4 A Brief Conclusion

The proper role of economic theory in empirical research has been and remains a controversial issue. In these two lectures, I have tried to illustrate the value for inference of connecting data and theory. In the first lecture (chapter 2), I showed how theory can be used to perform *ex ante* policy evaluation in the most challenging circumstance in which there is no direct policy variation. Theory was shown to be critical in providing a structure within which existing variation that provides an analogue to direct policy variation can be exploited. In the case of a wage tax, wage variation was the analogue to introducing a tax; and in the school attendance subsidy program, less obviously, joint variation in the child wage and parental income was the analogue to introducing a subsidy. In both cases, I considered assumptions under which nonparametric estimation was feasible. I showed how parametric assumptions permit the use of richer behavioral models in performing *ex ante* policy evaluation and concomitantly expand the types of policies that can be considered.

The main aim of the second lecture (chapter 3) was to show how the quality of inference in existing literatures could be improved by explicitly connecting theory and data analysis. Two of the examples, one on estimating the impact of unemployment benefits on unemployment duration and the other on estimating the impact of welfare benefits on the labor market and demographic choices of women, showed the importance of that connection in developing econometric specifications, for example, with respect to what variables should or—just as importantly—should not be included. The third example, using natural experiments to recover the relationship between school attainment and wages, illustrated how theory clarifies assumptions under which an interpretation is valid and also enables disparate results to be reconciled. The last example, the use of a field experiment to estimate the

impact of class size on student learning, showed how theory can shed light on the generalizability of experimental effects to other populations. Although each of the examples was used to focus on a specific methodological issue, all of them are relevant for understanding the role of theory in interpretation, specification, and external validity.

What I hope my lectures have accomplished is to convince empirical researchers (and journal editors and referees) that sharper inferences do indeed result from drawing explicit connections to coherent and relevant theory. The empirical approach, be it structural, quasi-structural, quasi-experimental, or experimental, is of secondary importance. Theory provides the only way to fully appreciate the behavioral assumptions that underlie inference from data.

Notes

Chapter 1

1. Tjalling C. Koopmans (1947), p. 172. Italics added.

2. I am obviously not referring to the large set of experiments, those conducted both in the laboratory and in the field, the goal of which has been to test the validity of predictions obtained from theory. Those experiments require a strict adherence to theory in their design and implementation.

3. This is the view taken by Angrist and Krueger (1999).

4. Actually, there was one exception in which a behavioral model was relegated to the appendix. The issue contained papers based on quasi-experimental and experimental (randomized controlled trial) approaches.

5. Jacob Marschak (1953), p. 26.

Chapter 2

1. I address this issue in the next chapter. There is also the additional assumption that the variation induced by the "experiment" is independent of unobserved determinants of the outome—the question of "what's in the error." The distinction between randomness and exogeneity is drawn in Rosenzweig and Wolpin (2000).

2. See, however, Bourgignon and Spadaro (2006) for a discussion of microsimulation methods for *ex ante* evaluation.

3. This section draws from and extends the arguments in Todd and Wolpin (2008).

4. In contrast to a parametric approach (see Heckman, 1974), no assumption is made about the way unobserved preferences affect the marginal rate of substitution. Usually, a monotonicity assumption is made in order to obtain a unique cutoff value of ε, given observables, that determines the choice of zero versus positive hours. No such assumption is made here in part because, as shown below, the goal of *ex ante* policy evaluation does not require the estimation of the utility function or even the full estimation of the hours of work function.

5. The analysis applies equally well to a wage subsidy, which is a negative tax ($\tau < 0$).

6. When viewed as a function of the tax rate, τ, as described by the η function, the hours of work function would not be considered structural with respect to the tax policy as

it is not invariant to changes in the tax rate. However, viewed as a function of the net wage, $(1 - \tau)w$, as described by the φ function, the hours of work function would be considered structural as it affects the value of a variable in the function and not the function itself.

7. In terms of notation, throughout this chapter, when it is clarifying, the optimal level of a choice variable, h in the current example, in the absence of the policy is denoted with an asterisk (as in equation 2.3). The same choice variable at its optimum in the presence of the policy is denoted with a double asterisk (as in equation 2.5). Choice functions that are invariant to the policy (and can thus be given a structural interpretation) are denoted by $\varphi(\cdot)$, and those that are not invariant (and thus cannot be given a structural interpretation) are denoted by $\eta(\cdot)$.

8. Note, however, that the effect of y or X on expected hours of work cannot be obtained from equation 2.6 without the additional independence assumption that $f(\varepsilon \mid y, X) = f(\varepsilon)$.

9. Ichimura and Taber (2000) also draw an analogy between their proposed method of nonparametrically recovering policy impacts and matching.

10. This observation is just a more general statement of what Heckman (2000) calls Marschak's maxim, namely, that *ex ante* policy evaluation may not require recovery of the entire structure, in this case the φ function.

11. See Todd and Wolpin (2008) for a discussion of the relationship of matching in this context to the more conventional matching estimator (e.g., Rosenbaum and Rubin, 1983).

12. Analogous to the use of nonparametric estimation of the conditional mean hours function, if the support of w and the support of $\tilde{w}$ do not overlap, the average impact based on the matched samples may differ from the population impact. See Ichimura and Taber (2000) for more discussion on this point.

13. See, for example, Haerdle and Linton (1994) and Heckman, Ichimura, and Todd (1997).

14. For the same reason, it would also not be possible to estimate the effect of introducing the tax on the participation rate.

15. Although our interest is only in recovering the wage offer function, estimation of the hours function would be problematic with only a wage error (that is, without the preference unobservable). Conditional on the wage, which is observed for participants, hours of work would be deterministic. Thus, for any given set of parameters, the likelihood contribution for at least one participant, that is, the joint density of hours and the wage, would almost surely be zero. To avoid degeneracy in the likelihood function would require an additional error, for example, measurement error in reported hours of work or in the reported wage.

16. The problem that arises from including the preference unobservable, ε, is that the marginal rate of substitution is now a function also of ε, that is, given by $M(y; X, \varepsilon)$. In that case, after integrating over ε as in equation 2.12, H is no longer a function of the participation probability.

17. It is also assumed that ξ is distributed independently of y and X.

18. One can also obtain an estimate of the measurement error variance given a distributional assumption, for example, if the measurement error in equation 2.10 is additive and normal.

19. Note that because the hours function depends parametrically on the fixed cost (assuming there are no available measures), the hours of work function would not be considered structural, that is, policy invariant, in the presence of a policy that provides a subsidy for the fixed cost.

20. See also Blomquist, Eklof, and Newey (2001).

21. The β would have to be bounded from above in order that the tax is not turned into a subsidy along any segment.

22. This result also holds in the Blomquist and Newey approach. Knowledge of the expected hours function given in equation 2.22 would still not allow a prediction of the effect of changing individual or subsets of the tax rates.

23. Blomquist and Newey (1997, 2002) do not make such extrapolations.

24. Nonconvexities arise, for example, in public welfare programs such as AFDC or the EITC (see Blundell and MaCurdy, 1999).

25. Note that the school attendance function could also be written, less revealingly, as a function of $y + \tau$ and $y + w$.

26. I return to this point in the later discussion of parametric models.

27. For convenience, I rule out the case that there are multiple children in the household eligible to attend a given grade.

28. If subsidy levels also depend on other child characteristics such as gender, then the configuration of children in the household would also condition on those characteristics.

29. Without the additional unobservable preference shifter, the number of choices will be more than the number of errors.

30. If family income were proportional to the number of children so that the budget constraint contained the term $(y + \tau)n$ instead of $y + \tau n$, the demand function with and without the subsidy would be the same. However, this restriction would be a rather unusual one to impose on the model.

31. The claim in Todd and Wolpin (2008) to the contrary is incorrect.

32. If, instead, households can freely borrow and lend, the budget constraint (with zero rate of interest) would be

$$\sum_{t=1}^{T} C_t = \sum_{t=1}^{T} y_t + \sum_{t=1}^{T} w_t(1-s_t) + \sum_{t=1}^{T} \tau_t s_t$$
$$= \sum_{t=1}^{T} (y_t + \tau_t) + \sum_{t=1}^{T} (w_t - \tau_t)(1-s_t). \qquad (2.57)$$

The school attendance demand function would then depend on $\sum_{t=1}^{T}(y_t + \tau_t)$, and matching would be done on that basis rather than the basis of the period-by-period income profile.

33. It is of course necessary to maintain the assumption that ε is orthogonal to income and the child wage in order for the matching estimator to remain valid.

34. If income is serially correlated, that is, if $f(y_2 \mid y_1) \neq f(y_2)$, then s_1 will also depend on y_1 (as well as $y_1 + \tau$). It is obviously not possible to match on both y_1 and $y_1 + \tau$ simultaneously for given τ.

35. The two unobservable preference shifters, ε_s and ε_c, affect the marginal utility of school attendance and child leisure, respectively.

36. It is also not possible to infer the effect of the program based on the nonparametric estimation of the school attendance demand function, which in this case is not policy invariant, that is, not structural.

37. The parameters α and/or β can be made parametric functions of observable preference shifters, X, without affecting the analysis.

38. Although the distributional assumption is unnecessary in this case (clearly, the matching estimator does not require it), it is useful to establish results with the distributional assumption because it will be necessary when dynamic models are considered.

39. The * and ** notation, indicating optimality without and with the subsidy, is not useful in the parametric case and so is dropped.

40. If the linear consumption term entered the utility function as γC, then either $\gamma = 1$ or $\sigma_\varepsilon = 1$ could serve as a normalization. Because the subsidy is in monetary units, the normalization of the utility function facilitates interpretation.

41. Note that if there is no income effect on school attendance, $\beta = 0$, then the nonparametric matching estimator can be implemented without matching on household income. Moreover, in that case, the effect of an increase in the subsidy amount is identical to the effect of an equal reduction in the child wage.

42. Keane (1992) has shown in Monte Carlo experiments that the model without exclusion restrictions may be difficult to identify in practice. Exclusion restrictions in this model could take the form of variables that appear in α, for example, distance from school, but not in β. An alternative restriction would be to set the covariance of the unobserved preference shifters to zero, that is, $\sigma_{\varepsilon v} = 0$.

43. The additive error (ξ_{it}) is convenient in calculating choice probabilities and is maintained for illustrative purposes. However, the additive structure is fragile. It is lost, for example, if the wage function, as is conventional, takes on the semilog form or if the utility function is nonlinear in consumption (e.g., CRRA or logarithmic).

44. Although I refer to $\Pr(s_{it} = 1, w_{it} \mid \Omega_{it})$ as a probability, it is actually a mixed probability for s_{it} and a density for w_{it}. The Jacobian of the transformation from the wage density to the wage error density is 1.

45. If the structure does not yield an additive (composite) error, the latent variable function becomes $v_t^*(\Omega_{it}, \eta_{it}, \varepsilon_{it})$. Calculating the joint regions of η_{it}, ε_{it} that determine the probabilities that enter the likelihood function and that are used to calculate the $E\max(\Omega_{it})$ function must, in that case, be done numerically.

46. Although the example assumes that the heterogeneity is in the intercept terms, more general specifications in which slope coefficients are heterogeneous are also possible.

47. The existence of unobserved heterogeneity creates an initial-conditions problem (see Heckman, 1981, and Keane and Wolpin, 1997, for an application to a DCDP model).

48. Heckman and Navarro (2007) develop a semiparametric estimator for a specific application.

49. The program also provided a transfer contingent on visiting a health clinic as well as nutritional supplements for children under the age of 2. The school attendance component of the tranfer amounted to more than 75 percent of the total.

50. A peso at that time was equivalent to about one-tenth of a U.S. dollar.

51. Todd and Wolpin (2008) use a biweight kernel function:

$K(s) = (15/16)(s^2 - 1)^2$ if $|s| \le 1$
$= 0$ else,

which satisfies the standard assumptions $\int K(s)ds = 1$, $\int K(s)sds = 0$, and $\int K(s)s^2ds < \infty$. Asymptotic consistency of this estimator requires that the smoothing parameters satisfy $n\lambda_n^w\lambda_n^y \to \infty$, $\lambda_n^w \to 0$ and $\lambda_n^y \to 0$ as $n \to \infty$.

52. This procedure is similar to that used in Heckman, Ichimura, and Todd (1997).

53. The exact matching estimator, which would have to match households with the same distribution of children by grade and gender, was not feasible given sample sizes.

54. Sample sizes of single ages were too small to be reliable for nonparametric estimation. The bandwidth was set equal to 200.

55. Azevedo, Bouillon, and Yanez-Pagan (2009) report a p-value of 0.604. The percentage of overlapping support was 62.2.

56. Infecundity is assumed to occur at age 45 and to be known with certainty at the beginning of the marriage.

57. To demonstrate how the subsidy effect can be estimated from the control sample ($\tau = 0$), Todd and Wolpin (2006) used an example with $\beta = 0$, which may have led to some confusion as to the necessity of that assumption as well as to whether Todd and Wolpin's (2006) specification actually assumed the absence of an income effect.

58. As with most DCDP models with a large choice set and state space, the Todd and Wolpin (2006) model is solved numerically using an approximation method in which the $E\max(\Omega_{it})$ functions are evaluated at a subset of the state points and interpolation is used to evaluate $E\max(\Omega_{it})$ at other values of the state space. This approach requires that the $E\max(\Omega_{it})$ interpolating functions be specified parametrically, for example, by a regression function in the state space elements. This approach was originally proposed by Bellman, Kalaba, and Kotkin (1963) and extended to models generally of interest to economists by Keane and Wolpin (1994).

59. A nonparametric estimator of the type probabilities would be exact.

60. The parameters of the approximation, although treated in estimation as unrestricted, are actually themselves functions of the structural parameters. The estimation method does not take into account these restrictions and is thus not fully efficient even if the approximation is close to exact.

61. Todd and Wolpin (2006) follow this strategy as opposed to allowing for child-specific wage shocks to avoid having to integrate over all of the child shocks in solving the dynamic programming problem. As they note, the problem of degeneracy exists more generally, namely that with family level preference shocks, some choices may not be generated by the model. However, Todd and Wolpin (2006) placed a number of restrictions on the choice set, for example, assuming that no older child can attend school without a younger child also attending school, which reduces the likelihood of this event but does not eliminate it necessarily. Estimation is feasible when such events occur because the likelihood is smoothed over zero likelihood events. After estimating the model, Todd and Wolpin (2006) verified that simulations of the model could generate all of the outcomes that were observed in the data.

62. There were 1,885 households in the treatment villages, 5,561 children ever born, and 2,694 children aged 6–15.

63. The utility function contained about 90 parameters.

64. This argument is not meant to imply that the model has not been overfit. Indeed, there is some indication of overfitting given that a significant number of the parameters are imprecisely estimated. Overfitting arises in DCDP models when parameters are added that, although intended by the researcher to capture population characteristics, actually capture idiosyncratic characteristics of the sample.

65. I follow Attanasio, Meghir, and Santiago (2011) in not normalizing utility, that is, in not setting $\theta_w = 1$.

66. In the specification given by equation 2.56, some components of Y^w and Y^s would be related directly to preferences and other components to parental income, y, which would be estimable using data on y. α would arise from a constant term in the coefficient vector for X_{it}. The identification of β would require that the family income and utility components of Y^w and Y^s be specified, which is not done by Attanasio, Meghir, and Santiago (2011).

67. Attanasio, Meghir, and Santiago (2011) assert that their specification corresponds to a collective model in which the husband and wife have control over different sources of income. The attendance subsidy, for example, is paid directly to the wife. They do not show that the identification of a collective model, if actually specified, would require using the treatment data.

68. Notice that one can write the school attendance decision (equation 2.56) as:

$s = 1$ iff $\varepsilon \geq -\alpha + \theta_w(w - \tau) - \beta(y + \tau) - \lambda\tau$,
$= 0$ otherwise.

Thus, if $\lambda > 0$, because τ enters as a separate argument, it is not possible to use the matching estimator based on control villages.

69. To fully characterize equilibrium effects of the PROGRESA program, one would need to account for changes in adult wages and labor supply separately from child wages and labor supply, given that parental income affects the school attendance decision.

70. To explicitly incorporate equilibrium effects using only the control sample (as in Todd and Wolpin, 2006) would require an exogenous village-level shifter of child labor supply. One possibility would be the existence of a secondary school in the village.

71. In calculating the failure rate, it is assumed that all children who did not attend a particular grade level would have failed had they attended. However, their model predicts nonattendance for other reasons, for example, a particular preference shock. Todd and Wolpin estimate the failure probability function jointly with the attendance decision and allow it to depend on household type.

72. The alternative, as followed by Todd and Wolpin (2006), is to estimate the wage offer function and failure probability function jointly with the school attendance decision. In estimating the wage function, the assumption in Attanasio, Meghir, and Santiago (2011) that the schooling choice is governed by a probit is not consistent with the assumptions of the decision model, which would contain a composite logistic and normal error.

73. Under that assumption, the expected wage needed to compute the expected utility of working would be $\exp(\mu_w)$ times the exponential of 0.5 times the variance of the log wage offer error. Alternatively, if it were assumed that the wage error arose only from measurement error, then no integration is necessary, and one could simply substitute the exponential of the predicted wage offer for the wage in the utility of work and a selectivity correction would be unnecessary. Attanasio, Meghir, and Santiago (2011) do not discuss which of these assumptions is made.

74. This procedure requires an exclusion restriction, a variable that affects accumulated schooling but does not affect the school attendance decision. Attanasio, Meghir, and Santiago (2011) use distance to a secondary school in the period preceding the program for this restriction.

75. These counterfactuals in Todd and Wolpin (2006) pertain to households prior to having children and reflect the "long-run" effect of the program. Predictions of the one-step-ahead effect of the subsidy that takes as given the current household composition reflects the "short-run" effect of an unanticipated policy change.

76. Todd and Wolpin (2006) also find that the proportion of children who complete at least sixth grade actually falls below the nonsubsidy level, suggesting that, in order to earn the bonus, families are substituting more schooling for some children and less for others. In fact, as evidence for this behavior, the within-family coefficient of variation in completed schooling increases by 25 percent with the bonus as compared to the no-subsidy case. Thus, the effect of the bonus would be largely to induce children who were already attending junior secondary school to complete ninth grade.

77. This amount is close to the maximum benefit that families may currently receive under the program in any year and represents about a 50 percent increase in annual family income.

78. Attanasio, Meghir, and Santiago set the distance to newly built schools to 3 kilometers.

79. This figure is an estimate the author made of the effect by averaging over single-age effects reported in Attanasio, Meghir, and Santiago (2011).

80. Todd and Wolpin (2006) assumed the change in the child wage to be the same independent of gender and age.

81. This experiment does not appear in the Attanasio, Meghir, and Santiago (2011) paper. I thank the authors for performing this experiment at my request.

82. A similar pattern exists for girls. The predictions are increases of 2.35 times for the single-child estimator in Todd and Wolpin (2008) and 2.28 times for the estimator in Todd and Wolpin (2006) but only 1.27 times for the multiple-child estimator in Todd and Wolpin (2008). Note that although a naive estimate of doubling the effect would be reasonably close for boys, based on the three similar estimates, that estimate would be considerably worse for girls based on the two similar estimates.

83. This discussion is taken from Keane and Wolpin (2007).

84. The idea is an old one in psychology. Mosier (1951) named the procedure "validity generalization," that is, validation by generalizing beyond the sample. Recently, Busemeyer and Wang (2000) have argued for its more widespread adoption in psychology and provide Monte Carlo evidence on its performance in model selection. The use of models to forecast out-of-sample behavior is also common in the marketing literature, where considerable effort has been devoted to forecasting demand for new products. Few of the papers in that literature, however, compare predictions to subsequent demand after the product is introduced.

85. The models and methods discussed above have been partial equilibrium. Although estimation of the partial equilibrium effects of a policy intervention is clearly a necessary ingredient for performing an *ex ante* policy evaluation, such estimates may be far off the mark if there are important general equilibrium effects. Although that does not seem to be the case for the PROGRESA evaluation of Todd and Wolpin (2006), as discussed below, the development of methods that incorporate general equilbrium effects should be a priority.

Chapter 3

1. The papers by Ehrenberg and Oaxaca (1976) and Burgess and Kingston (1976) were not published in this volume but were presented at the conference.

2. Flinn and Heckman (1982) had previously structurally estimated a stationary search model.

3. Burdett (1978) extended the standard unemployment search model to allow for search on the job.

4. It is a common theme in the structural literature to build on and extend the theoretical literature in developing estimable models. This is the case with Rendon's (2006) paper, which builds on the earlier work of Danforth (1979). Lentz (2009) also structurally estimates a sequential search model with savings. Unlike the standard model, the wage offer distribution is taken to be degenerate, and agents choose their search intensity, which affects the rate at which job offers are received. Lentz uses the model to empirically determine the optimal unemployment insurance scheme.

5. Detailed reviews of this literature can be found in the chapter by Mortensen and Pissarides in the *Handbook of Labor Economics* (Volume 3b, 1999) or the more recent survey by Eckstein and van den Berg (2007).

6. This section draws on Wolpin (1995).

7. The LHS is linearly increasing in V and passes through the origin. The RHS is monotonically decreasing in V until it reaches b and is then constant. There will be a unique intersection, and a unique $V > 0$, as long as $b > -\beta q \text{Emax}(W(w), 0)$.

8. In a continuous time model in which the arrival of offers follows a Poisson process with parameter λ, the implicit reservation wage equation is identical except that the instantaneous arrival rate (λ) replaces the offer probability, q.

9. Wolpin (1995) provides a proof.

10. The distribution of accepted wages is the truncated distribution of wage offers, namely, $f(w \mid w \geq w^*) = \dfrac{f(w)}{1 - F(w^*)}$.

11. It has been shown that the latter effect dominates for distributions that are log concave (Burdett and Ondrich, 1985; Flinn and Heckman, 1983).

12. There are statistical issues better handled by specifying the hazard function, such as dealing with incomplete spells and time-varying regressors. See Meyer (1990) for an example of this approach. The issues I raise, however, are easier to demonstrate in a regression framework but hold in the hazard framework as well.

13. The assumption that new unemployment spells are renewal processes rules out any structural connection between spells; for example, it rules out that the benefit level depends on the preunemployment wage. In that case the correlation between the wage offer distribution and the benefit level could arise, for example, if the more generous states were those with higher wages.

14. An alternative would be to assume that once the terminal period is reached, the individual accepts the next offer that arrives, in which case the reservation wage at T (and after) is zero.

15. See Wolpin (1987) for the particular case in which $F(w)$ is either normal or log normal.

16. Meyer also presents a static labor supply model as an alternative to the search model.

17. Whether an individual is eligible for an extension may also depend on an individual's work history, which I ignore.

18. Note that in equation 3.8, the probability that the eligibility horizon will change at $t + 1$ is written as a function of the state of the economy at t. To motivate that formulation, assume that the horizon is changed at $t + 1$ if $Z_{t+1} \geq \underline{Z}$ and that Z_{t+1}, which is unknown at t, follows a first-order autoregressive process. Then $\omega(Z_t) = \Pr(Z_{t+1} \geq \underline{Z} \mid Zt)$.

19. Any variable that belongs in the reservation wage also belongs in the hazard rate.

20. In the more general case where there than can be more than one extension, the probability of an extension may differ for each extension.

21. In order to better fit the spikes observed at typical exhaustion periods, Meyer includes in the hazard rate specification an indicator variable equal to 1 in period t of an individual's employment spell if benefits would have lapsed in that period: $I(t = t_c)$. He ascribes this effect to advanced planning on the part of firms to recall workers when benefits are exhausted and/or to workers postponing the start dates of new jobs. This behavior is not part of either model Meyer references.

22. This section follows the development in Flinn and Heckman (1982) and Wolpin (1995). The identification results here carry over to the richer model with benefit exhaustion.

23. The minimum observed wage is a superconsistent estimator of the reservation wage in that it converges at rate N. This leads to nonstandard asymptotics in the likelihood estimation of the search model (see Flinn and Heckman, 1982, and Christensen and Kiefer, 1991).

24. Setting $w_t^* = w_{t+1}^* = w^*$ and $h_{t+1} = h$ yields the infinite horizon implicit reservation wage function.

25. The implicit reservation wage equation given above would no longer hold in this case. In particular, the integration would also have to be taken over the distribution of the unobserved cost of search, recognizing that the reservation wage would be a function of that cost.

26. In some instances, wage rates that are directly reported in hours or weeks are inaccurate. In other cases, wage rates are derived from a division of earnings reported over a longer period, say annually, and hours worked reported over that period. The inaccuracy arises from a seeming mismatch in the time period between earnings and hours.

27. Actually, this waiting period is only for those who are unemployed through layoff. Those who quit or were fired had a waiting period of 5 weeks. Ferrall assumes the waiting period to be 2 weeks independent of the reason for the unemployment spell.

28. As previously noted, the lower bound of the support for the Pareto distribution cannot be identified. Ferrall fixes that value. Kiefer and Christensen (2009) also use the Pareto distribution and impose, based on the wage-posting model of Burdett and Mortensen (1998), the individual's reservation wage as the lower bound.

29. The TANF program was promoted by the Clinton administration as an end to "welfare as we know it."

30. The amounts are in real 1987 New York dollars.

31. To see that, rewrite equation 3.15 as $y_{its} = \alpha_1 B_{its} + (\alpha_2 - \alpha_1)B^T_{its} + \beta X_{its} + u_{its}$ and note that B_{its} and B^T_{its} are positively correlated.

32. This method avoids the integration problem that arises with missing state variables.

33. Even that figure is not comparable because work experience and schooling, which influence wage offers, are affected by preferences shocks that alter demographics.

34. These last two experiments attempt to provide reasonable approximations to time-limit and work requirement rules adopted by states after TANF was introduced (see Fang and Keane, 2004).

35. This is consistent with the findings in Fang and Keane (2004) that time limits as they have actually been applied under TANF can account for very little of the fall of welfare participation since 1996.

36. Because state dummies were included in the model, in order to simulate the data a model was estimated with the same parameters incorporating the data from Texas and allowing for Texas-specific effects.

37. This value-added specification, in which the previous stock is a sufficient statistic for all prior inputs, is prominent in the literature on estimating cognitive achievement production functions, as discussed in a later section. The inclusion of ability in a value-added specification is not common, although it can be motivated (see below).

38. Hause (1972) did not explicitly associate the constant term in equation 3.24 with the (log) rental price.

39. Card (1999) provides a summary of the results from that more recent literature.

40. This section draws on Rosenzweig and Wolpin (2000) and Wolpin (2003).

41. The additive separability between schooling and experience in equation 3.26 is made for convenience and has no effect on the basic conclusion.

42. The derivation uses the approximation that $\log(1 + x) = x$ for small x, for both $x = r$ and

$$x = \frac{c}{V(s=0|S_0)}.$$

This latter case will hold if the direct cost of schooling is small relative to the present value of earnings with completed schooling S_0.

43. Rosenzweig and Wolpin (2000) draw a distinction between natural "natural experiments," such as the variation utilized in Angrist and Krueger (1991), and natural experiments that use nonnatural variation such as differences in state laws.

44. The assumption that quarter of birth is random has not gone unchallenged (Bound and Jaeger, 1996). Notice, however, that with enough data, the instrument could be defined by dates of birth that differ by only 1 day (as in the illustration), effectively eliminating the instrument validity issue.

45. Rosenzweig and Wolpin (2000) question the validity of the instrument on the grounds that the existence of a sister is increasing with fertility. If fertility is a choice that is affected by parental ability and ability is heritable, then the existence of a sister will be correlated with ability. They propose that using the gender of the first birth as the instrument for the schooling of subsequent children would circumvent this problem, They provide evidence that the gender of the first birth is random and, thus, would not be correlated with the ability of higher-order children.

46. The interested reader is referred to Rosenzweig and Wolpin (2000) for a discussion of the biases that arise due to the endogeneity of work experience and to discussions of alternative approaches, for example, the use of twins, to the estimation of the schooling effect on wages.

47. The reason for including this psychic cost to returning to school is that the fraction of individuals who do so in the data is very low. Adding such a term illustrates the natural, and quite necessary, interplay between model development and data as discussed in chapter 2.

48. It is not necessary that the age range begin at age 16, nor that there be no gaps in the data. The estimation approach would need to be modified.

49. I discuss some of that literature in the next section.

50. Multiplying this likelihood by $\Pr(S_{i,16})$, which is not dependent on parameters, would not affect any of the parameter estimates. There is no assumption as to the interpretation of the correlation between type and initial schooling. Keane and Wolpin (1997) also allow for normally distributed additive mean zero measurement error in wages to reduce the impact of outlier wages on parameter estimates.

51. As seen from the structure of utility and the budget constraint, $\bar{\alpha}_m$ and γ_m $(m = 4,5)$ cannot be separately distinguished; γ_m, for example, can be normalized to zero for all m. In addition, one α_m must be normalized; for example, if α_3 is normalized to zero, all other α_m's will be relative to the nonpecuniary value of the military occupation.

52. In the extended model estimated by Keane and Wolpin, there are over seventy-five parameters.

53. All figures are in 1987 dollars.

54. It is, however, 20 percent higher for the first type as measured from age 26 as lifetime utility from that age becomes dominated by life cycle earnings paths.

55. High school graduation bonuses have been advocated by Robert Reich, former Secretary of Labor in the Clinton administration.

56. As discussed in the previous chapter, the credibility of structural estimation exercises hinges on providing evidence on model validation. Keane and Wolpin (1997) use

data on nearby cohorts from the CPS to assess the ability of the model to forecast occupational choices later in the life cycle. They find that the forecasts of white-collar employment are quite close, whereas those of blue-collar employment are somewhat overstated.

57. Krueger (1999), p. 528.

58. Nye, Hedges, and Konstantopoulos (2000), p. 147.

59. I_1 is also a function of $\bar{S}_1$, so equation 3.37 is based on having substituted in equation 3.38. Notice that one could also eliminate A_0 from both equations 3.37 and 3.38 by substituting the input demand equation for I_0 into equation 3.36 and the resulting equation into equations 3.37 and 3.38. Thus, the input demand equations can be written solely as a function of family endowments, in wealth, the child's endowment, and parental preferences.

60. The change in the school input in the first school year can have effects on achievement not only in the first year but also in later years (see the discussion below). Follow-up studies of the effect of Project STAR on achievement in years subsequent to the termination of the experiment have found long-run effects (e.g., Krueger and Whitmore, 2001; Nye, Hedges, and Konstantopoulos, 2000).

61. Notice that although A_0 enters the demand function for I_1 as written, the population can be fully described solely as a function of the endowment distribution, which determines the distribution of A_0.

62. The comparison does provide evidence on the value of the average indirect effect:

$$\int \frac{\partial g_1}{\partial I_1} \frac{\partial I_1}{\partial (S_1 - \bar{S}_1)} f(A_0, W, \mu, \varepsilon) .$$

63. Theories of child development posit important links between experiences during infancy and early childhood and later childhood cognitive and noncognitive outcomes.

64. Their main goal is to assess the role of home and school inputs and of maternal "ability" in producing cognitive achievement and in explaining the black-white and Hispanic-white test score gaps This assessment is carried out through the use of out-of-sample forecasts in which the inputs of minorities are replaced by those of whites. The use of cross-validation seems particularly appropriate for this exercise.

65. This specification would arise if the following restrictions in equation 3.47 hold: $\alpha_k - \gamma\alpha_{k-1} \neq 0$; $\beta_a - \gamma\beta_{a-1} = 0$; and $\varepsilon_{ija} - \gamma\varepsilon_{ij,a-1} = \eta_{ija}$ is *iid*.

66. Recall that Todd and Wolpin (2003) take the cumulative specification as the starting point. With two skills, the corresponding cumulative specifications would presumably be:

$$A_a^C = \varsigma_0^C + \varsigma_1^C X_a^C + \ldots + \varsigma_a^C X_1^C + \delta_1^C \mu_a^C + \delta_2^C \mu_o^N + \varepsilon_a^C ,$$

$$A_a^C = \varsigma_0^N + \varsigma_1^N X_a^N + \ldots + \varsigma_a^N X_1^N + \delta_n^N \mu_a^C + \delta_2^N \mu_o^N + \varepsilon_a^C .$$

Although backward substitution in equation 3.55 would reverse engineer a function that looks like the above cumulatives, the coefficients would not have the same skill-specific interpretation.

67. Cunha, Heckman, and Schenach (2010) extend the factor model of Cunha and Heckman (2008) to a nonlinear nonparametric setting.

68. The literature on assessing the impact of maternal employment and child care on child outcomes is extensive. Reviews of that literature can be found in Blau (1999), and Bernal and Keane (2011).

69. Identification of the achievement production function is achieved in part with the exclusion restriction that the local unemployment rate affects maternal employment but not cognitive achievement.

70. Brown and Flinn (2011) also study the relationship between divorce and child quality. I discuss Tartari because of the closer connection of her model to the DCDP approach.