemployment spells would contribute only the first term to the likelihood.

Observe that for any given value of the reservation wage (or the parameters that determine it), whether or not an individual works conditional on a wage draw is deterministic, that is, its probability, $\Pr(w_i^a > w^* \mid w_i^a)$ in equation 3.13, is either 1 or 0. In order that the likelihood not be degenerate, the reservation wage must be less than the lowest accepted wage in the sample; for this reason, the minimum observed wage in the sample is the maximum likelihood estimate of the reservation wage. That would not create an issue if observed wages were all reasonable. However, in most survey data sets the lowest reported wage is often quite small, less than \$1 or even only a few cents. Such outliers would potentially have an extreme effect on the estimates of the structural parameters. One remedy would be to trim the wage data, say by whatever percentage led to a "reasonable" lowest wage. However, that would be essentially choosing the reservation wage by fiat. A second alternative would be to add another error to the model, for example, by allowing the cost of the search to be stochastic, in which case the reservation wage would be stochastic.[25] A very low accepted wage would be consistent with the individual having drawn a very high search cost.

Of course, adding another source of error does not deal with what is the likely root cause of the problem, which is that wages are not accurately reported.[26] That "fact" has led researchers to directly allow

for measurement error in the reported wage. Introducing measurement error not only accounts for a real feature of the data, it is also convenient in that it does not require any change in the solution of the search problem. The reservation wage is itself unaffected by the existence of measurement error. If we take into account the existence of measurement error and let w_i^{aR} be the reported accepted wage and w_i^{aT} the true accepted wage, the likelihood function becomes

$$L_i = \left[q\Pr(w_i^{aT} < w^*) + (1-q)\right]^{d_i-1} \times q \int_{w_i^{aT}} \Pr(w_i^{aT} > w^*, w_i^{aR})$$
$$= \left[q\Pr(w_i^{aT} < w^*) + (1-q)\right]^{d_i-1}$$
$$\times q \int_{w_i^{aT}} \Pr(w_i^{aT} > w^* \mid w_i^{aR}, w_i^{aT}) g(w_i^{aR} \mid w_i^{aT}) dF(w)$$
$$= \left[q\Pr(w_i^{aT} < w^*) + (1-q)\right]^{d_i-1} \times q \int_{w_i^{aT}} \Pr(w_i^{aT} > w^* \mid w_i^{aT}) g(w_i^{aR} \mid w_i^{aT}) dF(w),$$

where $g(w_i^{aR} \mid w_i^{aT})$ is the distribution of the measurement error and where the third equality emphasizes the fact that it is the true wage only and not the reported wage that affects the acceptance probability. The most common assumption in the literature is that the measurement error is multiplicative, that is, proportional to the true wage.

The estimation of the finite horizon search problem when there are extreme low-wage outliers is particularly problematic. Recall that the reservation wage is declining with duration. Thus, if an outlier observation occurs at an early duration, the entire subsequent path of reservation wages must lie below the reservation wage at that early duration. The incorporation of measurement error is, therefore, critical for estimation. The analogous likelihood contribution for an individual for the finite horizon model, which takes into account that the reservation wage is duration dependent, is

$$L_i = \prod_{j-1}^{d_i-1} \left[q\Pr(w_j^{aT} < w_j^*) + (1-q_j)\right] \times q\Pr(w_{d_i}^{aT} > w_d^*, w_d^{aR}).$$

The estimation of the partial equilibrium search model involves an iterative process as in any DCDP model, namely numerically solving a dynamic programming problem at trial parameters and maximizing the likelihood function. The generality of the DCDP approach has allowed researchers considerable flexibility in modeling choices. Thus, researchers have adopted different distributional assumptions and have extended the standard search model in a number of directions, some of which have already been mentioned.

As has been generally true in the DCDP literature, the theoretical models that serve as their foundation cannot directly be taken to the data. In the case of the standard search model, neither the infinite horizon nor the finite horizon model can fit the generally observed fact that the hazard rate out of unemployment declines with duration. Recall that the hazard rate is constant in the infinite horizon case and increasing in the finite horizon case. There are several ways to deal with this mismatch between the data and the models. In the infinite horizon model, introducing unobserved heterogeneity in model fundamentals, such as the cost of search, the offer probability, and/or the wage offer distribution, can produce negative duration dependence in the population hazard while maintaining constancy at the individual level. In the finite horizon case, allowing for time dependencies in model fundamentals, such as allowing offer probabilities to decline with duration, can create negative duration dependence, in this case at the individual level as well as at the population level.

A DCDP Example

One great advantage of the structural approach is that it enables the researcher to incorporate into estimation an exact representation of program rules. This advantage is particularly valuable in policy evaluation. It is reasonable to suppose that the closer a model mimics existing program rules, the better the model will fit the actual behavior, and the more accurately the model will be able to evaluate prospective policy changes.

3.1.1.3 A Structural Application

Ferrall (1997) structurally estimates a DCDP model of job search, concentrating on the transition from school to work, that integrates all of the major features of the Canadian UI system. In Canada, as in the United States, the period of search for one's first job after leaving school is usually not covered by the UI system. Although that spell of search unemployment is not insured, there is still the potential for the UI system, given its structure, to affect search behavior. To understand why, consider the structure of the UI system in Canada that prevailed during the time period studied by Ferrall. In that system:

1. An unemployed worker who is eligible for insurance must wait 2 weeks after becoming unemployed before collecting benefits.[27]

2. The benefit level depends on the previous wage through a fixed replacement rate (0.60 at the time). The insurable weekly wage is bounded from below by $106 and from above by $530. Thus benefits are $0 if the wage on the previously held job was less that $106, 0.6 times the wage if the wage is between the bounds, and $318 for wages at or above $530.

3. To be eligible for UI benefits, a person must have worked on insurable jobs a certain number of weeks in the 52 weeks prior to becoming unemployed. The number of weeks depends on the regional unemployment rate and the person's previous employment and UI history.

4. The number of weeks of benefits depends on the number of weeks worked on the previous job and on whether the individual is eligible for extended benefits but is capped at 52.

There are two elements of the UI system that would affect search during an uninsured spell. First, because benefits are paid during insured spells, there is an incentive for individuals in an uninsured spell to become employed to be eligible for benefits during future unemployment spells. Thus, the UI system reduces the reservation wage in an uninsured spell; further, the reservation wage will be lower the higher are the benefits (Mortensen, 1977). On the other hand, because the level of benefits increases with the wage on the prior job, there will be an incentive for someone in an uninsured spell to wait for a higher wage offer and thus to have a higher reservation wage. This incentive, however, only applies to individuals whose reservation wage would otherwise be below the maximum insurable wage. Moreover, the magnitude of these incentive effects depends on the probability that an individual will be laid off from future jobs.

The model estimated by Ferrall, aside from the explicit incorporation of UI rules, differs from the standard single-spell search model in a number of ways. The model allows for a search period during school, an initial uninsured spell after leaving school, the first job spell, and a subsequent insured unemployment spell if a layoff occurs. The individual maximizes the expected present value of the log of consumption, where consumption equals the wage while working and the sum of UI benefits net of the cost of search plus the value of home production. In each period of unemployment, the individual receives a job offer with some positive probability. However, a job offer comes not only with a wage offer but also with a layoff rate. The wage offer function is Pareto, and the layoff rate can take on a fixed number of values, randomly drawn.[28] Individuals differ, according to their unobserved type, in their market skill level and in their value of home production.

The solution method is by backward recursion, where the value function for the infinite horizon search problem when UI benefits are exhausted after a layoff occurs serves as the terminal value function for the insured unemployment spell at the time of benefit exhaustion. The model is solved backward from there as a finite horizon problem until the beginning of the search period while in school. The estimation is by maximum likelihood. Ferrall provides evidence of model fit.

Ferrall performs a number of counterfactual experiments that vary the parameters of the UI system. The most extreme is the elimination of the UI system. Recall from the earlier discussion that there was no unambiguous prediction of how reservation wages of ineligibles would be affected by such an experiment. Ferrall finds that for those with at most a high school education residing outside of the Atlantic region, reservation wages rise; the expected duration of unemployment after leaving school (including those who have no unemployment spell) is estimated to increase by about 50 percent. Similarly, for those with some college residing outside of the Atlantic region, the increase is 40 percent. However, there is almost no effect on the expected duration for those residing in the Atlantic region regardless of education.

3.1.2 The Effect of Public Welfare on Female Labor Market and Demographic Outcomes

3.1.2.1 *Background*

The Aid for Dependent Children (AFDC) program has been among the most widely studied programs in the social sciences. Introduced in 1935, and since replaced in 1996 by the Temporary Assistance for Needy Families (TANF), AFDC provided means-tested welfare benefits to (primarily) single women with children.[29] As a by-product of providing income assistance, it has been argued that the AFDC program created a dependent underclass characterized by low levels of schooling, high fertility, low marriage rates, and low employment rates (Murray, 1984). Evidence on the effect of AFDC on these demographic and labor market outcomes has been reviewed by Moffitt (1992, 1998) and Hoynes (1997). They report two major empirical findings from the literature: (1) AFDC reduces labor supply by 10 to 50 percent, and (2) AFDC had a negative

effect on marriage and a positive effect on fertility, but there is no consensus on the magnitude.

Much of the modern empirical literature on estimating welfare effects is based on microlevel longitudinal data. Given that program rules for AFDC differ by state and vary over time, both between- and within-state variation can be exploited in estimation. The existence of these two sources of variation has led to a conundrum in the regression-based literature—should state-fixed effects be included, thus exploiting only within-state variation, or should they be omitted? The rationale for including state-fixed effects is to be able to better account for unobservable individual preferences that might differ across states as well as other omitted state-level programs, whose existence and generosity might influence behavior and be related to AFDC program rules. It was often found that regression estimates of welfare effects on demographic outcomes became smaller and statistically insignificant when state-fixed effects were included in the regression. This was interpreted as correcting for the bias from omitted variables (see Hoynes, 1997, for example).

The notion that the two sources of variation yield theoretically different effects was alluded to in a footnote by Moffitt (1997), but it was not seriously considered within an estimation strategy until Rosenzweig (1999) did so. Keane and Wolpin (2002), in a simulation exercise, provided an explicit demonstration of the mechanisms through which between- and within-state variation in welfare benefits can differentially affect behavioral outcomes. The simple insight is that decisions about behaviors that have long-term benefits and/or costs will be less affected by transitory changes in welfare generosity than by permanent changes. For example, a woman will be less likely to be induced to have a child when a change in welfare benefits is perceived to be of a short duration than if perceived to be permanent. Cross-state variation consists of both a permanent and a transitory component. Controlling for permanent differences using state-fixed effects leaves only within-state transitory fluctuations in benefits.

Figure 3.2 presents the time-series plot of the benefit amounts between 1971 and 1990 available to women who have two children and zero earnings in two states, New York and Texas. As seen, both the levels and patterns differ. Although benefits declined in both states over the entire period, the trend was not monotonic in either. Benefits in New York rose from about $850 per month in 1971 to $1,050 in 1976, then fell to $640 and remained constant.[30] In Texas, benefits were less

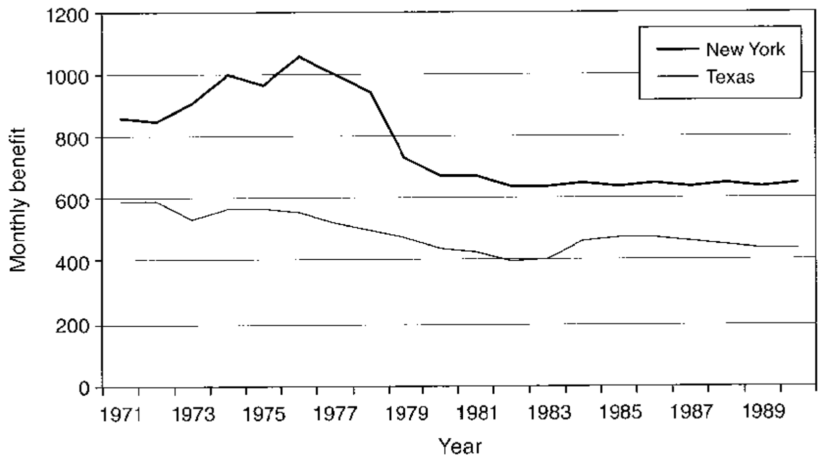

Figure 3.2
Welfare benefit in New York and Texas.

volatile, remaining roughly constant from 1971 to 1976 at around \$550, then falling to about \$400 in 1982, and then rising and remaining constant at about \$450 from 1985.

3.1.2.2 *The Regression-Based Approach*

Regression specifications have not distinguished between these two types of effects, conventionally taking the form (Hoynes, 1997)

$$y_{its} = \alpha B_{its} + \beta X_{its} + u_{its}, \tag{3.14}$$

where y_{its} is the decision variable for an individual i at time t residing in state s, B_{its} is the level of welfare benefits, X_{its} are other determinants of the decision, and u_{its} are unobserved factors determining the decision. Including state dummies in X implies that only within-state variation in B is used to identify α. A linear regression equation that distinguishes these two effects would include both permanent and transitory components of benefits, B^P_{is} and B^T_{its}, where $B_{its} = B^P_{is} + B^T_{its}$:

$$y_{its} = \alpha_1 B^P_{its} + \alpha_2 B^T_{its} + \beta X_{its} + u_{its}. \tag{3.15}$$

The implicit assumption, given equation 3.14, is that the same parameter governs the effect of both permanent and transitory changes, that is, $\alpha_1 = \alpha_2 = \alpha$. If state-fixed effects are included in equation 3.14, then α_2, the effect of the transitory change, is recovered. Researchers have interpreted the estimate of the transitory effect, α_2, as if it were the permanent effect, α_1. McKinnish (2008) interprets the lower estimate using state-fixed effects as evidence of measurement error and suggests using an IV approach as a remedy. Of course, if what is being estimated is α_2, in the presence of measurement error, IV recovers a consistent estimate only of α_2, not α_1. If state dummies are not included in equation 3.14 and the transitory component of benefits is omitted (even if there are no other omitted variables), then, assuming that $\alpha_1 > \alpha_2$, a downward-biased estimate of the permanent effect, α_1, is estimated.[31]

Rosenzweig's approach is to calculate for different cohorts of women the average benefit in their state of residence over a period of 9 years, from ages 14 to 22. Variation in this average benefit amount arises because different cohorts of women *within* a state span different years over those ages. Thus, unlike the approach described above, this variation in benefits persists in the presence of state-fixed effects. Rosenzweig estimates the impact of that variation in benefits on early nonmarital childbearing and finds larger effects than the previous literature. The idea is that this measure of benefits reflects differences (across cohorts) that are more permanent than reflected in year-to-year variation.

Keane and Wolpin (2002) illustrate the dependence on the benefit stochastic process of the impact of benefits based on within-state variation by specifying and simulating a DCDP model of fertility (contraceptive use) and welfare take-up. To outline the DCDP model, in each period a woman decides whether to contracept and whether or not to receive welfare (if eligible). A woman who chooses to contracept avoids a pregnancy that period with certainty. A woman who chooses not to contracept has a birth one period later with a known probability that depends on the woman's age. The woman receives a utility flow that depends on a composite consumption good, the number of children ever born, on whether or not she contracepts, and on whether she receives welfare (reflecting a potential stigma effect). The "marginal" utility of contraception and welfare receipt are subject to *iid* stochastic shocks. Consumption in any period depends on that period's earnings, which are stochastic and depend on the woman's age and number of children and on the amount of welfare benefits received that period. Current and past welfare benefits are known, but future benefits are not. The woman maximizes the expected present discounted value of lifetime utility.

Keane and Wolpin (2002) specify a more complete picture of the benefit schedule than is adopted in the literature, which relies on a summary statistic such as the level of benefits given to women who have two children and no earned income. Specifically, the welfare benefit available to a woman residing in state s who is age a at calendar time t is approximated by a rule that depends on the number of children residing in her household (n) and on her earnings (y) according to:

$$B_{at}^{s} = b_{0t}^{s} + b_{1t}^{s} n_{at} - b_{2t}^{s} n_{at} - b_{3t}^{s} y_{at} \text{ if } B_{at}^{s} > 0 \text{ and } N_{at} > 0 \tag{3.16}$$

$$= 0 \text{ otherwise,}$$

where $b_{0t}^{s}, b_{1t}^{s}, b_{2t}^{s}$ are the "guarantee" parameters and b_{3t}^{s} is the benefit reduction (or tax) rate. An increase in b_{0t}^{s} augments benefits independent of the number of children, whereas increases in b_{1t}^{s} (or a reduction in b_{2t}^{s}) augment the benefit with respect to additional children. B_{at}^{s} is denoted as the benefit rule and the b's as the benefit rule parameters. Benefit rule parameters are estimated in each year for each of the two states. The mean value of the constant term in equation 3.16 is about twice as large in New York as in Texas, \$454 versus \$228. In addition, New York has a larger payout for each additional child that increases essentially linearly with each extra child ($\bar{b}_1 = 161$, $\bar{b}_2 = 0.6$), whereas the payoff in Texas for additional children increases more slowly and at a diminishing rate ($\bar{b}_1 = 134$, $\bar{b}_2 = 4.9$).

Changes over time in the level of benefits arise from changes in the underlying benefit rule parameters. Benefit rule parameters evolve according to a general vector autoregression (VAR),

$$\mathbf{b}_t^{s} = \boldsymbol{\lambda}^{s} + \boldsymbol{\Lambda}^{s}\mathbf{b}_{t-1}^{s} + \mathbf{u}_t^{s}, \tag{3.17}$$

where $\mathbf{b}_t^{s}$ is a 4×1 column vector of the benefit rule parameters, $\boldsymbol{\lambda}^{s}$ is a 4×1 column vector of regression constants, $\boldsymbol{\Lambda}^{s}$ is a 4×4 matrix of autoregressive parameters, and $\mathbf{u}_t^{s}$ is a 4×1 column vector of *iid* mean zero shocks drawn from a stationary distribution with variance-covariance matrix Ω^{s}. The evolution of the benefit rule parameters also differs across the two states. The differences are highlighted in table 3.1, which shows the impulse response functions for a change in b_0 of \$83.33 per month (\$1,000 annualized, the period length of the decision model) and in b_1 of \$41.67 (\$500 annualized). Women are assumed to use equation 3.17 to update their beliefs about future benefits when deciding on contraception and welfare take-up in any period.

In New York, a once and not for all increase in b_0 of \$1,000 annualized in any period is followed immediately in the next period by a decline in b_0 of \$16 and further declines that peak at \$163 six (and

Table 3.1
Impulse Responses to Innovations in the Overall and Per-Child Guarantees

	New York		Texas	
Period	b_0	b_1	b_0	b_1
1	1,000	500	100	500
2	−16	498	1,148	274
3	−127	521	1,142	139
4	−147	559	1,068	56
5	−155	573	955	4
6	−163	569	841	−24
7	−163	547	727	−36
8	−157	517	625	−44
9	−155	474	528	−40
10	−143	429	443	−8
.				
.				
.				
25	11	−49	33	−4

Source: Keane and Wolpin (2002).

seven) periods later until the effect of the initial increase slowly dissipates. The same increase in Texas is followed by additional increases, peaking at \$1,148 in the next period and slowly declining. The increase is still \$443 in period 10 and falls to \$33 25 periods after the initial innovation. The effect of an innovation of \$500 (annual) in b_1 is exactly reversed. In New York, the innovation is followed by slightly higher values in the next seven periods (with the exception of period 2), whereas in Texas, the initial innovation is followed by an increase by \$274 in period 2, \$139 in period 3, and then small positive and negative fluctuations. In addition to having different impulse response functions, the two states differ in the contemporaneous correlation of the innovations to b_0 and b_1. The contemporaneous correlation of the innovations between b_0 and b_1 is positive with a value of 0.5 in New York but is negative and much smaller in Texas, with a value of −0.19.

To best illustrate the different impacts of permanent and transitory changes in welfare benefits, Keane and Wolpin (2002) parameterize the model to deliver large welfare take-up rates. To isolate the impact of the structure of welfare benefit rules, the other parameters of the model

Table 3.2
Effect of Permanent and Transitory Changes in Guarantee Parameters on Number of Pregnancies

	New York		Texas	
	Total Pregnancies by Period 20	Pregnancies in Period 1	Total Pregnancies by Period 20	Pregnancies in Period 1
Permanent				
$\Delta b_0 = 1{,}000$	0.78	-	0.76	-
$\Delta b_1 = 500$	1.37	-	0.39	-
Transitory				
$\Delta b_0 = 1{,}000$	−0.14	−0.03	0.47	0.05
$\Delta b_1 = 500$	1.20	0.25	−0.06	−0.005

(for example, preferences) are assumed to be the same in both states. The decision horizon is 20 periods. Table 3.2 shows the impact of a permanent $1,000 increase in b_0 on the mean number of pregnancies over the decision horizon in the two states. This effect corresponds to α_1 in equation 3.15 or how researchers were aiming to interpret α in equation 3.14. As seen, the effect of the increase in b_0 is to increase pregnancies similarly in both states, by 0.78 in New York and 0.76 in Texas.

A permanent change in b_1 of $1,000, also shown in table 3.2, is equivalent to a $500 change in b_1 for a mother with two children, which, as noted, is often the summary measure of welfare generosity used in the literature. Unlike the effect of increasing b_0, the increase in the per-child increment differs substantially between the two states, increasing pregnancies by 1.37 in New York and by 0.39 in Texas. Thus, even if the permanent effect could be estimated as was the intention, the effect of welfare benefits on fertility that would be estimated using the summary statistic would depend on the composition of the benefit variation between b_0 and b_1.

Table 3.2 also shows the responsiveness of fertility to transitory changes in b_0 (of $1,000) and b_1 (of $500). Two cases are considered, one in which the change occurs in only one period (the first) and one in which the change occurs in each period of the 20-year horizon, an extremely unlikely event given the *iid* nature of the innovations. Keane and Wolpin (2002) consider the latter case to determine the extent to which a change in the actual mean level of benefits generated from transitory changes would mimic the effect of the permanent change already discussed. This is one interpretation of the effect estimated by Rosenzweig (1999), albeit an extreme version. Alternatively, if women had perfect foresight (or if they believed, incorrectly, that all innovations were permanent), these changes would be viewed as permanent, and Rosenzweig's estimate would correspond to the permanent effect, subject to the caveat associated with using the summary measure of benefits.

The effect of a transitory change in b_0 in the first decision period is essentially zero in both states, trivially reducing pregnancies in New York and increasing pregnancies in Texas. Repeated innovations of the same amount in each of the twenty periods induces larger effects, a fall in pregnancies of 0.14 in New York and an increase of 0.47 in Texas. Recall that the estimate of the permanent effect was to increase mean pregnancies by about three-fourths in both states. The results, as is consistent with the impulse responses, are reversed for a transitory increase in the per-child benefit. The effect of that increase is negligible in Texas regardless of whether the increase occurs only in the first period or in all twenty periods. In New York, the innovation in only the first period increases pregnancies by 0.25, whereas increases in all twenty periods increase pregnancies by 1.20. Relative to the permanent increase, the persistent innovation is close for New York but not for Texas. Thus, if women have imperfect foresight and use updating rules consistent with the actual evolution of benefits over time, the effect on fertility of once and not for all transitory increases in benefits is considerably smaller than the effect of a permanent change in benefits. Innovations lack the persistence necessary to mimic permanent effects. Thus, in this case, state-fixed effect estimates that estimate transitory effects will differ from permanent effects that use cross-state variation in benefit rules.

Keane and Wolpin (2002) demonstrate that the results from the regression-based literature carry over to the data simulated from the model. In particular, in a logit specification of the contraceptive (fertility) decision, the benefit effect (based on the summary measure) is reduced by two-thirds with the inclusion of state-fixed effects. Further, if the benefit is disaggregated into the overall guarantee parameter (b_0) and the per-child guarantee parameter (b_1), the per-child parameter is twice as large, and eight times as large with state-fixed effects. Thus, the prediction of how increasing the benefit to a woman with two children would affect the number of pregnancies depends on how the increase is structured; there would be a many times larger impact if the increase came from a change in the per-child guarantee than from the overall guarantee.

There are a number of lessons to be drawn from these simulations for interpreting the regression-based literature. First, estimates based on cross-section variation confound the impact of permanent and transitory changes in welfare benefits. Second, estimates based on year-to-year fluctuations in benefits within states, which estimate the effect of transitory changes in benefits, may differ greatly, even in sign, from the effect of permanent changes. Third, in a dynamic setting, it matters how expectations are formed about future benefits. Fourth, estimates based on summary measures of welfare generosity confound the effect of different components of welfare rules, for example, the effect of changes in the overall level of benefits and changes in the benefits accruing to having additional children. Fifth, because states have different benefit rules, and they evolve differently over time, the effect of changes in welfare generosity is heterogeneous across states and across cohorts. Thus, the effect that is estimated depends on the composition of the sample in those dimensions. These issues can all be dealt with within a structural estimation approach, to which I now turn.

3.1.2.3 *Structural Approach*

There are a number of papers that have adopted the DCDP paradigm in estimating the impact of the AFDC program. Sanders (1993) and Miller and Sanders (1997) consider work and welfare participation, and Swann (2005) also includes marriage and welfare participation decisions. Fang and Silverman (2009) also consider work and welfare participation but allow for time-inconsistent agents. Keane and Wolpin (2007, 2010) allow for choices over work, welfare participation, marriage, fertility, and schooling.

As was the theme of chapter 2, consider the possibility of performing an *ex ante* policy evaluation of the AFDC program. Is it possible to do so nonparametrically as was the case in the *ex ante* evaluation of the wage tax or school attendance subsidy program? Ignore some of the complexity of the actual AFDC program and assume that there is a welfare benefit schedule, $b(n, y)$, offered to unmarried women with children, that depends on the number of children, n, and on the woman's non-earned income, y. In that case, assuming no saving or borrowing, the woman's budget constraint is

$$C = y + w(1-l) + b(n,y)l, \tag{3.18}$$

where C is consumption, $l = \{0, 1\}$ indicates nonemployment, and w is the wage offer. In the absence of a welfare program, $b = 0$. Assume that program eligibility depends only on unearned income not exceeding some limit. Notice that equation 3.18 can be written as

$$C = (y + b(n,y)) + (w - b(n,y))(1-l)$$

$$= \tilde{y} + \tilde{w}(1-l),$$

which takes the same form as the budget constraint without welfare, $C = y + w(1 - l)$.

The welfare program is identical in terms of the budget constraint to the school attendance subsidy program analyzed previously with multiple children, where the subsidy was in that case $\tau = bn$. As in that case, variation in wage offers (for female as opposed to child labor) and in non-earned income can be sufficient to identify the impact of introducing a (new) welfare program on labor supply; benefit variation is unnecessary. The same limitations of the nonparametric approach apply; for example, there can be no direct utility effect on welfare participation, and fertility cannot be endogenous.

The discussion of the parametric approach to *ex ante* evaluation of the school subsidy program also directly applies. Keane and Wolpin (2007, 2010) provide the most complete behavioral model. Let me touch on some of the major features (ignoring some of the details of the actual specification).

1. Welfare rules: The modeling of the welfare benefit rule, although still an approximation to the actual rules, accounts for more complexities than are seen in equation 3.16. In particular, the benefit rule is a five-parameter function that varies by state (s) and calendar time (t) given by

$$B_t^s(n_{at}^{18}, y_{at}^o, y_{at}^z) = b_{0t}^s + b_{1t}^s n_{at}^{18} - b_{3t}^s \beta y_{at}^z z_{at} \quad \text{for } y_{at}^o < y_{at}^{s1}(n_{at}^{18}) \tag{3.19}$$

$$= b_{2t}^s + b_{4t}^s n_{at}^{18} - b_{3t}^s[(y_{at}^o - y_{at}^{s1}) + \beta y_{at}^z z_{at}] \quad \text{for } y_{at}^{s1}(n_{at}^{18}) \le y_{at}^o < y_{at}^{s2}(n_{at}^{18})$$

$$= 0 \text{ otherwise.}$$

In the first line segment in equation 3.19, the grant is seen to increase linearly in the number of minor children n_{at}^{18}, and if a woman coresides with parents, $z_{at} = 1$, a part, $\beta < 1$, of parents' income, y_{at}^z, is taxed at rate b_{3t}^s. In general, benefits are taxed away if the woman has positive earnings, y_{at}^o. However, due to work expense deductions and child care allowances, the tax is not assessed until earnings exceed a (state- and time-specific) "disregard" level, denoted by $y_{at}^{s1}(n_{at}^{18})$. The amount of benefits, once earnings exceed this level, is given by the second line segment in equation 3.19. The benefit tax rate or "benefit reduction" rate is taken to be the same as the tax on parents' earnings, b_{3t}^{s}. Finally, $y_{at}^{s2}(n_{at}^{18})$ is the level of earnings at which all benefits are taxed away. Keane and Wolpin (2007, 2010) assume that the stochastic process generating the evolution of the benefit rule parameters within each state follows a VAR as in Keane and Wolpin (2002).

2. Choice set: A woman makes joint decisions at each age of her life about the following set of discrete alternatives: whether or not to attend school, work part-time, or work full-time in the labor market (if an offer is received), be married (if an offer is received), become pregnant (if she is of a fecund age), and receive government welfare (if she is eligible). Thus, a woman chooses from as many as 36 mutually exclusive alternatives at each age during her fecund life cycle stage and 18 during her infecund stage. The fecund stage is assumed to begin at age 14 and to end at age 45; the decision period extends to age 62. Decisions are made at discrete 6-month intervals up to age 45, that is, semiannually, and then annually up to age 62. A woman who becomes pregnant in any period has a birth in the subsequent period. Consumption is determined by the alternative chosen and the woman's state variables.

3. Preferences: The woman receives a utility flow at each age that depends on her consumption and her five choices: (1) work, (2) school, (3) marriage, (4) pregnancy, and (5) welfare participation. Utility also depends on past choices (as there is state dependence in preferences), for example, on the number of children already born, their current ages (which affect child-rearing time costs), and their current level of completed schooling (which affects the utility from attending school). Marriage and children shift the marginal utility of consumption. Preferences are assumed to evolve with age and to differ among individuals by birth cohort, race, and state of residence.

There is also a vector of five permanent unobservables, determined by a woman's latent "type," that shift her tastes for leisure, school, marriage, pregnancy, and welfare participation. In addition, there are age-varying preference shocks to the disutility of nonleisure time (i.e., the sum of time spent working, attending school, child rearing, or collecting welfare), as well as direct utilities or disutilities from school, pregnancy, and welfare participation (unrelated to the time cost) and a fixed cost of marriage.

4. Budget constraint: The budget constraint is assumed to be satisfied each period. Resources available to the woman depend on her marriage and parental coresidence status. In addition to the welfare benefits for which a woman is eligible (as in equation 3.19), a woman who (1) is neither married nor living with her parents receives her labor earnings, (2) is married (and by assumption not also residing with parents) receives a share of the combined household earnings, or (3) is living with her parents receives her own earnings and a share of her parents' income. The shares in the latter two conditions are estimated parameters of the model. A woman who attends college or graduate school pays a tuition cost.

Her husband's earnings depend on his human capital stock. Conditional on receiving a marriage offer, the husband's human capital is drawn from a distribution that depends on the woman's characteristics, some of which, like her schooling, depend on the woman's choices. In addition, there is an *iid* random component to the draw of the husband's human capital that reflects a permanent characteristic of the husband unknown to the woman prior to meeting (which induces marriage market search on the part of the woman). Husband's earnings evolve with a fixed (quadratic) trend subject to a serially independent random shock. Parents' income depends on education and race.

In each period a woman receives full-time and/or part-time job offers with given probabilities. Each offer rate depends on the woman's previous-period work status. If an offer is received and accepted, the woman's earnings are the product of the offered hourly wage rate and the number of hours worked (either part-time hours or full-time hours). As in the Ben-Porath (1967)-Griliches (1977) formulation, the hourly wage rate is the product of the woman's human capital stock and its per-unit rental price, which may differ between part- and full-time jobs. Her human capital stock is a function of completed schooling , work experience, and her skill "endowment" (at age 14).

5. Residence status: In each period a single woman draws an offer to marry with some probability that depends on a subset of her characteristics. If the woman is currently married, then, with some probability she receives an offer to continue the marriage that depends on the duration of marriage and her age. If she declines to continue the marriage, the woman must be single for one period before receiving a new marriage offer. A single woman lives with her parents at any age according to a draw from an exogenous probability rule.

6. Endowment heterogeneity: Both permanent preferences and skill endowments differ for black, Hispanic, and white women, by state of residence, and by a woman's unobserved type. Note that with state of residence controlled, only within-state variation in welfare benefits contribute to identification. However, identification is not dependent on the particular source of benefit variation, and, as shown above, the existence of benefit variation is not needed for identification in some circumstances. All random shocks to preferences and the human capital stock are mutually serially independent and jointly normal.

A woman maximizes the expected present discounted value of remaining lifetime utility at any age conditional on her state space, which depends on her initial (age-14) preferences and skill endowments, her history of prior choices, and the previous period's benefit rule parameters (equation 3.19). The solution procedure, as has been described previously, uses the approximation method developed by Keane and Wolpin (1994). The estimation method is based on Keane and Wolpin (2001), which uses an unconditional simulation of choices to form the likelihood function.[32] The model is estimated using data from the 1979 cohort of the National Longitudinal Surveys (1979) on women who have resided since age 14 in one of five states: California, Michigan, New York, North Carolina, and Ohio. The goal of the estimation is twofold: (1) as is usual, to assess the impact of alternative policies, and (2) specific to this enquiry, to perform counterfactuals that will shed light on the extent to which the observed large racial disparities in demographic and labor market outcomes is the consequence of differences in marital prospects, in labor market opportunities, or in preferences.

Before discussing those issues, it is interesting to contrast the importance of unobserved heterogeneity in this case to the findings about men in Keane and Wolpin (1997) and Sullivan (2010), which are discussed in a later section. For women as well as men, unobserved heterogeneity plays an important role in explaining differences in behavioral outcomes. In the case of women, six latent types were sufficient for the model to provide a reasonable fit to all of the key features of the data. Estimates of several of the structural parameters suggest that types differ greatly and manifest significantly different behavioral outcomes. For example, comparing the behaviors of the two extreme types, types 1 and 6, black women of type 6 have spent 7 more years on welfare by age 30 than have those of type 1, have worked about 8

fewer years, have about 4½ years less education, and have two more children. Differences in welfare receipt between types are smaller for Hispanic and white women but still substantial (i.e., about 5 and 3 years, respectively), and differences in work experience, schooling, and fertility are about as large as for blacks. Type 6 women are a larger group than type 1 women by 10, 6, and 1 percentage points for black, Hispanic, and white women, respectively. Indeed, type 6 women are the largest group for all races.

Keane and Wolpin (2010) find that unobserved heterogeneity is the most important of the initial conditions in accounting for the variance in behaviors. Unobserved type alone accounts for 65 percent of the variation in completed schooling (by age 30). Whatever the process by which these unmeasured preferences and endowments are formed by age 14, they are critical in determining completed schooling levels. In contrast, race/ethnicity accounts for only 2 percent of the variance, state of residence 4 percent, and parental schooling (which affects both type and parental income) 11 percent. Together, initial conditions account for 70 percent of the variance in completed schooling at age 30, with the other 30 percent due to idiosyncratic shocks. However, welfare participation is more volatile. At age 30, only 33 percent of the variance in the total number of (6-month) periods on welfare is explained by type. At age 40 the comparable figure is 36 percent. Being black, Hispanic, or white explains 7 and 9 percent of variance at those two ages, respectively, and parental schooling explains 6 and 5 percent. Although the percentage of variance explained by the different initial conditions varies across different outcomes, in all cases except for years of marriage, type explains the largest percentage.

In general, initial conditions explain a larger percentage of the variation in human capital outcomes, namely schooling, work experience, and wages, than in demographic outcomes, namely children ever born, marriage duration, and the income of potential husbands. Finally, unobserved type, although by far the most important single factor, explains only 32 percent of the variance in utility for white women, 9 percent for blacks, and 17 percent for Hispanics, considerably smaller than for men. In contrast, both Keane and Wolpin (1997) and Sullivan (2010) considered only labor market outcomes. A closer comparison would therefore be the explained variance in full-time wage offers, which is about 65 percent for women.[33] The key reason that type explains less of the variance of lifetime utility for women than it did for men is that demographic outcomes such as fertility and marriage are apparently governed by inherently more noisy processes than labor market outcomes.

Accounting for Minority-Majority Differences in Outcomes

To what extent do differences in marriage market opportunities, labor market opportunities, and tastes account for minority-majority differences in life-cycle outcomes? To answer this question, Keane and Wolpin (2010) perform four counterfactual experiments that involve altering sets of parameters related to (1) the marriage market, (2) the labor market, (3) welfare stigma, and (4) parental schooling. Each experiment reveals how close the outcomes for blacks and Hispanics would be to those for whites if each category of parameters, taken one at a time, were set equal to those of white women. I concentrate the discussion on blacks and refer the interested reader to the paper for a parallel discussion for Hispanics.

Table 3.3 reports some of the results. The first two columns show the baseline model predictions. Columns labeled (1)–(4) show the effects of (1) equalizing potential husband's income and women's preferences for marriage, (2) equalizing offer wage functions (i.e., eliminating labor market discrimination), (3) equalizing welfare stigma, and (4) equalizing (the distribution of) types.

Adopting the marriage market parameters of white women has a large impact on behavior of black women, for whom both the change in preferences and the improved husband's income distribution increase the incentive to marry. For blacks, marriage rates increase almost to parity with whites (for example, from 28.5 percent to 55.7 percent at ages 26–29.5, compared to 65.4 percent for whites). As marriage increases, welfare participation falls, reducing the gap with white women by over a quarter at ages 22–25.5 and by over a third at ages 26–29.5. The gap between white and black women in the percentage of women with an out-of-wedlock pregnancy falls by about one-third at ages 18–21.5 and by almost one-half at ages 22–25.5. A higher probability of marriage reduces the return to human capital investment by lowering the probability of employment. As a result, black women's employment rates fall along with the rise in their marriage rate, doubling the gap with white women. Also reflecting the forward-looking nature of the model, mean schooling falls by a third of a year. Black-white differences in wage offers can result from either or both different skill endowments (at age 14) and differing skill rental prices due to racial discrimination. Providing black women with the same wage offer

Table 3.3
Accounting for Racial Differences in Choices

	Baseline		Counterfactual			
	White	Black	1	2	3	4
Pct. on Welfare						
Age						
15–17.5	1.3	5.1	5.4	4.1	4.1	4.2
18–21.5	4.7	16.8	15.1	12.5	14.0	14.3
22–25.5	7.1	26.5	20.9	17.9	22.8	23.3
26–29.5	7.1	29.7	21.4	19.6	26.4	26.1
Pct. in School						
Age						
15–17.5	85.3	84.4	80.7	87.7	84.2	85.2
18–21.5	29.8	29.6	25.0	30.6	29.8	33.1
22–25.5	8.3	8.1	6.0	9.0	8.1	9.0
26–29.5	3.4	3.5	2.6	3.7	3.5	3.9
Pct. Working						
Age						
15–17.5	28.3	16.9	15.5	31.0	17.0	16.3
18–21.5	63.8	51.9	42.4	68.5	52.8	53.0
22–25.5	70.3	57.4	44.7	71.2	59.1	61.6
26–29.5	69.8	55.7	42.3	70.2	57.3	60.5
Pct. Married						
Age						
15–17.5	5.0	1.1	3.6	1.7	1.1	1.0
18–21.5	28.2	9.6	24.6	12.7	10.0	8.7
22–25.5	52.3	21.7	45.1	27.3	22.5	20.6
26–29.5	65.4	28.5	55.7	36.5	29.4	27.6
Pct. Out-of-Wedlock Pregnancy						
Age						
15–17.5	1.8	3.0	3.0	2.5	2.9	2.8
18–21.5	3.1	5.9	4.9	5.0	5.8	5.7
22–25.5	2.4	5.8	4.2	4.9	5.7	5.7
26–29.5	1.6	4.9	3.1	4.0	4.7	4.7

1. Black women face the same marriage market as white women.
2. Black women have the same wage offer as white women.
3. Black women have the same welfare stigma as white women.
4. Black women have the same type proportions as white women.
Source: Keane and Wolpin (2010).

function, that is, with the same skill endowment and skill rental price, as whites reduces welfare participation more than does equalizing marriage market parameters, closing 44 percent of the black-white gap at ages 22–25.5, while the employment rate for black women reaches parity with that of whites. Marriage rates increase, and out-of-wedlock pregnancies fall, by about 40 percent among women below age 18 and by 25 percent at ages 22–25.5. The effect on welfare participation of replacing the level of welfare stigma of black women with that of white women is relatively small, as is the effect on other outcomes as well. Thus, differences in preferences for welfare participation appear to play little role in explaining differences in outcomes. Finally, equalizing type proportions has a small effect on choices, not too dissimilar to the effect of equalizing welfare stigma.

The Incentive Effects of Altering Welfare Rules

Keane and Wolpin (2010) perform counterfactual simulations to assess the impact of various hypothetical changes in welfare rules on behavior. Results are reported for the 25 percent of black women who are of type 6, women whose preferences, endowments, and opportunities induce them to choose welfare much more frequently than other types. Table 3.4 highlights some of the results. Baseline outcomes under the welfare rules actually in effect for the sample are shown in column 1. About 65 percent of the type 6 black women receive welfare between the ages of 22 and 29.5, and only about a fifth are working. At ages 26–29.5, the marriage rate for black women is only 21.8 percent, while fertility rates are high; on average, 1.5 children have been born to these women by age 24. The low marriage rates combined with the high fertility rates imply that, between the ages of 18 and 21.5, out-of-wedlock pregnancies account for almost 90 percent of all pregnancies for black women of this type. The fraction of women who are high school dropouts ranges from 40 to 50. Welfare benefits, on average, comprise over 40 percent of the total income of black women between the ages of 26 and 29.5.

The behavioral outcomes that result from eliminating welfare, the most extreme contrast, are shown in column 2. In columns 3 and 4, time limits are imposed, first a strict 5-year limit and then a 3-year limit, after which benefits are reduced by a third. Column 5 presents an experiment in which benefits are reduced by 20 percent, and column 6 introduces a work requirement: after having been on welfare for 6 months, a woman has to work or engage in "work-related activities"

Table 3.4
The Effect of Welfare on Choices—Black Women (Type 6)

	1	2	3	4	5
Pct. on Welfare					
Age					
15–17.5	13.2	0.0	13.0	13.2	8.0
18–21.5	39.6	0.0	36.4	39.9	27.2
22–25.5	61.2	0.0	35.6	60.5	45.2
26–29.5	68.1	0.0	16.5	66.5	55.1
Pct. Working					
Age					
15–17.5	9.5	10.4	9.6	9.6	11.9
18–21.5	26.9	36.5	27.2	26.7	42.6
22–25.5	20.8	42.7	25.9	21.4	54.1
26–29.5	15.1	43.4	31.2	18.1	62.2
Pct. Married					
Age					
15–17.5	0.4	0.5	0.4	0.4	0.4
18–21.5	7.8	10.6	8.0	7.8	9.0
22–25.5	16.5	25.3	19.0	16.8	18.9
26–29.5	21.8	36.6	28.9	22.6	25.0
Pct. Out-of-Wedlock Pregnancy					
Age					
15–17.5	5.0	4.5	5.2	5.2	4.9
18–21.5	8.5	7.6	8.6	8.6	8.3
22–25.5	8.5	7.1	8.2	8.6	8.2
26–29.5	7.3	5.4	6.6	7.3	6.9
Pct. High School Dropouts	45.4	36.2	44.0	45.2	43.7

1. Baseline.
2. No welfare.
3. Five-year time limit.
4. Three-year time limit followed by one-third reduction in benefits.
5. Twenty-five hours/week work requirement after 6 months on welfare.

Source: Keane and Wolpin (2010).

for 25 hours per week to continue to receive benefits.[34] Eliminating welfare has a substantial impact on employment. The percentage of black women working between the ages of 26 and 29.5 nearly triples, from 15.1 to 43.4 percent. Also, school attendance rates increase, leading to a decline in the high school dropout rate from 45 to 36 percent for blacks. Marriage rates increase substantially, from 22 to 37 percent for blacks at ages 26 to 29.5, and fertility falls modestly from

2.40 to 2.24 births by age 30. There is a nonnegligible decline in out-of-wedlock pregnancies. In the four age categories, the prevalence of an out-of-wedlock pregnancy falls by 10.0, 10.6, 16.4, and 26.0 percent for blacks.

A strict 5-year time limit (column 3) reduces welfare participation negligibly at young ages but by a large amount thereafter as the limit becomes binding. By age 26 through 29.5, participation falls from 68 to 17 percent for blacks. Increases in employment follow the same age pattern but are of smaller magnitude as marriage rates also increase with the imposition of the time limit. There is a negligible change in fertility (overall and out-of-wedlock) and in schooling. The weaker (and more realistic) time limit (column 4), where benefits are only partially reduced when the limit is reached, has very small effects relative to the baseline.[35]

Imposing a "work" requirement of 25 hours per week to maintain eligibility for welfare benefits beyond 6 months, which is modeled as increasing the time cost of welfare receipt, increases employment by 47 percentage points for black women at ages 26–29.5 (from 15 to 62 percent) but reduces welfare receipt by only 13 percentage points (from 68 to 55 percent). Thus, most of the women who start working after the imposition of the work requirement remain eligible for welfare. However, welfare comprises, on average, only 25 percent of total income after the work requirement is introduced as opposed to 43 percent before, and earnings increase by a factor of 3.

There are two major conclusions from this analysis. (1) There is no simple answer to the question of what causes black (or Hispanic)-white differences in employment, schooling, and demographic outcomes. Labor market opportunities, that is, labor market discrimination and/or skill endowments in place by age 14, and marriage market opportunities, that is, the earnings potential of prospective husbands, as well as the interaction of these differential constraints with the effects of the welfare system, all provide important parts of the explanation. (2) The AFDC program created strong incentives for a significant minority of women, white, black, and Hispanic, that led to lower school attainment, less employment, lower marriage rates, and more out-of-wedlock births.

Validation

It is clear that the structural parametric approach is a powerful tool for *ex ante* policy evaluation. The question remains, as I have discussed

before, as to the credibility of the results from the counterfactuals. Keane and Wolpin (2007) pursued a range of approaches to model validation. As a first step, the within-sample fit of the DCDP model was examined across a number of dimensions and compared to a group of quasi-structural models, specified as four alternative multinomial logits (MNL) estimated on a subset of the choice data. These quasi-structural models are functions of the same set of state variables that enter into the decision rules derived from the DCDP model and thus are consistent with a myopic random utility model or a flexible approximation to a DCDP model. They are differentiated by whether they include state-fixed effects and by the complexity of the specification of welfare rules, one in which the rules are characterized by one parameter and the other by five parameters). Using a RMSE criterion (the number of parameters are similar), there seemed to be no clear winner in this cross-model competition.

The second step involved an external validation using data from a nonrandom holdout sample, specifically data from the state of Texas, which had a very different welfare policy regime from the five states that were used in estimation.[36] This external validation exercise is in the same spirit as the use of the treatment observations in Todd and Wolpin (2006) except that the holdout sample is not a random sample. Indeed, the Texas sample was chosen because the benefit schedule differed considerably from those of the other states in the same way as the treatment households in PROGRESA differed in that they were offered the school attendance subsidy and the estimation sample (controls) were not.

In terms of the same RMSE criterion, the model without state-fixed effects with the five-parameter welfare benefit rule produced predictions for Texas that were highly inaccurate. Further, the two models with state-fixed effects fit the data from Texas better than either the DCDP model or the remaining logit, although there was no clear ranking of the two better-fitting models. However, as was made clear in the discussion of Keane and Wolpin (2002) above, the state-fixed effect specification identifies the behavioral effects of only transitory changes in welfare benefits. Given the results from the simulations in Keane and Wolpin (2002), which was based on a similar VAR representation of the benefit rule parameters, those specifications should not have provided a better prediction of the effect of Texas's permanently different welfare rules. This result led to a third method of validation, which was to predict the effect of a policy intervention that has no analogue in the historical data, namely to give the five estimation states the same welfare rules as Texas.

Given Texas's considerably less generous benefits, welfare participation should have fallen. However, in the specifications with state-fixed effects, welfare participation actually increased. The only reason that the state-fixed effects specification had performed better in the external validation was that, in that exercise, Texas state dummies had been included in the estimation. Moreover, the one logit model without state-fixed effects that performed about equally well as the DCDP model in the external validation also produced unexpected results, forecasting that the decline in welfare participation would be considerably less than the increase in employment. On the basis of these results, Keane and Wolpin (2007) concluded that, in contrast to the DCDP model, all four MNL models were unreliable for policy prediction.

As a final observation, Keane and Wolpin (2007) conjectured that most economists would have professed a greater a priori faith in the ability of the MNL models to forecast behavior than in the DP model. That is, they would be concerned that, because the many assumptions invoked in setting up the DP model could all be questioned, it is unlikely such a model could forecast accurately. In contrast, they would view the MNL models, which simply model the value of each alternative as a flexible function of the state variables, as being much less restrictive. The evidence presented in the paper would seem to contradict that view.

3.2 The Role of Theory in Inference Based on Instrumental Variables—Estimating the Effect of School Attainment on Earnings

3.2.1 Background

The initial applications operationalizing the concept of human capital as an investment in embodied skills focused on estimating the internal rate of return (or marginal efficiency of investment) to schooling. The first empirical calculations of rates of return were based on comparisons of earnings profiles of synthetic cohorts with different levels of schooling (Carnoy, 1967; Hanoch, 1967; Hansen, 1963). The magnitudes of these estimated internal rates of return provided evidence of the credibility of the investment view of education.

A major innovation was the development of a shorthand way of estimating the internal rate of return based on a regression of earnings on school attainment (Becker and Chiswick, 1966; Mincer, 1974). What

has come to be called the "Mincer earnings function" is perhaps the single most estimated relationship in economics. However, as discussed in detail by Heckman, Lochner, and Todd (2008), the interpretation of the schooling coefficient in an earnings or wage regression as an internal rate of return is fragile, depending on a special set of assumptions (for example, no direct costs of schooling, perfect foresight). Yet, it is commonplace to motivate the earnings function as Mincerian even in settings in which the assumptions do not obviously hold.

In hindsight, this tradition is surprising given an alternative interpretation of the wage function that is, in fact, more closely tied to the original notion of human capital as an investment good and to economic theory. Ben-Porath (1967), following the foundational work of Becker (1962, 1964, 1967) and Mincer (1958, 1962), provided a formalization of the acquisition of human capital within an explicit optimization framework. The essential idea was that human capital is produced according to a technology that combines an individual's time and market goods. In Ben-Porath's formulation, the amount of human capital produced over some time period depended on the beginning-of-period stock of human capital, the amount of time (the fraction of the total available time) an individual devoted to producing new human capital, the amount of purchased inputs (for example, books, equipment), and the individual's innate ability.[37] The cost of producing human capital is the forgone earnings associated with the time withdrawn from market work and the direct cost of the purchased inputs. An individual's market wage is the product of a competitively determined rental price per unit of human capital times the amount of human capital employed in the labor market (the stock of human capital times the fraction of time not devoted to producing new human capital).

A basic result from this model is that there would be a beginning period of specialization, when the human capital stock is low, in which individuals would spend all of their time in the production of human capital, and that after that period the production of human capital would decline (due to the finite horizon). Mincer (1974) used the on-the-job training accounting framework developed in Becker and Chiswick (1966) together with the result from Ben-Porath that human capital production would decline over the life cycle to derive the quadratic relationship between (potential) work experience and earnings. Mincer combined the schooling model, based on equating the present value of income streams across schooling levels, with the accounting model to obtain the familiar logarithmic form of the earnings function.

Griliches (1977) used the production function concept and competitive labor market assumption of Ben-Porath to formulate a wage-schooling function with a very different interpretation than that in Mincer. Specifically, letting w be the wage, r the human capital rental price, H the stock of human capital, and S years of schooling, the wage function is derived as follows:

$$w = rH, \tag{3.20}$$

$$H = e^{\beta_0 + \beta_1 S} e^{v}, \tag{3.21}$$

where v are other determinants of human capital including ability. Substituting equation 3.21 into equation 3.20 and taking logs yields the semilog form

$$\log w = \log r + \beta_0 + \beta_1 S + v. \tag{3.22}$$

The interpretation of the schooling coefficient in equation 3.22 is not as an internal rate of return but rather as a production function parameter. It thus reflects the combined productivity of school inputs and student time in producing marketable human capital or skills. Note that the constant term is the sum of the log of the skill rental price and a skill endowment (β_0). A learning-by-doing model of on-the-job skill accumulation leads to an augmented wage function in schooling and actual work experience that mimics the Mincer formulation, namely

$$\log w = \log r + \beta_0 + \beta_1 S + \alpha_1 X - \alpha_2 X^2 + v, \tag{3.23}$$

where, with positive and declining marginal product, $\alpha_1, \alpha_2 > 0$.

It was recognized quite early that the estimation of equation 3.23 was problematic because unmeasured ability, contained in v, would likely be correlated with years of schooling. Hanoch (1967) explicitly noted that his estimates of the internal rate of return would be biased if more able individuals obtained more schooling. Proposed solutions to the problem included using test scores as (imperfect) measures of ability (e.g., Griliches and Mason, 1972; Hause, 1972) and using siblings to control for family-level unobservables (Chamberlain, 1977). Scores of papers were written trying to assess the extent of the bias based on the specification in equation 3.23.

Hause (1972) argued that the proper specification of the wage function would include an interaction between schooling and ability. Letting A be ability, he proposed (and estimated with a measure of ability) the wage function

$$\log w = \log r + \beta_1 S + \beta_2 A + \beta_3 S \cdot A + Z\gamma + u \tag{3.24}$$

where Z included family background variables. The rationale for the interaction effect was that ability was factor augmenting, enhancing the productivity of schooling inputs, implying that β_3 should be positive.[38]

More recent attempts to estimate the effect of schooling on wages free of ability bias have used instrumental variable approaches that rely on natural experiments or quasi-experiments.[39] The IV approach to the identification of the "causal effect" of schooling on earnings has been studied by Card (1995, 1999), Heckman and Vytlacil (1998), and Heckman, Lochner, and Todd (2006) in the situation exemplified by equation 3.24, which can be rewritten in the form of a regression model with random coefficients, namely

$$\log w_i = \beta_{0i} + \beta_{1i} S + Z\gamma + u_i. \tag{3.25}$$

In equation 3.25, $\beta_{0i} = \log r + \beta_2 A_i$ and $\beta_{1i} = \beta_1 + \beta_3 A_i$. These papers show the sources of bias associated with the OLS estimator that ignores unmeasured ability and the conditions under which an IV estimator can identify the mean schooling effect, $\beta_1 + \beta_3 \bar{A}$.

In general, IV estimators do not recover mean effects when effects are heterogeneous (Heckman, 1997; Imbens and Angrist, 1994). It follows then that alternative instruments estimate different effects. It is precisely in this circumstance that theory is of value. Yet, it is an aspect of the IV literature that theory is usually eschewed. Without a theory incorporating the role of the instrument, it is not possible to interpret the estimates or to tell what assumptions are implicit in any given interpretation.

3.2.2 Instrumental Variables Estimators of the "Causal" Effect of Schooling on Earnings

Angrist and Krueger (1991)[40] present a prominent application of the natural experiment approach. They provide a stark example of the value of using (even quite simple) received theory to interpret empirical work.

Consider a standard model of schooling choice incorporating ability heterogeneity. Assume that the (annual) wage at any age a, w_a, depends on three factors—the level of school attainment, S, the amount of work experience at age a, X_a, and ability, μ, according to

$$w_a = f(S, \mu) + g(X_a, \mu), \tag{3.26}$$

where f and g are monotonically increasing in their arguments.[41] All individuals are assumed to work full time after completing schooling. To conform to the natural experiment in the Angrist and Krueger (1991) paper described below, the model explicitly incorporates a mandated school entry age a_e and a mandated minimum school-leaving age a_l. Thus, school attainment at age a_l, "initial schooling," is $S_0 = a_l - a_e$.

It is convenient for what follows to limit the decision horizon by assuming that the individual decides on whether to attend school only for one period beyond the school-leaving age. School attendance in the period following a_l, period 1, is denoted by $s_1 = 1$ and nonattendance by $s_1 = 0$; completed schooling, at the end of period 1, S_1, is therefore either $S_0 + 1$ or S_0. An individual who decides not to attend school in period 1 works that period and all subsequent periods; that is, the individual works from period $a = 1$ to the end of working life $a = A$, whereas an individual who attends school does not work in that period but works in all subsequent periods, that is, from $a = 2$ to $a = A + 1$. There is a direct cost of attending school denoted by c. The assumption that school attendance precludes working implies that school attendance entails an opportunity cost in terms of the earnings that are forgone.

The individual is assumed to make the choice of whether or not to attend school according to which option maximizes the present discounted value of lifetime earnings. Denoting the discount factor as δ ($= 1/[1 + r]$), the present discounted value of lifetime earnings (the annual wage) under the two schooling alternatives, $V(s = 1 \mid S_0)$ and $V(s = 0 \mid S_0)$, are

$$V(s_1 = 1 \mid S_0) = \exp[f(S_0 + 1, \mu] \sum_{x=0}^{X} \delta^{x+1} \exp[g(x, \mu)] - c$$

$$V(s_1 = 0 \mid S_0) = \exp[f(S_0, \mu] \sum_{x=0}^{X} \delta^{x} \exp[g(x, \mu)].$$

The decision rule is to attend school if $V(s_1 = 1 \mid S_0) \geq V(s_1 = 0 \mid S_0)$, which reduces, after some manipulation, to[42]

$$s_1 = 1 \text{ if } f(S_0 + 1, \mu) - f(S_0, \mu) \geq r + \frac{c}{V(s_1 = 0 \mid S_0)} \tag{3.27}$$

$= 0$ otherwise.

Thus, the individual attends school if the percentage increase in earnings from attending school for the additional period is sufficiently greater than the interest rate.

If the increase in wages with additional schooling is greater for the more able, $\frac{\partial(f(S_0 + 1, \mu) - f(S_0, \mu))}{\partial \mu} > 0$, then there exists a cutoff value of ability, μ^*, such that individuals with ability at or above that cutoff attend school and those below it do not. Thus, the difference in earnings among the two school completion groups will reflect, in part, these induced ability differences. Although ability is distributed randomly in the population, optimizing behavior creates a positive correlation between ability and completed schooling, which implies that the average schooling effect calculated from (log) earnings differences between the two schooling groups (for example, from an OLS regression earning on schooling) overstates the schooling effect averaged over the ability distribution, that is,

$$E_\mu[f(S_0 + 1, \mu \mid \mu \geq \mu^*)] - E_\mu[f(S_0, \mu \mid \mu < \mu^*)] > E_\mu[f(S_0 + 1, \mu) - f(S_0, \mu)]. \tag{3.28}$$

The challenge has been to obtain an estimate of $E_\mu[f(S_0 + 1, \mu) - f(S_0, \mu)]$, the right hand side of equation 3.28. Whether the value of meeting that challenge is a worthwhile goal is an issue that will be discussed below. First, however, I consider whether the approach suggested by Angrist and Krueger (1991) achieves the goal.

Angrist and Krueger (1991) use natural variation in dates of birth, in conjunction with the existence of a birth date cutoff for school entry and a minimum compulsory school-leaving age, as an "instrument" for completed schooling to identify the mean schooling effect. To illustrate, consider a particular locality defined by a school entry age and a minimum school-leaving age in which a child must attain the age of 6 as of September 1 in order to enter the first grade and cannot leave school prior to attaining the age of 16. Thus, in the previous notation, $a_e = 6$ and $a_l = 16$. Then, a child whose birthday falls on September 1 will have 10 years of schooling at the minimum school-leaving age ($S_0 = 10$). However, a child born in the same calendar year but a day later, on September 2, will enter the first grade one year later and will have completed only 9 years of schooling on reaching age 16 ($S_0 = 9$). Thus, if children in the two (day of birth) cohorts are otherwise identical and at least some of them prefer nine or fewer years of schooling, average completed schooling of the two cohorts will differ because some of the children in the September 1 cohort will have been forced to obtain an additional year of schooling.

An additional simplification of the model illuminates how the Angrist and Krueger (1991) instrumental variable estimator works. Assume that there are just two ability types, denoted as μ_1 and μ_2, with the first type of higher ability ($\mu_1 > \mu_2$). Type 1's comprise π_1 proportion of the population, and type 2's $\pi_2 = 1 - \pi_1$. The insight of Angrist and Krueger (1991) is that, given laws governing the ages at which children can enter and leave school, completed schooling will vary with birth date for some part of the population. Date of birth is assumed to be a random variable uncorrelated with ability (as transmitted intergenerationally). Because the instrument relies only on within-state variation in dates of birth, laws governing school-entering and school-leaving ages, set by each state, can be correlated with other state-specific unobservables that influence earnings without inducing bias. Thus, variation in state schooling laws are not treated as natural experiments, can be endogenously determined, and do not contribute to identification.[43] It is variation in date of birth within states that serves to identify schooling effects. Having information on multiple states with differing laws merely adds precision to the estimates.

Angrist and Krueger (1991) present Wald estimates of the schooling effect on earnings based on comparing (log) weekly earnings and school completion levels for two cohorts of men aged 41–50 who differed in their quarter of birth, specifically comparing those born in the first quarter of a calendar year to those born in the other three quarters (of the previous calendar year).[44] In terms of the model, it is easy to show what their Wald estimator identifies. Suppose, according to equation 3.27, that the optimal level of schooling for type 1's, the high-ability type, is $S_0 + 1$ and that for type 2's is exactly S_0, given the school entry age (a_e) and the minimum school-leaving age (a_l). In that case, type 1's obtain $S_0 + 1$ years of schooling and have mean earnings of $f(S_0 + 1, \mu_1)$ regardless of date of birth. Type 2's, however, obtain different levels of schooling depending on when they were born. Older type 2's, those who make the deadline (born on September 1 in the example), obtain $S_0 + 1$ years of schooling and have mean earnings of $f(S_0 + 1, \mu_2)$, whereas younger type 2's who miss the deadline (those born on September 2 in the example) obtain only S_0 years of schooling and have mean earnings of $f(S_0, \mu_2)$. Thus, the mean earnings of younger children is

$$\pi_1 f(S_0+1,\mu_1)+(1-\pi_1)f(S_0+1,\mu_2),$$

that of older children

$$\pi_1 f(S_0+1,\mu_1)+(1-\pi_1)f(S_0,\mu_2),$$

and the difference in mean earnings between the younger and older children

$$\Delta Ew = (1-\pi_1)[f(S_0+1,\mu_2)-f(S_0,\mu_2)]. \tag{3.29}$$

Similarly, the difference in mean level of schooling in the population is

$$\Delta ES = \pi_1(0)+(1-\pi_1)\cdot 1 = (1-\pi_1). \tag{3.30}$$

Thus, the Wald estimator, the ratio of equation 3.29 to equation 3.30 is

$$\frac{\Delta Ew}{\Delta ES} = f(S_0+1,\mu_2)-f(S_0,\mu_2),$$

which is the marginal effect of schooling on earnings for the less-able children, those induced to change their schooling by the instrument.

Under the maintained assumption that ability augments the effect of schooling on human capital (and thus wages), the Wald estimator is thus less than the population average effect $E_\mu[f(S_0 + 1, \mu) - f(S_0, \mu)]$. The two would coincide only if schooling and ability did not interact in the human capital production function, that is, if $f(S_0 + 1, \mu) = f_1(S) + f_2(\mu)$. Notice that if this latter condition held, there would be no need for the instrument, as the model would predict that the schooling decision is unrelated to ability (as long as the rate of interest and the direct cost of schooling did not vary with ability).

Angrist and Krueger (1991) found that for the group of men aged 41–50 in 1970 (1980), those born in the first quarter of a calendar year obtained 0.1256 (0.1088) fewer years of schooling on average than those born in the other quarters. Presumably, those men born in the first quarter of a calendar year were more likely to have had to delay school entry than those born in the previous three quarters. Dividing those differences by the corresponding differences in (ln) weekly wages implied a return to schooling of 0.0715 in 1970 and of 0.1020 in 1980. In contrast, the OLS estimate was larger in 1970, 0.0801, and smaller in 1980, 0.0709.

Butcher and Case (1994) suggest that natural variation in the sex of siblings, in particular whether a girl has any sisters, can be used to obtain an estimate of the average schooling effect on earnings (of women) that is free of ability bias. They discuss several reasons for the gender of siblings to affect parental human capital investments in a given child. In particular, the gender of siblings may affect the cost of investing in a child's human capital through the existence of borrowing constraints if there are exogenous gender differences in the return to human capital. According to Butcher and Case (1994), in the presence of borrowing constraints and assuming that boys receive a higher return to each level of schooling, "we should expect to see not only that boys receive more education but also that the presence of sons reduces the educational attainment of daughters." Also, independent of preferences, there may be exogenous differences in the cost of raising girls relative to boys, thus affecting the household's overall budget constraint. In addition, they argue that the gender of one's siblings may affect a child's preferences for schooling investments, with girls who have brothers perhaps adopting "masculine" traits and vice versa. Finally, parents may simply prefer to invest differentially in girls and boys depending on the overall gender composition of their children.[45]

To illustrate the Wald estimator of the return to schooling based on the gender of the first birth instrumental variable, in the context of the model used to elucidate the AC (1991) experiment, assume, as is consistent with the Butcher and Case (1994) argument, that the cost of schooling to a secondborn girl is higher if the firstborn is a boy as opposed to a girl, specifically c_b and c_g, respectively, with $c_b > c_g$. Ability is assumed independent of sex. Suppose that given these costs, it is optimal for the girl with a firstborn brother not to attend school regardless of her ability type, so that she completes only the minimum, S_0, years of schooling. On the other hand, suppose that among girls with a firstborn sister, who face a lower cost of schooling, type 1's optimally choose to attend school, thus completing $S_0 + 1$ years, whereas the type 2's still do not attend and complete only S_0 years. In this case, mean schooling of those girls with firstborn brothers is $\pi_1 f(S_0, \mu_1) + (1 - \pi_1)f(S_0, \mu_2)$ and for those with firstborn sisters, $\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1)f(S_0, \mu_2)$. Thus, the difference in mean earnings of girls with firstborn sisters and firstborn brothers is $\pi_1[f(S_0 + 1, \mu_1) - f(S_0, \mu_1)]$, and the difference in mean schooling levels is π_1. Therefore, the Wald estimator is

$$\frac{\Delta Ew}{\Delta ES} = f(S_0 + 1, \mu_1) - f(S_0, \mu_1),$$

the effect of the additional schooling on earnings for the more able group. Thus, opposite to the Angrist and Krueger (1991) instrument,

the variation in completed schooling that arises from the sex of the firstborn sibling overstates the mean schooling effect. The reason that the Butcher and Case (1994) instrument leads to an overstatement of the mean schooling effect, in contrast to the understatement using the Angrist and Krueger (1991) instrument, is due not to the Butcher and Case instrument reflecting variation in the cost of schooling but rather to the relationship of ability to the schooling decision as the instrument varies.

Card (1995), for example, uses variation in the proximity of a 4-year college when growing up as an instrument for school attainment that, like the Butcher and Case instrument, is naturally interpreted as reflecting variation in the cost of schooling, c. Clearly, as seen from equation 3.27, conditional on ability, those who grew up in proximity to a college and thus faced a lower schooling cost would be likely to obtain more schooling than those who were not in close proximity. Maintaining the simplification of only two ability types, we then assume that the more able type, μ_1, would attend college, that is, choose $S_0 + 1$ years of schooling regardless of whether they grew up near a college, whereas the low-ability type, μ_2, would obtain $S_0 + 1$ years of schooling if they grew up near a college and S_0 years if they did not. In that case, mean earnings of those who grew up near a college would be $\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1)f(S_0 + 1, \mu_2)$, and mean earnings of those who did not $\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1)f(S_0 + 1, \mu_2)$. This is exactly the situation that arises in the Angrist and Krueger (1991) case and, as in that case, leads to a Wald estimator reflecting the effect of schooling on earnings for the less able.

Both the Angrist and Krueger (1991) and Butcher and Case (1994) instrumental variables should yield the same estimate if there is no interaction between schooling and ability. However, as Rosenzweig and Wolpin (2000) discuss, even in that case, the interpretation of the IV estimator is still dependent on specification. Both specifications of the earnings function control for age rather than, as suggested by Mincer (1974), (potential) work experience. Rosenzweig and Wolpin (2000) show that this seemingly innocuous specification choice leads to bias in both the Angrist and Krueger (1991) and Butcher and Case (1994) estimates and that the biases are of opposite sign. The existence of a theoretically valid natural instrument thus does not mean that the specification of the equation of interest does not matter for identification. That is, bias can still arise due to misspecification.[46]

3.2.3 Why Estimate the Causal Effect of Schooling on Wages?

The mark of a mature literature is when the rationale for a new study relies solely on reference to past studies. Identifying the effect of schooling on wages is by that definition very ripe indeed; the preoccupation of labor economists over the last 40 years with the estimation of β_1 in equation 3.22 has become sufficient justification for attaching great value to the development of new methods whose sole purpose is to provide a better estimate. As noted, an important rationale for the early literature on estimating the internal rate of return to schooling was to validate the notion of schooling as an investment. The identification of that parameter as a measure of the internal rate of return perhaps accounts for that preoccupation. However, knowledge of that parameter, when viewed in the context of richer behavioral models, is of considerably less value.

To illustrate this last proposition, consider a simplified version of the DCDP model of schooling, work, and occupational choice in Keane and Wolpin (1997). In that model, individuals choose at each age (a = 16 to 65) whether to work in any of three occupations (white collar, blue collar, or the military), to attend school, or to neither attend school nor work (the home option). These alternatives are mutually exclusive and exhaustive. Per-period utility flows are linear in consumption (C) and nonpecuniary values attached to each option (denoted below by α's). Specifically, letting $d_{ms} = 1$ if alternative m is chosen and equal to zero otherwise, where $m = 1$ indicates working in a white-collar occupation, $m = 2$ working in a blue-collar occupation, $m = 3$ working in the military, $m = 4$ attending school, and $m = 5$ staying home, flow utility at age a is

$$U(C_a, \mathbf{d}_a) = C_a + \sum_{m=1}^{5} \alpha_m d_{ma} + \alpha_6 d_{4a}(1 - d_{4,a-1}), \tag{3.31}$$

where $\mathbf{d}_a$ is the vector of alternatives d_{ma}. The utility flow from school attendance includes a potential psychic cost of returning to school (α_6).[47]

Consumption is the occupation-specific wage, w_{ma}, if the individual works plus any (alternative-specific) monetary transfers if the individual does not work, γ_4 and γ_5, minus the cost of attending college, t_1, given that alternative is chosen, namely

$$C_a = \sum_{m=1}^{3} w_{ma} d_{ma} + \sum_{m=4}^{5} \gamma_m d_{ma} - t_1 I(12 \le S_a < 15) d_{4a} \tag{3.32}$$

where S_a is the grade attained by age a. The Ben-Porath/Griliches wage offer function in each occupation is the product of a competitively determined skill rental price (assumed to be constant over time) and the amount of embodied occupation-specific skill. The number of occupation-specific skill units possessed by an individual at age a depends on the years of schooling completed by that age and on the number of years of work experience in each occupation at that age. If we let e_{ma} be the number of skill units, x_{ma} the number of years of work experience, $e_{16,m}$ the endowment of skill units at age 16, and ε_{ma} a shock, the production functions for skill units is given by

$$e_{ma} = \exp(e_{16,m} + \beta_{m1} S_a + \sum_{m=1}^{3} \beta_{m2} x_{ma} - \beta_{m3} x_{ma}^2 + \varepsilon_{ma}) \text{ for } m = 1,2,3.$$

This specification leads to a log wage function of a standard form,

$$\log w_{ma} = \log r_m + e_{16,m} + \beta_{m1} S_a + \sum_{m=1}^{3} \beta_{m2} x_{ma} - \beta_{m3} x_{ma}^2 + \varepsilon_{ma}. \tag{3.33}$$

Note that the constant term is the sum of the (log) skill rental price, r_m, and the individual's skill endowment at age 16. The utility flows to work are stochastic because wage offers are stochastic. Utility flows to school and home (the α's) are also assumed stochastic, that is, $\alpha_{4a} = \overline{\alpha}_4 + \varepsilon_{4a}$ and $\alpha_{5a} = \overline{\alpha}_5 + \varepsilon_{5a}$. These parameters may be thought of as reflecting productivity differences in school attainment (an effort cost) and in home production (leisure). The vector of shocks, $\boldsymbol{\varepsilon}_a = [\varepsilon_{1a}, \varepsilon_{2a}, \varepsilon_{3a}, \varepsilon_{4a}, \varepsilon_{5a}]$ is assumed to be jointly serially independent and distributed joint normal.

Individuals begin the decision process at age 16 with a set of permanent endowments consisting of their occupation-specific skills, $e_{16,m}$ (m = 1, 2, 3). and their preferences over attending school, $\overline{\alpha}_4$, and of remaining at home, $\overline{\alpha}_5$. Endowment heterogeneity is captured by assuming there to be K discrete types, with each type distinguished by a vector of endowments, $\mathbf{e}_{16}^k = [e_{16,1}^k, e_{16,2}^k, e_{16,3}^k, \overline{\alpha}_4^k, \overline{\alpha}_5^k]$. Thus, individuals may have comparative advantages among the different alternatives, including in acquiring schooling and in home production, that are known to them. Each person also starts with S_{16} years of schooling and zero work experience ($x_{16} = 0$). Unobserved endowments reflecting premarket skills (including innate abilities) enter the occupation-specific wage functions in the same way as ability does in the "causal effects" literature.

At any age a, the individual's objective is to maximize the expected present discounted value of remaining lifetime utility. Define $V_a^k(\Omega_a^k)$, the value function, as the maximal expected present value of remaining lifetime utility of an individual of type k, where Ω_a^k is the state space, that is, the history relevant for determining remaining expected lifetime utility. On substituting the wage function (equation 3.33) into the budget constraint (equation 3.32) and the resulting budget constraint into equation 3.31, the flow utility at age a for an individual of type k can be written as a function of the state variables $\Omega_a^k = [\mathbf{e}_{16}^k, S_a^k, x_a^k, d_{1,a-1}^k, \boldsymbol{\varepsilon}_a]$. Thus,

$$V_a^k(\Omega_a^k) = \max_{\{\mathbf{d}_a\}} E\sum_{t=a}^{65}\delta^{t-1}\sum_{m=1}^{5}U_{ma}(\Omega_a^k)d_{ma}. \tag{3.34}$$

The maximization in equation 3.34 is achieved by choosing the optimal sequence of control variables, $\mathbf{d}_a$, for $a = 16$ to 65.

As previously discussed, one can rewrite the value function as the maximum over alternative-specific value functions,

$$V_a^k(\Omega_a^k) = \max_m\{V_{ma}^k(\Omega_a^k)\},$$

where

$$V_{ma}^k(\Omega_a^k) = U_{ma}(\Omega_a^k) + \delta E[V_{a+1}^k(\Omega_{a+1}^k) \mid \Omega_a^k, d_{ma} = 1] \text{ for } a < 65, \tag{3.35}$$

$$= U_{m,65}(\Omega_{65}^k).$$

The expectation in equation 3.35 is taken over the distribution of the random elements in Ω_{a+1}^k conditional on Ω_a^k, that is, over the unconditional distribution of $\boldsymbol{\varepsilon}_{a+1}$, given that $f(\boldsymbol{\varepsilon}_{a+1} \mid \boldsymbol{\varepsilon}_a) = f(\boldsymbol{\varepsilon}_{a+1})$. Schooling at a + 1 is given by $S_{a+1} = S_a + d_{4a}$ and occupation-specific work experience at a + 1 by $x_{m,a+1} = x_{ma} + d_{ma}$ for $m = 1, 2, 3$.

The solution of the optimization problem serves as the input into estimating the parameters of the model given data on choices and wage offers for workers. Although the solution is deterministic for the individual, it is probabilistic from the researcher's view because contemporaneous shocks are not observed. Consider having data on a sample of N individuals from the same birth cohort who are assumed to be solving the model described above and for whom choices are observed over at least a part of their lifetimes. In addition, assume, as is the case, that wages are observed only in the periods in which market work is chosen and only for the occupation that is chosen. Thus, for each individual, n, the data consist of the outcomes $\{d_{nma}, w_{nma}d_{nma} : m = 1, 2, 3\}$

and $\{d_{nma}: m = 4, 5\}$ for all ages within the age range $\{16, \bar{a}_n\}$.[48] Let c_a denote the outcome data at age a and let $\Omega_a^{k(-)}$ denote the components of the state space observed by the researcher, that is Ω_a^k net of the shocks. The probability of any sequence of outcomes for an individual n given the individual is of type k and has initial schooling $S_{n,16}$ provides the likelihood contribution, L_{nk},

$$L_{nk} = \Pr(c_{n,16}, \dots, c_{n,\bar{a}_n} \mid S_{n,16}, type = k).$$

Because the researcher does not observe an individual's type, the likelihood function is a mixture of these type-specific likelihoods,

$$\prod_{n=1}^{N}\sum_{k=1}^{K}L_{nk}\Pr(type = k, S_{n,16}).$$

It is unlikely that initial schooling (at age 16) is exogenous, in which case conditioning the likelihood on it as though it were nonstochastic is problematic. One remedy would be to specify the optimization problem back to the age at which initial schooling was zero for everyone (say, age 3) and solve for the correct probability distribution of attained schooling at age 16 (conditional on age 3 endowments). However, such a model would have to focus on parental decision making with respect to investments in children (including fertility decisions) and would be very demanding in many dimensions (modeling, computation, and data).[49] The alternative Keane and Wolpin (1997) follow is to assume that initial schooling is exogenous conditional on the age 16 endowment vector. Given this assumption, the likelihood function can be conditioned on initial schooling as follows:

$$\prod_{n=1}^{N}\sum_{k=1}^{K}L_{nk}\Pr(type = k \mid S_{n,16}),$$

where $\Pr(type = k \mid S_{n,16})$ is estimated nonparametrically.[50] Identification in this model is achieved by a combination of exclusion restrictions and functional form and distributional assumptions.[51] The model is estimated using data from the 1979 cohort of the National Longitudinal Surveys of Labor Market Experience (NLSY79).

The first thing to note about this model relative to the "causal effect of schooling" literature is that there is no single causal effect. An increase in schooling affects wage offers in all three occupations. If wage offers were observed and a pooled wage regression estimated, given that occupation is a choice, the effect of additional schooling on wages would depend on the distribution of endowment types across occupations and would not convey anything of interest. The second thing to note is that these causal effects, the β's, are only three of the structural parameters of a rich model of behavior.[52] By themselves, these schooling parameters reveal little, if anything, that is of direct substantive interest.

In contrast, the point and strength of the structural approach are that the estimates of the model enable researchers to address important issues of consequence, issues that cannot be addressed in the causal effects literature. Before we consider some examples, it is useful to summarize some of the results of the estimation in order to better understand the mechanisms at work. Keane and Wolpin (1997) allow for four endowment types and for two levels of initial schooling (at age 16). A principal finding is that endowments differ and matter (a lot). Compared to type 1's, wage offers (at the same levels of schooling and experience) in the white-collar occupations are 8.7 percent lower for type 2's, 60.9 percent lower for type 3's, and 52.0 percent lower for type 4's. Although type 1's are endowed (at age 16) with the most white-collar skill, type 2's have the largest blue-collar skill endowment. Relative to type 1's, type 2's wage offer in that occupation is 30.5 percent higher, type 3's is 21.1 percent lower, and type 4's 5.5 percent lower. In addition, there is considerable heterogeneity in the (net) nonpecuniary values attached to attending school and staying at home. Type 1's value a year of schooling at \$11,031 and a year at home at \$20,242. The comparable figures are \$5,667 and \$18,107 for type 2's, \$2,131 and \$5,564 for type 3's, and \$9,562 and \$17,330 for type 4's.[53] Keane and Wolpin (1997) calculate that 90 percent of the total variation in expected lifetime utility is the result of between-type variation, with the rest a consequence of within-type idiosyncratic shocks (the sequence of ε_a's).

These endowment differences translate into different behaviors and welfare. Table 3.5 shows the levels of completed schooling and choice probabilities at age 24 and the expected present discounted value of lifetime utility (in 1987 thousands of dollars) for each initial schooling and type. Sample proportions are given in the last row. As shown, of the four distinct types that are identified in the estimation, type 1's, comprising 17 percent of the sample, complete about 4 more years of schooling than any of the others, over 16 years on average, and specialize in white-collar employment. Based on the estimated white-collar wage function, those additional four years of schooling would translate

Table 3.5
Selected Characteristics at Age 24 and EPDV of Lifetime Utility by Type and Initial Schooling

	Initial Schooling ≤ 9				Initial Schooling ≥ 10			
	Type 1	Type 2	Type 3	Type 4	Type 1	Type 2	Type 3	Type 4
Mean Schooling	15.6	10.6	10.9	11.0	16.4	12.5	12.4	13.0
Proportion in: White-Collar	0.509	0.123	0.176	0.060	0.673	0.236	0.284	0.155
Blue-Collar	0.076	0.775	0.574	0.388	0.039	0.687	0.516	0.441
Military	0.000	0.000	0.151	0.010	0.000	0.000	0.116	0.005
School	0.416	0.008	0.013	0.038	0.239	0.024	0.025	0.074
Home	0.000	0.095	0.086	0.505	0.050	0.053	0.059	0.325
Expected Present Value of Lifetime Utility[1]	387	374	214	286	416	396	229	291
Sample Proportion	0.010	0.051	0.103	0.090	0.157	0.177	0.289	0.123

1. Thousands of 1987 dollars.
Source: Keane and Wolpin (1997), Tables 11, 12, and 13.

into a wage offer that is 28 percent higher for type 1's than that for the other types. Although that represents a large differential, recall that the difference in wage offers based on white-collar skill endowments was twice as large as for type 3's and type 4's. Type 2's, comprising another 23 percent of the sample, complete, on average, 12 years of schooling and specialize in blue-collar employment. Interestingly, the expected present value of lifetime utility is about the same for the two types.[54] The third type, 39 percent of the sample, is the only type to enter the military; about 13 percent are in the military at age 24. They also complete about 12 years of schooling and, other than military service, specialize in blue-collar employment. However, as compared to the second type, who have much the same schooling and employment patterns, because type 3's have considerably lower age-16 levels of premarket skills, their expected lifetime utility is only about three-fifths as great. The last type, the remaining 21 percent, also completes about 12 years of schooling on average, but about 40 percent choose to be at home at age 24, about five times as much as any other group.

Initial schooling also significantly affects choices. The difference in school attainment that has already emerged by age 16, a difference of about 1.4 years, that is due to exogenous events unrelated to endowments, for example, illnesses that occurred in the past or date of birth that affects age at entry given the minimum school entry age (as in Angrist and Krueger, 1991), is magnified over time for two of the four types. Although the initial difference is reduced to about 0.8 years for type 1 individuals and is essentially unchanged for type 3 individuals, the completed schooling difference increases to about 2 years for types 2 and 4. However, initial schooling differences have only a small impact on expected lifetime utilities (ranging over types from about 1 to 8 percent higher for the highest initial schooling group).

Under the assumption that the four types capture all unobserved differences in premarket endowments, it is possible to determine the extent to which conventional measures of family background are able to account for differences in lifetime utilities that arise from age-16 endowments. In the baseline, the difference in expected lifetime utility between the first type (the highest) and the third type (the lowest) is estimated to be $188,000 (1987 dollars). In comparison, the difference for individuals whose mothers were college graduates relative to those whose mothers were high school dropouts is estimated to be $53,000, the difference for those who lived with both parents at age 14 relative to those who lived in a single parent home is $15,000, and the difference

between those whose parental income was more than twice the median relative to those less than one-half the median $66,000. Although the association of expected lifetime utility with family income and mother's schooling appears to be strong, a regression of expected lifetime utility on those family background characteristics is reported to explain only 10 percent of the variation. Coupled with the fact that 90 percent of the total variation in expected lifetime utility is determined to be due to between-type variation, these family background characteristics are in actuality only imperfect proxies for premarket skills.

3.2.3.1 Ex Ante *Policy Evaluation*

Continuing with the theme of the chapter 2, I consider several applications of the Keane and Wolpin (1997) model to *ex ante* policy analysis. Keane and Wolpin (2000) applied the Keane and Wolpin (1997) model to analyze race differences in educational and labor market outcomes. The disparity in educational attainment between white and black males is substantial. For example, Keane and Wolpin (2000) report that for a subset of males in the 1979 youth cohort of the NLSY79 who were 13–16 years old by October 1977 (people born in the early 1960s), mean highest grade completed at age 23 (at which point education is mostly completed) was 12.7 for whites compared to 11.9 for blacks. In addition, 21.3 percent of whites had completed college, compared to only 6.8 percent of blacks and 35.9 percent of blacks were high school dropouts, compared to 24 percent of the whites. In terms of (accepted) wages, for that same cohort, at age 26 blacks working in a white-collar occupation earned only 61 percent as much as whites, and those working in a blue-collar occupation 75 percent as much.

To understand these differences, Keane and Wolpin (2000) estimated a restricted version of the model in which all parameters for blacks and whites were constrained to be equal except for (1) the type proportions and (2) the intercepts in the earnings functions. The type proportions capture race differences in age-16 skill endowments and preferences through a shift in type-specific population weights. The intercepts capture both race differences in skill rental prices due to discrimination and differences in the level of the skill endowments of each type. Strikingly, compared to a model in which all parameters were allowed to differ by race, this severe restriction on the model was not rejected. Keane and Wolpin (2000) find that the constant terms in the wage offer equations are 7.9 percent lower for blacks in the white-collar occupation and 4.7 percent lower in the blue-collar occupation. The differences in type proportions are even more pronounced: only 17.4 percent of blacks are types 1 or 2 (the high-wage types) compared to 43.2 percent of whites. Conversely, 41.6 percent of blacks are type 4's (which accounts for most high school dropouts) compared to only 20.3 percent of whites.

Keane and Wolpin (2000) examine the extent to which policies that provide financial incentives for school attendance can reduce racial differences in school attainment and wages. Based on the model estimates, they calculate that providing blacks with a bonus of $6,250 for finishing high school and an additional $15,000 for finishing college (both in 1987 dollars) would essentially equalize the distribution of completed education for whites and blacks.[55] They calculate that implementation of such a scheme would cost about $2 billion per year (in 1987 dollars).

One might expect that the equalization of schooling would go a long way to equalizing lifetime earnings, both because schooling directly increases earnings in all occupations and because those with higher schooling work in high-wage (white-collar) occupations. However, Keane and Wolpin (2000) find that the large increase in education induced by this scheme has very little impact on the mean present value of lifetime earnings for blacks. Under the baseline, Keane and Wolpin simulate that the present value of lifetime earnings is $285,000 for whites and $210,000 for blacks, so lifetime earnings for blacks is only 73.7 percent that of whites. The school attendance bonus scheme increases lifetime earnings for blacks only up to $214,000, which is 75.1 percent of that for whites. Thus, the bonus scheme eliminates only about 5 percent of the gap.

The result that reasonably modest attendance bonuses would substantially increase college attendance of blacks is consistent with a large literature that finds a rather elastic response of college attendance to college costs (see, e.g., Leslie and Brinkman, 1987, or Keane and Wolpin, 2001, for reviews). However, raising college attendance per se does nothing to alter the age-16 endowments or to reduce wage discrimination, the factors that account for the earnings differential between blacks and whites. The importance of these factors can be seen through the following two hypothetical experiments. Equalizing type proportions (that is, giving blacks the same type proportions as whites) would raise the present value of lifetime earnings for blacks to $254,000, whereas eliminating racial discrimination and/or equalizing type-specific skill endowments would raise it to $239,000. Thus, these hypothetical experiments would eliminate 52 percent and 39 percent of the earnings gap, respectively. These are large effects, particularly relative to the effects of equalizing school attainment.

The main conclusion that Keane and Wolpin (2000) draw from these results is that policies such as college financial aid and school attendance bonuses can only be effective ways to reduce lifetime earnings inequality between blacks and whites if they also act to change age-16 endowments. These endowments may be determined by many interrelated factors, such as parental education and income, parental investments in children, pre- and postnatal care, the quality of child care, preschools, primary schools and secondary schools, neighborhood effects, and others. The extent to which the kind of educational policies considered above induce behaviors that affect endowments is unknown. The development of effective policy requires opening the black box of what determines age-16 skill endowments. There is now a growing research agenda that is focusing attention on the formation of both cognitive and noncognitive skills during childhood. I turn to this issue below.

A further advantage of the structural approach is that it enables the researcher to address a number of policy issues within the same model. Keane and Wolpin (2000) also examine whether wage subsidies for low-wage workers would be effective as a means to reduce earnings inequality. Wage subsidy policies have been suggested, and sometimes pursued, as a means of reducing joblessness and providing income support for low-wage workers (see Katz, 1996). Keane and Wolpin (2000) implement a wage subsidy policy akin to that suggested by Phelps (1997). In that policy an individual's wage is augmented by 33 percent of the difference between $12 an hour and the accepted wage; for example, an individual who accepted a wage of $9 would actually receive 10 dollars per hour. The result of the policy is to increase the present value of earnings of whites by 6.4 percent and that of blacks by 16 percent. This is the intended result. However, the policy also has the unintended effect of inducing earlier withdrawal from school. The proportion of high school dropouts increases from 26.2 percent to 35.8 percent for whites and from 37.9 to 41.2 for blacks.[56]

Sullivan (2010) developed a DCDP model that combines a model of job market search (as in Wolpin, 1992, and Rendon, 2006) with a human capital model of schooling and occupational choice (as in Keane and Wolpin, 1997). The previous literature considered job search as a separate phenomenon from schooling and occupational choice, although the choices are clearly related. In Sullivan's (2010 model), workers decide in each period whether to attend school and/or work in one of five occupations or neither work nor attend school. An employed individual may stay at the current job or switch jobs either within the same occupation or with a change in occupation. Human capital accumulated through work experience is both firm- and occupation-specific. Individuals have heterogeneous skill endowments and preferences for employment in different occupations. Wage offers depend on an individual's unobserved occupation-specific skill endowment (modeled as in Keane and Wolpin, 1997), on a firm-worker match-specific component, and on an *iid* time-varying shock. The match-specific component introduces serial correlation into the wage offer process, as it is carried over from period to period in the current job, which is what makes job search valuable. Model parameters are estimated by simulated maximum likelihood using data from the NLSY79 on detailed job histories.

Sullivan (2010) finds that occupational and job mobility are critical determinants of life cycle wage growth, quantitatively more important than the accumulation of occupation-specific human capital. As in previous research, the results also indicate the importance of comparative advantage in understanding schooling and occupational choices. There are large differences in occupation-specific skill endowments at age 16; for example, type 1 individuals receive wage offers in the professional and managerial occupation that are 40 to 70 percent larger than the other (three) types. However, a significant part of the variation in wages is due to difference in match-specific draws, variation that in Keane and Wolpin (1997) would have been accounted for in endowment heterogeneity. Sullivan (2010) finds that endowment heterogeneity accounts for a relatively smaller, although still substantial, component of the variance in expected lifetime utility than found by Keane and Wolpin (1997), 56 percent as opposed to 90 percent.

Ex ante policy evaluation should be conducted within an equilibrium framework. Policies that are of consequence and that impact significant parts of the population can be expected to alter market prices that affect the behavior not only of those targeted by the policy but also of agents in the same market who were not targeted. It is thus necessary to have developed methods and models that account for equilibrium effects. There are few examples of *ex ante* policy evaluations carried out in an equilibrium framework. One example that illustrates again the importance of developing model selection criteria is the assessment of the effect of college tuition subsidies on educational attainment.

In contrast to a graduation bonus scheme, which rewards individuals for achieving a threshold number of years of schooling, tuition subsidies are based only on attendance. Leslie and Brinkman (1987), in a survey of twenty-five empirical studies based on cross-state or time-series variation in college costs, report the modal estimate of the response to a $100 tuition increase (in 1982–1983 dollars) to be a 1.8 percent decline in the enrollment rate of 18- to 24-year-olds. Other estimates range from a low of 0.8 percent for 18- to 24-year-old whites to a high of 2.2 percent for 18- to 24-year-old blacks (Kane, 1994). In Keane and Wolpin (1997), the comparable effect is a decline of 1.2 percent.

These policy effects assume that general equilibrium feedbacks are negligible. However, policies that serve to increase schooling levels and thus the aggregate skill level will reduce the skill rental price in equilibrium. Partial equilibrium estimates of the impact of those policies on schooling would, therefore, overstate the general equilibrium impact. The Keane and Wolpin (1997) framework is amenable to a labor market equilibrium setting. More concretely and for simplicity, assume that there is only a single occupation and that production is Cobb-Douglas, that is, $Y = AS^{\alpha}K^{1-\alpha}$. Aggregate skill is given by the sum of the skill levels over individuals, namely $S = \sum_a \sum_{j=1}^{N_a} S_{ja}$, where N_a is the number of individuals of age a in the population. As noted, it is not possible to separately identify, from data on choices and wages, both the skill rental price and the level of premarket skills. Moreover, given the non-observability of skill, the technology shifter, A, also cannot be identified. A baseline case can be established by normalizing either A or the rental price. With the assumption that $A = 1$, a baseline equilibrium rental price can be established that induces a level of aggregate skill that equates the marginal product of skill to that rental price. The general equilibrium impact of a tuition subsidy on schooling is determined by resolving jointly for the equilibrium rental price and optimal choices.

Donghoon Lee (2005) has implemented this procedure in the multi-skill setting, allowing as well for nonstationarity in population size, that is, for changing cohort size. Lee adopts an overlapping generations version of the occupational and schooling choice model of Keane and Wolpin (1997), assuming that individuals have perfect foresight about equilibrium outcomes but imperfect foresight about their idiosyncratic shocks to preferences and skills. Each cohort therefore faces a known, but different, sequence of occupation-specific skill rental prices. Individuals are assumed to face exogenous cohort and age-specific fertility rates and a constant college tuition cost. Aggregate production at any calendar time depends on the aggregate level of white-collar skill, the aggregate level of blue-collar skill, and capital. Capital grows exogenously at a known rate. The production function is Cobb-Douglas with time-varying factor shares (technology parameters), about which individuals also have perfect foresight.

Lee estimates the model by using aggregate data from the 1968–1993 Current Population Surveys (CPS). The model is fit to the empirical moments of school enrollment rates, occupation-specific employment rates, and educational attainment by age, year, and sex and to wage moments by age, year, sex, and educational attainment. The partial equilibrium effects are in line with those of other studies. Similar to Keane and Wolpin (1997), Lee finds that a $100 tuition increase reduces college enrollment rates by 1.12 percent for 18- to 19-year-old males and by 1.34 percent for 18- to 24-year-old males. General equilibrium effects on completed schooling for the 50 percent subsidy range from 92 to 95 percent of partial equilibrium effects on college enrollment rates for the $100 tuition increase.

In an earlier series of papers, Heckman, Lochner, and Taber (1998, 1999) performed a similar comparison based on a general equilibrium model with different characteristics. The data and estimation methods they adopted also differ significantly from those of Lee. As with Lee, I only briefly describe the essential features of their paper: (1) skill classes are defined by educational attainment; (2) in addition to making an initial discrete schooling decision, in each postschooling period individuals explicitly decide on their human capital investment time on the job as in Ben-Porath; (3) individuals make savings decisions in each period; and (4) labor is inelastically supplied in each period. As they noted, accounting in estimation for the restrictions that general equilibrium imposes given the structure of their model is not feasible with available data. In particular, estimation is hindered by the lack of microdata on consumption and on investment time on the job. For that reason the procedure adopted for recovering the parameters of the model involves a combination of calibration based on results from existing studies and new estimation using both aggregate CPS data (over the period from 1963 to 1993) and longitudinal data from the

NLSY. The partial equilibrium effect of a $100 increase in college tuition costs is 1.6 percent, well within the range of other studies and not much different from that of Lee (2005) or Keane and Wolpin (1997). However, the general equilibrium effect is reduced by an order of magnitude, to 0.16 percent.

Clearly, the bound established by Heckman et al. and Lee is extreme. Accounting for the divergence in their results in any quantitative manner is problematic because of the major differences in their models. Yet, in determining the efficacy of tuition subsidies as a means for increasing school attainment, it clearly matters which result one adopts or how the results are weighted. This example again illustrates the importance of establishing some criteria for model validation and model selection.

Although the early human capital literature recognized the importance of understanding skill formation in children (Leibowitz, 1974), the human capital literature became focused, overly so in my view, on estimating the impact of years of school attainment on wages. The recent revival of interest in early childhood skill acquisition is an important development discussed below.

3.3 The Role of Theory in Drawing Inferences to External Populations

3.3.1 Project STAR

Views about the effectiveness of reducing class size on school performance were significantly affected by the results of the Tennessee Student/Teacher Achievement Ratio experiment, more conventionally known as Project STAR. That experiment was mandated by the Tennessee state legislature in 1985 and initiated in the 1985–1986 academic year. All elementary schools were invited to participate. A little more than one-third of the school districts, including about 180 schools, responded positively. Given the class-size requirements of the experiment and other factors, eventually 79 elementary schools located in 42 of the 141 districts were incorporated into the program. The experiment randomly assigned both students and teachers to three kindergarten classes: a small class, ranging from 13 to 17 students, and two regular size classes, with and without a teacher aid, ranging from 22 to 26 students. The experiment was conducted through the third grade, after which random assignment of students and teachers was discontinued.

There were initial studies encompassing the 4 years of the project and later studies that followed students into high school and beyond.

Krueger (1999) and Nye, Hedges, and Konstantopoulos (2000) provide two of the most rigorous studies of the initial 4 years of the program. Both of these studies conclude that class size has a statistically significant and quantitatively large effect on student performance as based on test scores. Krueger (1999) contrasts the experimental results to those obtained from nonexperimental studies as follows: "One well-designed experimental study should trump a phalanx of poorly controlled, imprecise observational studies based on uncertain statistical specifications."[57] Concerning the value of the experiment and its place in the literature, Nye, Hedges, and Konstantopoulos (2000) conclude that "The STAR study provides an important (and perhaps the strongest) piece of the converging evidence about the effectiveness of small class sizes in promoting achievement."[58]

In this section, I address a number of issues concerning these studies, and social experiments more generally, that relate to the role of theory in the development, analysis, and interpretation of social experiments. To fix ideas, let A_{ig} be a continuous measure of achievement for student i at the end of grade g, and S_{ig} be the class size assignment of that student, equal to 1 if the child is assigned to a small class and equal to 0 otherwise. As in the STAR experiment, the class size assignment is made at the beginning of the school year, and the achievement assessment is done at the end of the school year. Given randomization, the (demeaned) regression

$$A_{ig} = \alpha S_{ig} + u_{ig} \tag{3.36}$$

would provide an estimate of α, the average difference in achievement between students in small and large classes. The question is how to interpret α.

3.3.2 Is α a Production Function Parameter?

The title of Krueger's (1999) paper is "Experimental Estimates of Education Production Functions," and one of his stated goals is to "use the experimental results to interpret estimates from the literature based on observational data," where the literature being referenced is the education production function literature. Nye, Hedges, and Konstantopoulos (2000) also compare their estimates to those that come from the estimation of education production functions using observational data. They find that meta-analyses of nonexperimental estimates are similar to

those obtained from the experiment. If one is to take comfort in that observation, it is important that the two sources of estimates are measuring the same effect.

To address this issue, consider the following stylized formulation of the process of knowledge accumulation (achievement) as presented in Todd and Wolpin (2003). Define $t = 0$ as the time interval from conception to first attending school, $t = 1$ as the first year of school (kindergarten in the STAR experiment), and $t = 2$ as the second year of school. Let I_0 be the set of family-provided inputs to the child during period $t = 0$ that affect the child's accumulation of knowledge (including, for example, the mother's diet during the pregnancy, the degree of mental stimulation during infancy and early childhood, and others), and μ the child's endowed mental capacity determined at conception. Note that I_0 is meant to be inclusive of all of the determinants of the child's knowledge by the start of school, A_0. Thus, the achievement production function during period $t = 0$ is

$$A_0 = g_0(I_0, \mu).$$

The family is assumed to make (utility-maximizing) preschool input decisions based on their "permanent" resources, W, on the child's endowment and on the family's preference over the child's achievement, ε; that is, $I_0 = g_0(W, \mu, \varepsilon)$.

In addition to family inputs, the household also makes a decision about the potential school inputs the child will be exposed to through a location choice, which, like family inputs, depends on W, μ, and ε. However, at the time of the location decision, parents have incomplete information about the actual level of inputs applied to their child, S_1, for example, the quality of the assigned teacher, which may vary within the school. To simplify notation, the potential school input is given by a single summary statistic, $\bar{S}_1$.

Achievement at the end of the first school year, A_1, depends on family inputs during the preschool period and during the first period of school, I_0 and I_1, and on actual school inputs, S_1, as well as on the child's endowment, namely

$$A_1 = g_1(I_0, I_1, S_1, \mu).$$

Along with the technology for combining inputs to create achievement outcomes, there is a decision rule for both parents and schools that determines the levels of inputs. The family decision rules concerning direct family inputs and potential school inputs associated with their location decision are given by

$$I_1 = \phi_I(A_0, W, S_1 - \bar{S}_1, \mu, \varepsilon) \tag{3.37}$$

$$\bar{S}_1 = \phi_S(A_0, W, \mu, \varepsilon), \tag{3.38}$$

where it is important to note that family input decisions are assumed to be made subsequent to the actual realization of the school inputs applied to their children.[59]

Todd and Wolpin (2003) assume that the school chooses input levels for a particular child purposefully, taking into account the child's achievement level and the endowment, that is,

$$S_1 - \bar{S}_1 = \psi(A_0, \mu).$$

For example, a child who enters first grade able to read may receive different school inputs than a child who is not able to read. Thus, a simple comparison of student achievement based on differences in realized school inputs, $S_1 - \bar{S}_1$, would be contaminated by differences among students in A_0 and μ because of the sorting decision by schools. Of course, a comparison of student achievement across schools would similarly be contaminated because of the parental location decision as embodied in equation 3.38.

Within this simple framework, it is possible to consider different kinds of parameters of interest in empirical research. To focus on class size, one question of interest is the following:

1. *How would an exogenous change in class size, holding all other inputs constant, affect achievement?*

Knowledge of the production technology would suffice to answer this question. The goal of most observational studies has, in fact, been to uncover features of the production technology, given above by g_0 and g_1.

An alternative question of interest is the non–ceteris-paribus effect of changing school inputs, namely

2. *What would be the total effect of an exogenous change in class size on achievement, that is, not holding other inputs constant?*

Suppose first that the change in class size is unanticipated by parents, as in the Project STAR experiment. The total effect of this unanticipated change includes both the effect holding other inputs constant as well as any indirect effects that operate through changes in the levels of other inputs. In the simple model with only one school input, the total effect of an exogenous change in the school input (class size) in the first year, S_1, on achievement at the end of the year, A_1, is given by

$$\frac{dA_1}{dS_1} = \frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_1}\frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}. \tag{3.39}$$ [60]

Knowledge of the technology is not sufficient to answer this question because the production function provides $\frac{\partial g_1}{\partial S_1}$ and $\frac{\partial g_1}{\partial I_1}$, but not $\frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}$. To get the last term requires knowledge of the family input decision rule (equation 3.37), which could in principle be estimated from nonexperimental data. The Project STAR experiment provides an answer to the second question but not the first. Thus, α in equatiion 3.36 corresponds to the expected value of $\frac{dA_1}{dS_1}$ in equation 3.39 and is not a production function parameter. It is the "policy effect" of an *unanticipated* change in class size taking into account adaptations in inputs that arise because of the differences in class size, namely

$$E\left(\frac{dA_1}{dS_1}\right) = \int \left[\frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_1}\frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}\right] f(W, \mu, \varepsilon) \tag{3.40}$$

where $f(W, \mu, \varepsilon)$ is the joint density of W, μ, and ε.[61] Thus, the Project STAR experiment provides the answer to a question that may be of direct policy interest.

This average policy effect defined above can be either larger or smaller than the ceteris paribus effect. For example, families in Project STAR whose children were assigned to small classes may have spent less time teaching their children at home, that is, if school and family inputs are substitutes in producing achievement. In that case, the average policy effect measured by the experiment would be less than the ceteris paribus effect. Alternatively, the effects could reinforce each other, for example, if families were encouraged to apply more family inputs by their child's greater learning at school, that is, if school and family inputs are complements. Only under the assumption that families do not take into account changes in school inputs in choosing family input levels would the experimental estimate correspond to a production function parameter, and it is only in that case that a comparison of estimates obtained in these two ways is warranted.[62]

The model assumes that parents choose among locations that differ in the (permanent) level of school inputs provided, $\bar{S}_1$. The Project STAR experiment was interpreted in this analysis as an unanticipated change in school inputs from $\bar{S}_1$ to $S_1 = \bar{S}_1 + (S_1 - \bar{S}_1)$. The family then may have adapted its own inputs, I_1, to that change. Suppose, however, that all school districts permanently changed their level of school inputs and that, for simplicity, such a change did not alter the location decisions of parents. In this new regime, parents with newborn children face a level of school inputs, S_1, that differs from the one that parents of newborn children faced under the old regime. In that case, parents in the new regime may be expected to choose a different level of their own inputs in the preschool years, I_0, as well as in the school years. The average policy effect of the change to the new regime would be

$$E\left(\frac{dA_1}{dS_1}\right) = \int\left[\frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_0}\frac{\partial I_0}{\partial \bar{S}_1} + \frac{\partial g_1}{\partial I_1}\left(\frac{\partial I_1}{\partial \bar{S}_1} + \frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}\right)\right] f(W, \mu, \varepsilon).$$

The Project STAR experiment does not provide evidence on this policy change. In addition, if the policy change occurred in only one district, parents would sort differently across school districts leading to a change in the joint density, $f(W, \mu, \varepsilon)$, and thus to a change in the policy effect. Presumably, the actual policy that would be implemented is a permanent reduction in class size. The Project STAR experiment only identifies a subset of the elements that comprise this policy effect.

3.3.3 External Validity

There are two forms of external validity, extrapolating experimental results to different populations and extrapolating results to treatments not considered in the experiment. With respect to the first, as is apparent in equation 3.40, the results of the experiment can generally not be applied to a population that differs from the experimental sample in endowment distribution, $f(W, \mu, \varepsilon)$. Both production function parameters, $\frac{\partial g_1}{\partial S_1}$ and $\frac{\partial g_1}{\partial I_1}$, depend on μ, and $\frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}$ depends on W, μ, and ε. Remaining within the experimental paradigm requires direct experimentation on other populations. Similarly, extrapolation of the experimental results to other treatments, such as a permanent change in class size as discussed above, is also not possible without additional experimentation.

Alternatively, one could design the experiment to include a broader data collection effort, as was done in the PROGRESA program, which would enable the estimation of a structure allowing both types of extrapolations. Such an endeavor requires at least a broad outline of a theory, perhaps like the one above, to guide data collection.

3.3.4 The Estimation of the Achievement Production Function Using Observational Data

These considerations make me much less sanguine than the authors cited above about the singular importance of the Project STAR experiment as a guide to policy. Unfortunately, however, the interpretation of the results from observational research, like that from the experiment, also suffers from a lack of a theoretical underpinning. Todd and Wolpin (2003) provide a detailed discussion of the foundations and assumptions of alternative specifications of the achievement production function commonly employed in the literature.

Todd and Wolpin (2003) distinguish between two parallel literatures. The early childhood development (ECD) branch seeks to understand the role of parental inputs in producing cognitive, and more recently noncognitive, skills. The education production function (EPF) branch seeks to determine the role of school inputs in the determination of cognitive skills. These literatures are directly relevant to the finding that unobserved endowments at young adult ages are critical for the determination of schooling and career choices

In both the ECD and EPF literatures, there is a remarkable lack of consensus about which inputs increase skill acquisition and to what extent (Hanushek, 2003; Hedges, Laine, and Greenwald, 1994; Krueger, 2003). Empirical studies employ a wide variety of estimating equations. Hanushek (1996, 2003) summarizes in several meta-analyses the findings of EPF studies and concludes that although some papers find statistically significant effects, the overall pattern suggests that none of the measured schooling quality inputs is reliably associated with achievement. Krueger (2003) argues that Hanushek's sample of estimates is biased toward his conclusion. Although they disagree about the conclusions that may be drawn from meta-analytic studies, they do agree on the importance of taking into account model specification in combining evidence across studies. As they note, estimating equations are often adopted with little theoretical justification, making it difficult to assess their coherency.

Todd and Wolpin (2003) develop a coherent conceptual framework for interpreting the different estimating equations found in the literature based on the notion of a human capital production function (Ben-Porath, 1967). Skills (cognitive or noncognitive) are acquired through a production process that depends on the levels of current and past individual, family, and school inputs in combination with an individual's genetic endowment (determined at conception). The inputs are determined within a choice-theoretic decision model. These two tenets, that the production process is cumulative and that inputs are choices, have critical implications for the interpretation and estimation of the specifications observed in the ECD and EPF literatures.

Following the formalization in Todd and Wolpin (2003), let T_{ija} be a measure of skills, for example, a test score, for child i residing in household j at age a. Denote the vector of parent-supplied inputs at a given age as I_{ija}, school-supplied inputs as S_{ija}, and the vectors of their respective input histories up to age a as $I_{ij}(a)$ and $S_{ij}(a)$. Further, let a child's endowment be denoted as μ_{ij0}. With allowance for measurement error in test scores (for example, random fluctuation in test performance), denoted by ε_{ija}, the skill production function is given by

$$T_{ija} = T_a(I_{ij}(a), S_{ij}(a), \mu_{ij0}, \varepsilon_{ija}), \qquad (3.41)$$

where the notation implies that the production function itself may depend on the age of the child.

The empirical implementation of the model described by equation 3.41 has foundered on two main problems: (1) the genetic endowment of mental capacity is not observed; and (2) data sets on inputs are incomplete—in particular, they have incomplete input histories and/or missing inputs. Todd and Wolpin (2003) clarify the different assumptions required for consistent estimation for the common specifications found in the literature.

3.3.4.1 *The Contemporaneous Specification*

The contemporaneous specification includes only contemporaneous school and family inputs. To justify that specification requires that either (1) only contemporaneous inputs matter to the production of current achievement or (2) inputs are time invariant, so that current inputs capture the entire history of inputs. In addition, contemporaneous inputs must be independent of the unobserved endowment. The contemporaneous specification can be written as

$$T_{ija} = T_a(I_{ija}, S_{ija}) + \varepsilon'_{ija}, \qquad (3.42)$$

with ε'_{ija} an additive measurement error. Neither of the sets of assumptions given above is very plausible.[63] Many inputs of potential importance in the development of cognitive skills vary temporally and may vary for systematic reasons with the child's age (for example, maternal employment) or be specific to particular ages (for example, maternal alcohol use during pregnancy). Estimating the contemporaneous specification when the true technology is that given in equation 3.41 requires that the omitted past inputs and endowed ability are uncorrelated with contemporaneous inputs. Such assumptions are inconsistent with economic models of optimizing behavior. For example, economic models in which parents care about a child's development imply that the amount of resources allocated to the child will be responsive to the parent's perception of a child's ability (for example, Becker and Tomes, 1976). Thus, although the contemporaneous specification can be implemented with only limited data, strong assumptions are required to justify its application.

3.3.4.2 *Value-Added Specifications*

The lack of data on input histories and on endowments led researchers to adopt a value-added approach to estimating achievement production functions. In its most common form, this specification relates achievement to contemporaneous school and family input measures and lagged (baseline) achievement. Thus, it differs from the contemporaneous specification only in the inclusion of the baseline achievement measure, which is taken to be a sufficient statistic for unobserved input histories as well as the unobserved endowment. Evidence based on the value-added specification has generally been regarded as more convincing than that based on a contemporaneous specification (see, for example, Hanushek, 1996 and 2003, and Krueger, 2000). However, the value-added formulation also imposes strong assumptions on the underlying production technology and strong restrictions for consistent estimation.

To simplify the notation, let X denote the vector of family and school inputs and $X(a)$ their input histories up to age a; the value-added specification assumes that equation 3.41 can be written as

$$T_{ija} = T_a(X_{ija}, T_{a-1}(X_{ij}(a-1), \mu_{ij0}, \varepsilon_{ija}), \eta_{ija}), \qquad (3.43)$$

where it is assumed the baseline achievement measure was administered at $a - 1$. Value-added regression specifications usually treat the arguments in equation 3.43 as additively separable and the parameters as non–age-varying, which leads to the estimating equation:

$$T_{ija} = X_{ija}\alpha + \gamma T_{ij,a-1} + \eta_{ija}. \tag{3.44}$$

A more restrictive specification sometimes adopted in the literature sets the parameter on the lagged achievement test score to 1 ($\gamma = 1$) and rewrites equation 3.44 as

$$T_{ija} - T_{ij,a-1} = X_{ija}\alpha + \eta_{ija}, \tag{3.45}$$

which expresses the test score gain solely as a function of contemporaneous inputs.

To understand the restrictions the value-added formulation implies for the true technology function, consider the linear regression analogue of the true technology (equation 3.41), namely

$$T_{ija} = X_{ija}\alpha_1 + X_{ija-1}\alpha_2 + \ldots + X_{ij1}\alpha_a + \beta\mu_{ij0} + \varepsilon_{ija}. \tag{3.46}$$

This specification imposes the assumption that the production technology, as in equation 3.44, is non–age-varying at least over the ages used in implementing the model. Note that α_1 refers to the effect of the contemporaneous inputs, α_2 to the effect of the inputs one period lagged, and so on.

Subtracting $\gamma T_{ij,a-1}$ from both sides of equation 3.46 and collecting terms gives

$$T_{ija} = X_{ija}\alpha_1 + X_{ija-1}(\alpha_2 - \gamma\alpha_1) + \ldots + X_{ij1}(\alpha_a - \gamma\alpha_{a-1}) + \gamma T_{ij,a-1} \tag{3.47}$$

$$+(\beta_a - \gamma\beta_{a-1})\mu_{ij0} + \varepsilon_{ija} - \gamma\varepsilon_{ij,a-1}.$$

Two conditions must be met for equation 3.47 to reduce to equation 3.44: (1) input coefficients must be geometrically (presumably) declining with distance, as measured by age, from the achievement measurement, and the rate of decline must be the same for each input ($\alpha_k = \gamma\alpha_{k-1}$ for all k); and (2) the impact of the endowment must be geometrically declining at the same rate as input effects ($\beta_a = \gamma\beta_{a-1}$). For the value-added specification based on the gain in achievement (equation 3.45) to be appropriate, it is required, in lieu of (1) and (2), that (1′) the effect of each input must be independent of the age at which achievement was measured ($\alpha_k = \alpha_{k-1}$ for all k), and (2′) the effect of the endowment must likewise be independent of the achievement age ($\beta_a = \beta_{a-1}$).

With respect to estimation, if the restrictions that lead to the gain specification (equation 3.45) are valid, OLS estimation would provide consistent estimates of input effects. With $\gamma \neq 1$, as in equation 3.44, however, in order that OLS estimation be consistent, the measurement error in test scores must be serially correlated, and the degree of serial correlation must exactly match the rate of decay of input effects (that is, $\eta_{ija} = \varepsilon_{ija} - \gamma\varepsilon_{ij,a-1}$ is an *iid* shock). In the conventionally assumed case that measurement error (ε_{ija}) is *iid*, the estimate of γ will be downward biased given the necessarily positive correlation between the lagged test score measure, $T_{ij,a-1}$, and its measurement error, $\varepsilon_{ij,a-1}$.

If we drop the assumption that the impact of the endowment declines at the same rate as the decay in input effects, then the error in equation 3.44 would include the endowment, that is, assuming that $\beta_a - \gamma\beta_{a-1} = \beta'$ is a constant independent of age, which yields

$$T_{ija} = X_{ija}\alpha + \gamma T_{ij,a-1} + \beta'\mu_{ij0} + \eta_{ija}, \tag{3.48}$$

instead of equation 3.44. Estimation of equation 3.48 by OLS is problematic. As with the contemporaneous specification, one requirement for OLS to be consistent is that contemporaneous inputs and the unobserved endowment be orthogonal. However, even if that orthogonality condition were not violated, an OLS estimation of equation 3.48 would still be biased because baseline achievement must be correlated with the endowment. If the endogeneity is not taken into account, then the resulting bias will not only affect the estimate of γ but may be transmitted to the estimates of all the contemporaneous input effects.

In an optimizing behavioral model, we would expect family and school input choices to be affected by baseline achievement, particularly if, as in equation 3.48, baseline achievement has persistent effects on achievement in future time periods. It is possible to estimate equation 3.48 consistently if there exists a third (earlier) observation on achievement, along with the data on the input set $X_{ij,a-1}$, using a simple differencing procedure.

So far, it has been assumed that there are no missing contemporaneous inputs. However, suppose that ε_{ija} contains unmeasured contemporaneous inputs. And, to make the argument most strongly, suppose that the missing inputs are orthogonal to the included inputs. In this case, neither OLS estimation of equation 3.44 nor applying OLS to a differenced form of equation 3.48 will provide consistent estimates of input effects. Recall that the residual in equation 3.44 or 3.48 is a composite of the underlying current and baseline period residuals, $\eta_{ija} = \varepsilon_{ija}$

$-\gamma\varepsilon_{ij,a-1}$. When ε_{ija} contains omitted inputs, baseline achievement, $T_{ij,a-1}$, will likely be correlated with the composite residual for two reasons. First, baseline achievement, $T_{ij,a-1}$, must be correlated with its own contemporaneous omitted inputs contained in $\varepsilon_{ij,a-1}$. Second, to the extent that omitted inputs are subject to choice, optimizing behavior will create a correlation between the contemporaneous omitted inputs and baseline achievement. For example, parents may respond to realized poor achievement by increasing family inputs (such as parental time or providing tutors).

In the form most commonly adopted, the data requirements of a value-added specification are only slightly more demanding than those of the contemporaneous specification. One additional variable—a baseline test score—is all that is required. As noted, evidence based on value-added specifications is generally regarded as being superior to that based on contemporaneous specifications; however, the benefits of a value-added approach seem less clear when the potential for omitted data on inputs and endowments is taken into account.

3.3.4.3 *The Cumulative Specification*

Direct estimation of the cumulative specification given by equation 3.46 requires data on both contemporaneous and historical family and school inputs. There are several different approaches to directly estimating the cumulative specification, assuming that we have data on current and past inputs but do not observe endowments. Under the assumption that any omitted inputs and measurement errors in test scores are uncorrelated with included inputs, the remaining estimation problem is that observable inputs will generally be correlated with the unobserved child endowment. A number of methods have been employed in the literature to deal with endowment heterogeneity, although few studies have incorporated contemporaneous and past inputs. One method makes use of observations on multiple children within the same household or family (siblings or cousins), and the other makes use of multiple measurements for the same child at different ages.

In describing these methods and the properties of estimators based on them, we take the following version of equation 3.46 as the baseline specification:

$$T_{ija} = X_{ija}\alpha_1^a + X_{ija-1}\alpha_2^a + \ldots + X_{ij1}\alpha_a^a + \beta_a\mu_{ij0} + \varepsilon_{ij}(a), \tag{3.49}$$

in which input effects vary not only with the distance between the application of inputs and the achievement measure (as indicated by the parameter subscripts) but also with age itself (as indicated by the parameter superscripts). As the notation indicates, the residual includes all current and past unmeasured inputs.

A class of estimators used to account for permanent unobservable factors makes use of variation across observations within which the unobservable factor is assumed to be fixed. Two such "fixed effect" estimators that are prominent in this literature use variation that occurs either within families (across siblings) or within children (at different ages). It is useful in what follows to rewrite equation 3.49 for two different ages, age a and a':

$$\begin{aligned} T_{ija} &= X_{ija}\alpha_1^a + X_{ij,a-1}\alpha_2^a + \ldots + X_{ij,a'+1}\alpha_{a-a'}^a \\ &\quad + X_{ij,a'}\alpha_{a-a'+1}^a + X_{ij,a'-1}\alpha_{a-a'+2}^a + \ldots + X_{ij2}\alpha_{a-1}^a + X_{ij1}\alpha_a^a + \beta_a\mu_{ij0} + \varepsilon_{ij}(a) \end{aligned} \tag{3.50}$$

$$T_{ija'} = X_{ija'}\alpha_1^{a'} + X_{ij,a'-1}\alpha_2^{a'} + \ldots + X_{ij2}\alpha_{a'-1}^{a'} + X_{ij1}\alpha_{a'}^{a'} + \beta_{a'}\mu_{ij0} + \varepsilon_{ij}(a'),$$

where the parameter α_x^a indicates the effect of an input on an achievement measure taken at age a that is $x - 1$ periods removed from age a. Note that contemporaneous inputs are, as a matter of convention, zero periods removed from the age of the measurement, although they may be thought of as being applied between a and $a - 1$.

As an example, suppose we are looking at a child's achievement at age $a = 6$. In equation 3.50, the effect of reading to a child at age 3 on the child's achievement at age 6 may differ from the effect of reading to the child at age 2 ($\alpha_4^6 \neq \alpha_5^6$) because the inputs differ in their distance from the achievement measure. In addition, the effect of reading to the child at age 3 on achievement at age 5 may differ from its effects at age 6 ($\alpha_3^5 \neq \alpha_4^6$), again because it is more distant from the achievement measure, and also may differ from the effect of reading to the child at age 4 on reading achievement at age 6 ($\alpha_3^5 \neq \alpha_3^6$) because the efficacy of reading to a child on subsequent performance may depend on the child's age.

3.3.4.4 *Within-Family Estimators*

Within-family estimators exploit the fact that children of the same parents (or grandparents) have a common heritable component. In particular, assume that the child's endowment can be decomposed into a family-specific component and an orthogonal child-specific component, denoted as μ_0^f and μ_0^c. Thus, siblings have in common the family component but have their own individual-specific child components. Rewriting equation 3.49 to accommodate this modification yields

$$T_{ija} = X_{ija}^{a}\alpha_1 + X_{ija-1}^{a}\alpha_2 + \ldots + X_{ij1}^{a}\alpha_a + \beta_a\mu_{ij0}^{f} + \beta_a\mu_{ij0}^{c} + \varepsilon_{ija}. \tag{3.51}$$

Now, suppose that longitudinal household data on achievement test scores and on current and past inputs are available on multiple siblings, and consider two types of data. In the first, data are available on siblings at the same age. Notice that unless the siblings are twins, the calendar time at which achievement measures are obtained must differ. In the second, data are available on siblings in the same calendar year, which generally means that they will differ in age.

Data Collected on Siblings at the Same Age

The within-family estimator is based on sibling differences, which eliminates the family-specific component of the endowment but not the child-specific component. Consider the estimator in the case of two siblings, denoted by i and i', observed at the same age a. Differencing equation 3.51 yields

$$T_{ija} - T_{i'ja} = (X_{ija} - X_{i'ja})\alpha_1^{a} + \ldots + (X_{ij1} - X_{i'j1})\alpha_a^{a} + [\beta_a(\mu_{ij0}^{c} - \mu_{i'j0}^{c}) + \varepsilon_{ij}(a) - \varepsilon_{i'j}(a)]. \tag{3.52}$$

In estimation, the residual term will include all the terms within the square brackets. Consistent estimation of input effects, therefore, requires that inputs associated with any child not respond either to own or to sibling child-specific endowment components.

Furthermore, given that achievement is measured for each sibling at the same age, the older child's achievement observation (say child i) will have occurred at a calendar time prior to the younger sibling's observation. Thus, the older sibling's achievement outcome was known at the time input decisions for the younger child were made, at the ages of the younger child between the older and younger child's achievement observations. Thus, consistent estimation by OLS requires that input choices be unresponsive to prior own and sibling outcomes; otherwise, the realizations of $\varepsilon_{ij}(a)$ will affect some of the inputs to sibling i'.

In essence, intrahousehold allocation decisions must be made ignoring child-specific endowments and prior outcomes of all the children in the household. The within-child estimator considered below relaxes this assumption.

Data Collected on Siblings at the Same Calendar Time

The within-family estimator based on siblings of different ages can be viewed as a special case of the within-child estimator based on test scores of the same child measured at different ages.

3.3.4.5 *Within-Child Estimators*

Within-child estimators are feasible when there are multiple observations on achievement outcomes and on inputs for a given child at different ages. Consider differencing the achievement technology at the two ages shown in equation 3.50. This procedure, after grouping inputs applied at the same age, yields

$$\begin{aligned} T_{ija} - T_{ija'} &= X_{ija}\alpha_1^{a} + X_{ija-1}\alpha_2^{a} + \ldots X_{ij,a'+1}\alpha_{a-a'}^{a} + X_{ij,a'}[\alpha_{a-a'+1}^{a} - \alpha_1^{a'}] \\ &+ X_{ij,a'-1}[\alpha_{a-a'+2}^{a} - \alpha_2^{a'}] + \ldots + X_{ij2}[\alpha_{a-1}^{a} - \alpha_{a'-1}^{a'}] \\ &+ X_{ij1}[\alpha_a^{a} - \alpha_{a'}^{a'}] + [\beta_a - \beta_{a'}]\mu_{ij0} + \varepsilon_{ij}(a) - \varepsilon_{ij}(a'). \end{aligned} \tag{3.53}$$

Without any restrictions on the relationship among parameters, the within-child estimator recovers (1) age-specific input effects for the inputs that are applied between the two age observations and (2) differences in parameters that depend on both age and time from the achievement measure for inputs applied contemporaneously or prior to the earlier achievement observation.

The parameters of equation 3.53 can be consistently estimated under the following assumptions. (1) The impact of the endowment on achievement must be independent of age ($\beta_a = \beta_{a'}$), in which case the differencing eliminates the endowment. In that case, orthogonality between input choices and endowments need not be assumed. Second, similar to the within-family estimator based on sibling observations at the same age, because any prior achievement outcome is known when later input decisions are made, it is necessary to assume that (2) later input choices are invariant to prior own achievement outcomes.

The difference between the within-child estimator and the within-family estimator based on observations for siblings that differ in age is that in the latter only the family component of the endowment disappears from equation 3.53. Thus, consistency of that estimator requires the same behavioral assumption with respect to intrafamily allocations as did the within-family estimator based on sibling observations at the same age.

Suppose, however, that the researcher is willing to impose the restriction that input effects are age invariant; that is, $\alpha_x^a = \alpha_x^{a'}$ as is often assumed in the application of fixed-effects estimators. Then, equation 3.53 can be rewritten as

$$T_{ija} - T_{ija'} = [X_{ija} - X_{ija'}]\alpha_1 + \ldots + [X_{ija'} - X_{ij1}]\alpha_{a'} \quad (3.54)$$

$$+ X_{ij,a-a'-1}\alpha_{a'+1} + \ldots + X_{ij1}\alpha_a + \varepsilon_{ij}(a) - \varepsilon_{ij}(a')$$

that would identify all of the parameters of the production function.

3.3.4.6 *Instrumental Variables Within-Child Estimators*

It is possible to relax the assumption that input choices do not respond to previous realizations of achievement. If the residuals in equation 3.54 consist only of unforeseen factors (for example, randomly being ill or randomly drawing a bad teacher), and if the impact of these factors on achievement has limited persistence, then input levels prior to the earliest achievement observation (a') can serve as instrumental variables in estimating equation 3.53 or 3.54. For example, if the achievement tests are taken at ages 8 and 5, then perhaps the set of inputs at ages earlier than age 3 might satisfy this requirement. However, even if that were the case, there are more parameters in equation 3.54 than instruments—at least as many as the number of measured inputs—so identification cannot be achieved with these orthogonality conditions alone. However, it is also the case that inputs associated with the child's siblings applied at a time sufficiently prior to the earliest observation used to implement the within-child estimator could also be used as instrumental variables. Thus, we can still possibly estimate equation 3.53 or 3.54 consistently using own prior and sibling inputs as instrumental variables.

Some researchers have used cross-sectional variation in prices and other location-specific characteristics as instrumental variables to estimate human capital production functions (Rosenzweig and Schultz, 1983). One potential problem with that approach if applied directly to the baseline specification (equation 3.49) is that state-level variation in the endowments of its residents will plausibly be correlated with the demand for different market or politically supplied services and products, for example, school inputs. If applied to the within-sibling specification given by equation 3.52, those instruments will be valid to the extent that location decisions are independent of child-specific (though not necessarily family-specific) endowments and are also independent of the actual achievement realizations of the siblings. Applying the same approach to within-child specification (equation 3.53 or 3.54), given that the sample includes children who have lived in different locations, avoids the biases from omitting child-specific endowments but would still be subject to the potential problem that families may change locations to find more suitable schools for their children based on their prior achievement.

Finally, none of the IV approaches are valid if omitted inputs are not orthogonal to the included ones. Omitted inputs that reflect choices are as likely to be correlated with an instrumental variable as are included inputs. Thus, any instrument that has power will also not be valid. It is therefore important to have data that contain a large set of inputs spanning both family and school domains.

3.3.4.7 *Choosing a Specification*

As noted, a possible reason for the range of findings in the ECD and EPF literatures is the diversity of specifications of the production function. This diversity is partly the result of data limitations, partly the result of a lack of a theoretical foundation and also partly the result of a lack of consensus on the best approach to model selection. With respect to the latter, in a recent paper, Todd and Wolpin (2007) explore the use of cross-validation criteria for the purpose of model selection.[64] Cross-validation methods choose the best-performing specification according to an out-of-sample forecasting criterion. They provide a useful alternative to specification testing when specifications are nonnested. The data set used by Todd and Wolpin (2007), the children born to the women in the NLSY79 (NLSY79-CS), contains an extensive history of family inputs starting from birth and cognitive achievement tests starting from early childhood, thus permitting the estimation of the achievement production function under a number of alternative assumptions as described above. In addition to home inputs, the NLSY79-CS also contains a measure of maternal "ability," the AFQT test.

Todd and Wolpin (2007) consider the following set of alternative specifications: (1) contemporaneous; (2) cumulative; (3) within-family fixed effect (siblings at same age); (4) within-child fixed effect; (5) value added; and (6) value added with lagged inputs.[65] They find strong support for the notion that skills are acquired through a cumulative process; indeed, the value-added with lagged inputs specification does best in the cross-validation exercise.

3.3.5 The Dynamic Factor-Model Approach

Cunha and Heckman (2008) begin from the observation that the most comprehensive data sets used in the study of early childhood development tend to have multiple measures of home inputs. The NLSY79-CS, for example, administers a battery of questions about the home environment, the Home Observation Measurement of the Environment—Short Form (HOME-SF). The HOME-SF consists of four instruments that depend on the child's age: ages 0–2, 3–5, 6–9, and 10 and above. Responses to the individual items and a scale based on a summation of responses that form a total raw score are available to researchers. There are about twenty items in any age group. To maintain estimation precision, researchers either must choose a (small) subset of the items, use the existing scale, or create their own scale. Cunha and Heckman (2008) suggest that a natural way to use these data that avoids an arbitrary choice of items or scales is to place the estimation problem within a dynamic factor-model framework.

In addition to the estimation approach, Cunha and Heckman (2008) depart from the ECD literature by considering the joint development of cognitive (C) and noncognitive skills (*N*) in early childhood. In contrast to the assumptions made in the conventional approach to identification as described above, their factor-model approach achieves identification through the use of covariance restrictions. The production functions of those skills take the form of value-added specifications of the form:

$$A_a^C = \gamma_0^C + \gamma_1^C A_{a-1}^C + \gamma_2^C A_{a-1}^N + \gamma_3^C X_a^C + \eta_a^C, \tag{3.55}$$

$$A_a^N = \gamma_0^N + \gamma_1^N A_{a-1}^N + \gamma_2^N A_{a-1}^C + \gamma_3^N X_a^N + \eta_a^N.$$

Note that the production function of each skill depends on the previous level of both skills and on skill-specific inputs (which may be the same).[66] Skills are latent factors and are unobservable. Test scores provide measurements of skills with error. Similarly, inputs are considered as latent factors and are also unobservable. Individual items on the HOME-SF would be measures, with error, of the lower-dimensional latent inputs.

As Cunha and Heckman (2008) show, identification is achieved through covariance restrictions on the measurement error equations (and sometimes also depending on assumptions with instruments). Given longitudinal data, it is not necessary to restrict the measurement errors to be serially uncorrelated. Distributional assumptions are also

not necessary for identification. It is also shown that the assumption of orthogonality between observed (latent) and unobserved (latent) inputs can be relaxed in the case that unobserved inputs do not vary over time. An additional measurement equation relating adult outcomes (earnings and schooling) to the level of skills achieved on reaching adulthood gives substantive content to input productivities and test scores.[67]

Estimation is carried out using data from the NLSY-CS. The results of the analysis are provocative and important. They find that (1) both cognitive and noncognitive skills exhibit strong persistence at all child ages (measured in standardized test score units, γ_1^C is generally about 0.95 and γ_1^N about 0.90); (2) cognitive skills 2 years prior affect noncognitive skill formation, although the effect is small ($\gamma_2^N = .03$ in standardized test score units), but prior noncognitive skills do not have a statistically significant effect on cognitive skill formation; (3) parental inputs affect both cognitive and noncognitive skill formation; (4) the effect of parental inputs are greater at younger child ages in the case of cognitive skill formation but greater at older child ages in the case of noncognitive skill formation; (5) estimates of input effects are sensitive to correcting for measurement error; and (6) the HOME scale, which weights responses equally, would not be a good representation of the latent HOME input.

There is much to recommend the factor-model approach. Given historical data on both inputs and outcomes, as in the NLSY79-CS, it would be desirable to estimate a cumulative specification using information back to conception. The NLSY79-CS includes data on child motor and social development and personality development outcomes starting from birth and data on inputs starting from conception (for example, smoking during pregnancy). The most significant shortcoming is the lack of information on school inputs. Todd and Wolpin (2007) merge school input data at the county and state level with the NLSY79-CS data, but that is clearly a gross proxy for the child-specific school inputs. The Early Childhood Longitudinal Studies (ECLS-B and ECLS-K) are both representative samples that could be combined within a dynamic factor model to study child cognitive and noncognitive skill formation from birth through the eighth grade.

3.3.6 Limitation of the Production Function Literature

We have now come full circle. Unlike the goal of the Project STAR experiment, the estimation of the cognitive (or noncognitive) achievement production function cannot tell us what the effect would be of any particular policy initiative to foster the development of skills during childhood. To do that requires estimating a model of the behavior that governs the choice of inputs. Developing an estimable model of family investments in skill acquisition, including schooling quality, within a dynamic setting and incorporating multiple children (and perhaps the choice of fertility), is an ambitious undertaking. To date, there have been only a few such attempts to incorporate the estimation of the achievement production function in a household dynamic optimization problem. I choose two examples to illustrate.

Bernal (2008) develops and structurally estimates a DCDP model of a woman's employment and child care decisions that directly affect a child's cognitive achievement.[68] Bernal adopts a cumulative specification for the cognitive achievement production function that includes an unobserved child endowment. Women have heterogeneous labor market skills.[69] The model is estimated by maximum likelihood using data from the NLSY79-CS. Bernal predicts the effects of policy interventions that include child care subsidies and maternity leave entitlements, which she finds, on average, have adverse affects on pre-school-age children's cognitive outcomes.

Tartari (2007) developed and structurally estimated a DCDP model of a married couple's decision about the investments they make in their children's "quality" and the decision of whether to divorce.[70] Child quality is a public good in the household and is produced by parental time and goods inputs and by the degree of conflict within the marriage, the latter also being a choice. The achievement production function is modeled as an extended value-added specification that includes a family-specific unobserved permanent component. Divorce is an endogenous outcome of the decision model, along with fertility, the wife's labor supply, and the level of child support if the couple divorces (and has children). Divorce does not directly enter into the child quality production function; instead, divorce indirectly affects child quality by reducing the amount of joint time the parents can spend with the child and, presumably, expenditures on the child. On the other hand, divorce shields children from the potentially negative consequences of exposure to parental conflict within the home. Tartari estimates the model by indirect inference with data from the NLSY79-CS. The model estimates are used to perform several counterfactual exercises. In one, the divorce option is removed, which addresses the question of whether children of divorced parents would be better or worse off had the marriage continued. Tartari finds that the cognitive achievement of children is increased because the positive effect of the increase in parental investment time and investment goods exceeds the negative effect of exposure to conflict.

An implicit assumption in these models is that the skill acquisition technology is known to parents. Clearly, that assumption is false; it cannot be the case that researchers are attempting to determine what everyone else already knows. The question, of course, is how the assumption of perfect knowledge affects estimates of policy effects. The problem in introducing imperfect knowledge is that a model in which parents differ in their beliefs about the productivity of child investment inputs is observationally equivalent to one in which there is variation across parents in the marginal product of inputs and/or in which parents have direct preferences over inputs that vary in the population. To distinguish among these alternatives as explanations for the variation in input choices of parents and the achievement outcomes of children would require collecting information on beliefs or preferences (or both).

Recently, there has been (renewed) interest in the collection of data on subjective beliefs (Manski, 1990, 2004). Data have been collected on expectations about choices, such as school attainment and retirement age, and expectations about future states of the world, such as realizations of future income. Recently, Cunha (2012) has taken a first serious step towards designing and administering questions about the productivities of alternative inputs in producing cognitive or noncognitive skills in children. Given the early nature of this work, whether subjective beliefs in this domain can be elicited in a meaningful way is unclear; however, the payoff is potentially large.

The power of combining data on subjective beliefs with structural estimation is nicely illustrated by Delavande (2008). Delavande (2008) designed and administered a survey that elicited subjective beliefs about the efficacy of alternative contraceptive methods in preventing pregnancies, beliefs about their side effects, their ability to prevent sexually transmitted diseases, the propensity of partners to disapprove, their expected monetary cost, and their mode of administration. Combining the subjective data with actual contraceptive choices, Delavande estimates a static random utility model from which she recovers underlying preferences for alternative contraceptive methods. She shows that assuming instead homogeneous beliefs, based on contraceptive failure rates computed from national statistics, would yield very different estimates for preferences.