

American Economic Association

Natural "Natural Experiments" in Economics

Author(s): Mark R. Rosenzweig and Kenneth I. Wolpin

Source: *Journal of Economic Literature*, Vol. 38, No. 4 (Dec., 2000), pp. 827-874

Published by: American Economic Association

Stable URL: <http://www.jstor.org/stable/2698663>

Accessed: 07/09/2011 17:56

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Economic Literature*.

<http://www.jstor.org>

Natural “Natural Experiments” in Economics

MARK R. ROSENZWEIG and KENNETH I. WOLPIN¹

1. *Introduction*

THE COSTLINESS OF and limitations on experiments involving human subjects have long been identified as major constraints on the progress of economic science. Indeed, it has been increasingly recognized that identification of many interesting parameters, such as the effects of schooling or work experience on earnings or of income on savings, requires attention to the fact that the variation in many of the variables whose effects are of interest may not be orthogonal to unobservable factors that jointly affect the outcomes studied. Such unmeasured or unmeasurable factors may include pre-existing or endowed skills (“ability”), preferences, or technologies that vary across individuals or firms in the economy. The possible existence of heterogeneity in these attributes means that almost all estimates are open to alternative interpretations in terms of self-selection by such traits. In determining the returns to schooling, for example, individuals cannot be considered to be randomly sorted among schooling levels. Thus, that more-schooled individuals have higher earnings may reflect the fact

that more able individuals prefer schooling or face lower schooling costs. Similarly, that fertility and female labor supply are negatively correlated may reflect variation in preferences for children and work in the population.

Economists have used experiments that purposively randomize treatments to assess their effects in the presence of heterogeneity. Among the issues that some of the most prominent experiments have addressed are the impact of a negative income tax on labor supply, the effects of class size on test outcomes, and the effects of job training programs on earnings. However, these “man-made” experiments are subject to the criticisms that they lack generalizability and, most importantly, often do not adhere in implementation to the requirements of treatment randomness. The most widely applied approach to identifying causal or treatment effects, which has a long history in economics, employs instrumental variable techniques. This approach essentially assumes that some components of non-experimental data are random. That is, it is assumed that some variable or event satisfies the criterion of “randomness”—the event or variable is orthogonal to the unobservable and unmalleable factors that could affect the outcomes under study. This assumption, along with a set of additional assumptions

¹ University of Pennsylvania. We are grateful to two anonymous referees and the editor for helpful comments on an earlier draft of this paper. Partial support for the research was provided by NIH grants HD30907 and AG11725 and NSF grants SBR95-11955 and SBR93-08405.

about behavior, yields estimates of economic parameters that are of general interest, e.g., the returns to schooling, the income elasticity of savings. Not surprisingly, the assumptions made, inclusive of randomness, are subject to much skepticism.

Given the difficulties of carrying out and financing experiments with near-perfect random treatments to answer questions of general importance and the lack of credibility of many of the assumptions of standard instrumental variable studies, economists as well as researchers in other fields have sought out "natural experiments," random treatments that have arisen serendipitously. These putative natural experiments are usually changes or spatial variation in rules governing behavior, which are assumed to satisfy the randomness criterion. Indeed, 72 studies using the phrase "natural experiment" in the title or abstract issued or published since 1968 are listed in the *Journal of Economic Literature* cumulative index.² Many of these studies do not produce easily generalizable results about behavior because of the specific nature of the rule changes that are studied, as in the randomized-experiment literature. The major problem with these studies, however, is that the assumption of randomness is not credible.

In recent years economists, in recognition that "nature" provides almost perfect randomness with respect to im-

² Among the "natural" experiments in these studies are changes in nineteenth-century Brooklyn welfare laws, cross-country differences in East African educational policies, the introduction of a Civil-War-era union pension plan, cross-state differences in wage distributions, trade policy reform in New Zealand, the passage of the U.S. Tax Reform Act of 1986, differences in salary restrictions across U.S. professional basketball and baseball, the introduction of Eastern European privatization rules, changes in Ohio rules for financing mental health care, and British dividend tax reforms.

portant variables, have ingeniously exploited naturally random events as instrumental variables.³ These natural experiment studies have attracted a great deal of attention both because of the appeal of the instruments and because the studies have addressed important questions in economics. Five major random outcomes that arise from biological and climate mechanisms have been used as instruments: twin births, human cloning (monozygotic twins), birth date, gender, and weather events. These natural outcomes, which are plausibly random with respect to at least two of the major sources of heterogeneity in human populations—tastes and abilities—have been used to study three issues: What are the returns to schooling and labor market experience? How sensitive are consumption, savings, and labor supply to temporary and permanent changes in income? How responsive is women's labor force participation to fertility change?

This review essay examines this recent literature exploiting natural events as instruments to assess to what extent it has advanced empirical knowledge. The advantage of the natural experimental approach is that the assumption of randomness for the instrumental variables employed is more credible than for those instruments used in almost all other studies. But a weakness of many of the studies that adopt this approach is that the necessary additional behavioral, market, and technological assumptions needed to justify the authors' interpretations of the estimates obtained are absent. The impression

³ Other scientists have also exploited natural events, the most prominent studies being those based on human cloning (twins) that attempt to identify the separate roles of genetics and environment (e.g., K. McCourt, T. J. Bouchard, David T. Lykken, M. A. Tellegen, M. Keyes 1999).

left by this literature is that if one accepts that the instruments are perfectly random and plausibly affect the variable whose effect is of interest, then the instrumental-variables estimates are conclusive. Indeed, part of the appeal of this literature is its simplicity, in contrast, for example, to the structural estimation literature, in which complex estimation strategies are used to identify the structural parameters of fully parameterized but highly restrictive models. A characteristic of this approach is that all assumptions are transparent, but the structure of these models incorporates, as does any economic model, strong and sometimes unappealing assumptions about behavior or markets, and about functional forms and statistical distributions, often made strictly for reasons of tractability.

However, the absence of models in the natural natural experiment literature does not mean that there are no important and implausible assumptions being implicitly used by the authors in interpreting the estimates they have obtained. Randomness and explanatory power are necessary but not sufficient conditions for identification of a parameter that is of interest. Moreover, in the few cases where an explicit model is used to motivate the specification and exclusion restrictions used, it is not always clear how robust is the study's interpretation to even minor changes in the model. In this review we (i) summarize the methodology and findings from twenty studies employing natural events as instruments; (ii) examine and clarify the set of assumptions, beyond the randomness of the natural events, that are implicitly made in these studies that lead to the authors' conclusions about the empirical results obtained; (iii) show how the relaxation of some of these restrictions, consistent with empirical evidence or with well-established mod-

els and/or evidence in the literature, changes the interpretation of the results; and (iv) provide additional empirical evidence on the validity of these implicitly made assumptions using the same set of natural events. We clarify the assumptions and the interpretation of the studies' findings by constructing simple economic models pertinent to the phenomena studied. The new evidence is used to suggest the additional research, empirical and theoretical, that is needed to obtain a more reliable understanding of the results from this literature and for future inquiries using the natural natural experimental approach.

2. Instrumental Variables, Natural Natural Experiments, and the Role of Economic Theory

To clarify the empirical methodology that underlies the natural natural experiment studies and to highlight the issues which we will illuminate by employing simple behavioral models in this review essay, it is useful to start with a presentation of the statistical model used in almost all of the natural natural experiment studies in its simplest form. The first element is the equation that contains the parameter of interest β :

$$Y = \alpha + \beta X + \varepsilon. \quad (1)$$

The canonical problem is that in the population X is correlated with the error term ε so that least squares estimates of β will be biased. The solution proposed is some variable Z that affects X but is not correlated with ε .⁴ This variable is then used as an instrument for X , and the instrumental variables estimator is:

$$\beta_{IV} = \frac{\text{cov}(Y, Z)}{\text{cov}(X, Z)} = \frac{\text{cov}(X, Z)\beta + \text{cov}(\varepsilon, Z)}{\text{cov}(X, Z)}. \quad (2)$$

⁴ Most studies in this literature in fact conveniently assume that only one X is correlated with ε , since then only one natural natural instrument is required.

A variant of (2), applied frequently in the natural experiment literature, is the "Wald" estimator, which is simply an instrumental variables estimator based on grouped data (Abraham Wald 1940).

As seen in (2), identification of β thus requires that Z and X covary but not Z and ϵ . Each study must therefore define what is in the error term in order to provide a plausible story for why the instrument Z is not correlated with it. The description of what is in the error term is thus the critical part of all studies employing instrumental variables. Because what is in the error term depends in turn on the specification of the equation of interest, it is important to have some explicit criteria or framework for deciding what constitutes a plausible specification of (1), in particular, what other variables besides X belong in (1) and what are the relationships between these other variables and the X , Z , and ϵ ? It is in providing a coherent framework for answering these questions that most of the studies employing natural experiments fall short. And, what is excluded from (1) turns out to be critical in most of the studies for achieving identification of β . Our strategy is to show using very simple behavioral models what assumptions about the characteristics of technology, preferences, and markets are required to justify these exclusion restrictions.

We divide our review essay into three main sections defined by the β of interest. In the first section, we examine studies in which Y is (log) earnings and β is the return to human capital investments in the form of either schooling or post-school investments. The natural experiments considered in this section include a child's date of birth, a child's gender, the date-of-birth draft lottery, and human cloning. We first show that these studies provide a wide range of estimates of β , are almost

evenly divided with respect to the direction of "ability bias," and do not often provide estimates that differ very much from those obtained using least squares. With respect to the properties of the instrumental-variable estimates, the two studies concerned with the returns to schooling assume that the only other variable in (1) besides schooling is age and that the only component of the error term correlated with schooling is ability. And the sex of one's sibling, or one's month of birth, the instruments used, are plausibly assumed to be uncorrelated with ability. We show using a very simple model of schooling choice that the date-of-birth and child gender instrument do identify, under these assumptions, the returns to schooling, but for different ability groups in the population. We also show, however, that (i) if work experience is a better proxy for post-school human capital than is age, the instrumental estimates of the returns to schooling from the two natural experiments are biased, and the biases go in opposite directions, and that (ii) if work choices after school are made optimally and work experience affects earnings, there is additional bias in the two IV estimates that will depend on the age of the sample population. We also present evidence from U.S. panel data, using the same natural instruments, that work experience is correlated with the error term in (1) and with schooling. Thus, the particular applications of the two instruments are likely to provide biased estimates of the return to schooling. Moreover, the two instrumental variables estimators will differ from each other when applied to the same population, especially if returns differ by ability group, and each instrumental variable estimate will provide different (biased) estimates for different-aged populations.

The seemingly innocuous but empirically dubious assumption used in the two natural natural experiment schooling studies, that age is the only other determinant of earnings besides schooling (and ability), turns out to be critical for identification even if the natural events used as instruments are uncorrelated with ability. Similarly, we show using the same behavioral model that the unstated critical assumptions in the draft-lottery study of the returns to civilian work experience (where ϵ is again defined to be ability and $X =$ experience in (1)) are (i) that schooling is uncorrelated with ability and (ii) schooling is uncorrelated with experience. The former assumption is inconsistent with that made in the natural natural experiment studies of schooling, and the latter we show to be empirically untrue. Finally, we also consider in this section the growing set of studies that make use of natural human cloning to estimate the return to schooling. We show that most of these studies provide downward biased estimates of the return to schooling if work experience is an important determinant of earnings, even if the critical assumption that the differences in the schooling within twin-pairs is purely random holds. We also show, however, that evidence on differing birthweights within pairs of monozygotic (genetically identical) twins and a plausible model of optimal job-taking suggest that even with information on work experience, twin-studies yield estimates of schooling returns that are likely to be biased, and it is not clear *a priori* whether the bias is upwards or downwards.

In the next section, we look at studies using weather events as natural instruments to estimate and test hypotheses about the effects of changes in permanent and transitory income on consumption and labor supply. In these

studies β is the permanent or transitory income effect, and weather events are plausibly assumed to be orthogonal to the error term, which implicitly is defined to contain farmer preferences. In all studies, the sample population consists of farmers. We show using a simple two-period model that identification of transitory income effects in these studies depends critically on a number of assumptions about the structure of preferences and labor markets. In particular, the studies have implicitly assumed that either family and hired labor are perfect substitutes, or, if not, that a market exists for all types of labor, that labor is perfectly spatially mobile, and/or that leisure and consumption are strongly separable. We show that if the assumption of perfect labor mobility is relaxed, that even if family and hired labor are perfect substitutes and markets are complete, the weather-based instrumental estimates of transitory income effects are biased upwards given the specifications used in two of the most well-known studies. We also show, however, using data from a setting similar to those used in these studies, that the assumption that family labor is perfectly substitutable for hired labor is rejected.

The third major section contains a review of studies in which the β of interest is the effect of an exogenous change in fertility (X) on the labor supply choices of married women (Y) and that rely on the occurrence of a twin first birth or the sex composition of the first two children as natural natural instruments. In these studies, the error term is defined to contain preferences, and it is plausibly assumed that neither having a twin on the first birth nor the sex of children are correlated with preferences. To illustrate plausible conditions under which the random sex-sameness of children or a twin birth may directly affect labor supply net of fertility (so

that Z and ϵ are not orthogonal as assumed), however, we again set out a bare-bones model, in this case one in which parents care about the sex composition of children and make labor supply and fertility decisions. The model illustrates the types of strong restrictions required so that the natural instrumental-variable estimate of the fertility effect on labor supply based on the sex-sameness of initial births or twin births corresponds to the experiment in which fertility itself is varied exogenously. Here we show that it is necessary to assume that the sex-sameness of births or having twins has no effect on the costs of children for identification to be achieved, or that strong assumptions about the separability of fertility, labor supply and sex-composition must be imposed. We adduce evidence from data from rural India that suggests that the sex-composition of children at least in that context has significant effects on child-rearing costs so that the sex-composition instrumental-variables estimate does not identify β .

In the final section, we assess the contribution of the natural natural experiment studies and propose a research strategy which makes use of the general insights of this literature.

3. Estimating the Return to Human Capital Investments: Schooling and Work Experience

The presence of "ability bias" in estimates of the return to human capital investments in the form of schooling is a longstanding problem. Early proposed solutions to the problem include using test scores as (imperfect) measures of ability (e.g., Zvi Griliches and William Mason 1972) and using siblings to control for family-level unobservables (Gary Chamberlain 1977). More recently, it has been suggested that natu-

ral experiments that induce variation in school attainment unrelated to ability can be used to eliminate ability bias in estimating the return to schooling. These natural natural experiments include date of birth, as used in Joshua Angrist and Alan Krueger (1991), and child gender, as used in Kristin Butcher and Anne Case (1994). In addition, seven studies have used differences in the schooling attainment of individuals in monozygotic twin pairs—human cloning—to eliminate the contamination of returns estimates from genetic differences in ability. And one study (Angrist 1990) has exploited the date-of-birth draft lottery during the Vietnam War to estimate the returns to civilian work experience.

Table 1 summarizes the results from these studies of the human capital determinants of earnings that exploit natural natural experiments. One interesting feature of the set of studies is that only one of the studies examines jointly the returns to both work experience and schooling. And, of the two studies providing estimates of experience returns, one excludes schooling from the specification. As we discuss, these omissions are not only (or at all in some cases) due to data restrictions, but are related importantly to the limitations of the natural natural experiment approach. What is also striking about the set of natural natural-experiment schooling returns estimates is that, not only do they have an extraordinary range—from less than one percent to over 18 percent—they also are almost exactly evenly divided between those that indicate a negative ability bias (4) and those that indicate a positive ability bias (6), although in many cases the estimated returns do not differ significantly when the instrument is used. It is difficult to infer from the pattern of estimates across the studies the precise role of

TABLE 1
ESTIMATES OF THE RETURNS TO SCHOOLING AND WORK EXPERIENCE USING NATURAL EVENTS AS INSTRUMENTS

Study	Natural Event	Data Set	OLS Return to Schooling	Instru- mented Return to Schooling	Instru- mented Return to Work Exp.	"Control" Variables
Angrist 1990	Date of birth + Vietnam draft lottery	U.S. Social Security Administration Continuous Work History Sample, for men	—	—	.102-.003*years of work exp.	
Angrist and Krueger 1991	Date of birth + school leaving age + school entry age	1970 and 1980 U.S. Censuses, men aged 41–50	.080 (1970) .071 (1980)	.072 (1970) .102 (1980)	—	Age, race, SMSA, married
Butcher and Case 1994	Gender (any sisters)	Panel Study of Income Dynamics, white women	.091	.184	—	Age
Behrman, Hrubeck, Taubman, and Wales 1980	Human cloning	U.S. NAS-NRC Twins Sample of White Veteran Males Born in 1917–27	.080	.003	—	Age
Ashenfelter and Krueger 1994	Human cloning + instruments (report by co-twin)	Twinsburg, Ohio Annual Twins Festival Sample, 1991—men and women	.116 ^a	.129	—	Age, tenure in last job, union, married
Behrman, Rosenzweig and Taubman 1994	Human cloning + instruments (report by twins' oldest child)	Biographical Questionnaire sample from the Minnesota Twin Registry, twins born in Minnesota 1936–55—men	.101 ^a	.050	—	Age
Miller, Mulvey and Martin 1995	Human cloning + instruments (report by co-twin)	Australian Twin Register 1980–2 and 1988–9	.073 ^a	.045	—	Age, married
Ashenfelter and Rouse 1998	Human cloning + instruments (report by co-twin)	Twinsburg, Ohio Annual Twins Festival Sample, 1991, 1993—men and women	.113	.100	—	Age, tenure in last job, union, married
Behrman and Rosenzweig 1999	Human cloning + instruments (report by co-twin)	1994 Minnesota Twin Registry survey sample, twins born in Minnesota 1936–55—men and women	.113 ^a	.104	.0084	Age, post-school full-time work experience
Rouse 1999	Human cloning + instruments (report by co-twin)	Twinsburg, Ohio Annual Twins Festival Samples, 1991–93, and 1995—men and women	.111	.119	—	Age, tenure in last job, union, married

^a Estimate using same instruments to correct for measurement error in schooling as used in cloning-based estimates.

the different natural experiments exploited, as some use samples of males only, one uses only females, and some use a sample including both males and females. However, as we show, an important factor in explaining the differences in findings is the natural-experiment based instrument used. One reason, as shown in human capital models by David Card (1994) and James Heckman (1997) and generalized in a large literature concerned with evaluation (e.g., Guido Imbens and Angrist 1994), is that when there is heterogeneity in returns, the "treatment" effect that is identified is for the group or groups whose behavior is influenced by the intervention. And the instruments affect different groups.

To show how these natural experiments are used to estimate the returns to human capital investments, the essence of the problem that is being solved, some of the factors that lead to differences among the estimates, and the assumptions made implicitly by the authors of the studies, we begin with a standard model of schooling choice incorporating ability heterogeneity. We assume that earnings at any age a , y_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

$$\ln y_a = f(S, \mu) + g(X_a, \mu), \quad (3)$$

where f and g are monotonically increasing in their arguments. We initially assume, as do all but two of the studies, that all individuals work full-time after completing schooling. Each individual enters school at a mandated school entry age a_e and must remain in school until the mandated minimum school leaving age a_K . Thus, school attainment at a_K is $S_0 = a_K - a_e$. It will be 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_K , period one, is denoted by $s_1 = 1$ and nonattendance by $s_1 = 0$; completed schooling, at the end of period one, S_1 , is therefore either $S_0 + 1$ or S_0 . An individual who decides not to attend school in period one works that period and all subsequent periods; i.e., the individual works from period $a = 1$ to the end of working life $a = A$, while an individual who attends school does not work in that period, but works in all subsequent periods, i.e., from $a = 2$ to $A + 1$. There is a direct cost of attending school denoted by c . In addition, the assumption that school attendance precludes working implies that school attendance entails an opportunity cost in terms of the earnings that are foregone.

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 $\beta (= \frac{1}{1+r})$, the present value of each alternative, $V_1(s_1 = 1 | S_0)$ and $V_1(s_1 = 0 | S_0)$, is:

$$\begin{aligned} V_1(s_1 = 1 | S_0) &= \exp[f(S_0 + 1, \mu)], \\ &\sum_{a=0}^{A-1} \beta^{a+1} \exp[g(a, \mu)] - c, \\ V_1(s_1 = 0 | S_0) &= \exp[f(S_0, \mu)]. \end{aligned} \quad (4)$$

$$\sum_{a=0}^{A-1} \beta^a \exp[g(a, \mu)].$$

The decision rule is to attend school if $V_1(s_1 = 1 | S_0) \geq V_1(s_1 = 0 | S_0)$, which reduces, after some manipulation, to:⁵

$$\begin{aligned} s_1 = 1 &\text{ if } f(S_0 + 1, \mu) - f(S_0, \mu) \\ &\geq r + \ln \left[\frac{c}{V_1(s_1 = 0 | S_0)} + 1 \right], \end{aligned} \quad (5)$$

$$s_1 = 0 \text{ otherwise.}$$

⁵ We use the approximation $\ln(1 + r) \approx r$ in deriving the decision rule.

The individual attends school if the percentage increase in earnings from attending school, the marginal return to schooling, is sufficiently greater than the interest rate.⁶ If ability increases the marginal schooling return,

$$\frac{\partial[f(S_0 + 1, \mu) - f(S_0, \mu)]}{\partial \mu} > 0,$$

then there exists a cut-off value of ability, μ^* , such that individuals with ability at or above that cut-off attend school and those below it do not.⁷ 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) return to schooling calculated from (\ln) earnings differences between the two schooling groups (given experience) overstates the (average) return to schooling, i.e., $E_\mu[f(S_0 + 1, \mu) | \mu \geq \mu^*] - E_\mu[f(S_0, \mu) | \mu < \mu^*] > E_\mu[f(S_0 + 1, \mu) - f(S_0, \mu)]$. The challenge is to obtain an estimate of the returns to schooling without this "ability bias" for every ability group in the population. Almost all of the empirical literature assumes that the return to schooling is identical for all ability groups, so that there is one true return to schooling.

⁶ If the direct cost of attending school is small relative to the present discounted value of lifetime earnings, then the right hand side can be approximated by $r + c / V_1(s_1 = 0 | S_0)$. Regardless, if the direct cost of schooling is zero, either condition reduces to the marginal return to schooling exceeding the interest rate. With continuous schooling choices and zero direct cost, the rule for optimal schooling is to equate the marginal earnings return to the rate of interest, i.e., $f_s = r$. The second order condition would require in that case that $f_{ss} < 0$.

⁷ Notice that when c is non-zero, the right-hand side is falling monotonically with ability. A cut-off ability level will exist, therefore, even if the marginal return to schooling is independent of ability.

3.1 The Returns to Schooling When Schooling Is a Choice: The Angrist-Krueger and Butcher-Case Natural Experiments

3.1.1 Angrist and Krueger: Age at Birth

Angrist and Krueger (AK) suggest that natural variation in dates of birth, in conjunction with the existence of a birth date cut-off for school entry and a minimum compulsory school leaving age, can be used as an "instrument" for completed schooling that provides an estimate of the returns to schooling without ability bias. This study exemplifies well the difference between the natural natural experiment approach and the more prevalent natural experiment approach which often relies on variations in government-determined laws or regulations. The AK study does not use the variation in mandatory school leaving ages across or within states or other administrative units as a source of identification. To illustrate the reasoning, consider a particular locality, defined by a set of fixed school entering and school-leaving ages, in which a child must attain the age of six as of September 1 in order to enter the first grade and cannot leave school prior to attaining the age of sixteen. Then, a child whose birthday falls on September 1 will have completed ten years of schooling at the minimum school leaving age. 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 nine years of schooling upon reaching age sixteen. 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 AK 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. Type 1's comprise π_1 proportion of the population and type 2's $1 - \pi_1$. The insight of AK 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 by AK to be a random variable uncorrelated with ability (as transmitted inter-generationally), while the laws governing school-entering and leaving ages, set by each state, can be correlated with other state-specific unobservables that influence earnings.⁸ These are subsumed in a state fixed-effect. Thus, variation in state schooling laws are not treated as natural experiments, can be endogenously determined, and do not contribute to identification. 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.

AK present Wald estimates of the return to schooling based on comparing (\ln) 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). In terms of the model, it is easy to show what their Wald estimator identifies.

⁸ John Bound and David Jaeger (1996) present evidence that date of birth is correlated with a number of attributes of children that might be directly related to their later earnings net of schooling effects, including personality, mental health, and parental income. They do not provide any behavioral model that suggests why these correlations exist.

Suppose, according to (5), that the optimal level of schooling for type 1's 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_k . Consider reducing the age at entry by one year, leaving the minimum school leaving age unchanged. In that case, both ability types will complete $S_0 + 1$ years of schooling; type 1's do so because it is optimal, while type 2's are forced to remain in school an extra year. The difference in expected (\ln) incomes associated with the alternative school entry ages divided by the corresponding difference in expected schooling levels, that is, the Wald estimator, is $f(S_0 + 1, \mu_2) - f(S_0, \mu_2)$, which is the marginal return to schooling of the less able type.

AK found that for the group of men aged 41–50 in 1970 (1980), those born in the first quarter of a calendar year obtained .1256 (.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 concomitant differences in (\ln) weekly wages implied a return to schooling of .0715 in 1970 and of .1020 in 1980. In contrast, the OLS estimate was larger in 1970, .0801, and smaller in 1980, .0709.⁹

3.1.2 Butcher and Case: Child Gender as a Schooling Instrument and the Quality–Quantity Tradeoff

Butcher and Case (BC) 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 schooling return (to women) that is free of ability-bias. They discuss

⁹ Two-stage least squares estimates that control for age trends tend to provide estimates that are also close to the OLS estimates.

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 BC, 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.

BC provide evidence from three different data sources, the Panel Study of Income Dynamics (PSID), the Current Population Survey (CPS), and the National Longitudinal Surveys Mature Women's cohort (NLSMW), in support of the hypothesis that gender composition affects human capital investments in children. Specifically, they find that girls who have any sisters, conditional on the number of siblings, have lower school attainment than do girls with no sisters.¹⁰ Given this result coupled with the inherent randomness of child gender, BC argue that the existence of any sisters can serve as a valid instrument

for school attainment in estimating the earnings return to schooling.

BC recognize, however, that the existence of sisters, even if gender is randomly determined, depends on the choice of family size. There are economic models in the literature that suggest that this choice may be correlated with ability. Although the gender of any particular child is random, the probability of having a sister obviously increases with the number of siblings. Assuming fertility is subject to control, to the extent that fertility is related to the ability of children, the "any sisters" instrument will be invalid, and inclusion of number of siblings as a regressor to control for this relationship requires an assumption that parents are indifferent about the sex-composition of their children. One such behavioral mechanism creating a correlation between the immutable component of the abilities of children and number of siblings arises in the context of the joint fertility-child investment decision models first proposed by Gary Becker and H. Gregg Lewis (1973) and elaborated by Becker and Nigel Tomes (1976) to include heritable endowments of parents and children.

In these models, which highlight the trade-offs between the "quality" and quantity of children, parents receive utility from the number of children they have (fertility) and from the average quality of their children. Child quality is produced through the application of parental (time and goods) inputs conditional on parental (innate and heritable) ability. Parents with higher ability for given inputs are assumed to be able to produce higher quality children. The essence of the quality-quantity model is that the cost of children depends on total child quality (the price per unit quality times quality per child times fertility), which implies that the shadow price of fertility is increasing in the

¹⁰On the other hand, the school attainment of boys is found to be unrelated to gender composition.

average quality of children. An implication of this model is that parents with higher ability will tend to have fewer children that are of higher quality.¹¹ Given the heritability of parental ability, family size and child ability will thus be negatively related.

One of the earliest uses of the birth twinning natural experiment (Rosenzweig and Wolpin 1980a) was for the purpose of testing the quality–quantity framework. In that study, data on households from rural India were used to assess whether an exogenous increase in the number of children, brought about by a twin birth, led to reduced schooling attainment for children. Explicit attention was paid to the fact that like the “any sisters” variable, having a twin birth is positively correlated with the propensity to have a large family by dividing the number of twin births by the total number of pregnancies. This measure of “extra” births clearly does not completely conform to the randomness criterion if family size is a choice variable and led to the subsequent use of having a twin on the first birth as a natural instrument, described below.

BC, recognizing the existence of a relationship between family size and having a sister, also incorporate family size in their instrumental variables estimation procedure. They note, however, that sex-composition may still not be random if fertility is chosen in a sequential decision process. In particular, if parents have a preference for mixed-gender families then among parents who have the same completed family size,

¹¹ The effect of higher ability on the level of child investments is ambiguous and depends on how ability affects the marginal product of inputs. The higher income of higher ability parents will also tend to reduce fertility if, as was posited by Becker and Lewis, the income elasticity of quality exceeds that of quantity. For a life-cycle model with the same implications see Rosenzweig and Wolpin (1980b).

those with same-sex children will differ from those with mixed-gender in some characteristic that leads to having fewer children, e.g., lower income. BC provide as evidence that the “any sister” instrument is not contaminated by fertility selection that, for two-child families, boys with a brother have no less education than boys with a sister even though girls with a sister have less education than do girls with a brother. Although this evidence clearly shows that boys and girls are treated differently, in the absence of *a priori* knowledge of how they are treated differently, it does not refute the possibility that parents prefer mixed genders. Indeed, Angrist and Evans (1998), as described below, present direct evidence that parents do prefer mixed-gender families, which they use as the basis for the identification of fertility effects on labor supply.

BC employ two specifications in estimating the returns to schooling using instrumental variables. In one specification, BC include the number of siblings as a determinant of both earnings and schooling. The validity of this estimator rests on the assumption that parents are indifferent about gender composition. In a second specification, they omit number of siblings from the earnings function but still include it as a determinant of schooling, thus using the number of siblings as an additional instrument. Clearly, if number of siblings is related to a child’s innate ability, as theory suggests, this specification will not provide a consistent estimate of the schooling return.¹² The distribution of the total number of children across households is not the outcome of a natural experiment.

For the BC instrument “any sisters”

¹² The estimate of the schooling coefficient is only statistically significantly different from zero in the specification that uses family size as an additional identifying instrument.

to be valid it is thus sufficient that fertility not be correlated with ability, that the models of fertility choice are incorrect in some way. For example, parents may not know their children's ability prior to making their fertility decisions. Without a measure of innate ability, it is of course not possible to demonstrate conclusively the existence of a correlation of ability with number of siblings (or having sisters). Indeed, if a test that perfectly measured innate ability were available, obtaining an unbiased estimate of the return to schooling would not be a problem in the first place. We can provide evidence, however, on whether having sisters is correlated with scores from tests administered to very young children that are designed to measure ability, recognizing that such measures are unlikely to be themselves immutable and that such a correlation might exist as a result of parental investment behavior. For this purpose, we make use of the "ability" measures present in the 1979 Youth and Child cohorts from the NLS.¹³

We restrict the sample to the firstborn children of women in the NLSY who had a first birth by the age of 29. The gender of firstborn children, absent sex-selective abortion, meets the randomness criterion for all households with any children. The sample consists of 2311 first births for whom there is a valid score on the Peabody Picture Vocabulary Test (PPVT) and for whose mothers there is a valid score on the Armed Forces Qualifying Test (AFQT). The PPVT is for children aged three to six and is designed to measure verbal

ability. The AFQT has been shown to be highly correlated with IQ-type tests. A necessary condition for this exercise to be minimally credible is that the gender of the firstborn not be correlated with either the firstborn child's nor the mother's test score. The first and third columns of Table 2 provide evidence consistent with that condition; the OLS coefficient from a regression of the firstborn child's PPVT score and the mother's AFQT score on the gender of the firstborn child (being a female) is neither statistically significant nor large in magnitude.¹⁴ The data confirm that the gender of the first child satisfies the condition of randomness.

The second and fourth columns of Table 2 provide evidence based on the same data, however, that calls into question the implicit behavioral assumptions that would render the BC "any sisters" instrument as valid. Restricting attention to the 1135 firstborn children who were female, we regressed the child's PPVT score and the mother's AFQT score on whether that child had any sisters. In contrast to the correlations with respect to the gender of the firstborn, it appears that the child's PPVT score and the mother's AFQT score are both negatively related to whether the firstborn female child has any sisters. Further, as seen in column five, the mothers of female children with more children have lower AFQT scores. These results together are consistent with the notion that the less "able" mothers have more children (as is demonstrated in column five) and that their lower "ability" is transmitted to their offspring.¹⁵

¹³ Note that while there is some empirical evidence calling into question the randomness of month of birth (Bound and Jaeger 1996), no behavioral model has been presented as to why ability and intra-annual birth timing should be correlated. The randomness of fertility with respect to parental ability is called into question by well-established economic models, but there is limited evidence.

¹⁴ The mean PPVT score for the sample is 90 with a standard deviation of 20. Similarly, the mean AFQT score is 611 with a standard deviation of 215.

¹⁵ All of the results in Table 2 are the same for boys. These results are available from the authors upon request.

TABLE 2
GENDER, SIBLING GENDER, AND "ABILITY" TEST SCORES OF CHILDREN AND THEIR MOTHERS
(MOTHERS AGED 29+ IN 1992 FROM THE NLSY79)

	Firstborn's PPVT Score		Mother's AFQT Score		
	All mothers	Mothers with firstborn girl	All mothers	Mothers with firstborn girl	All mothers
Firstborn is a girl	-0.462 (0.580) ^a	—	-6.44 (8.26)	—	—
Firstborn has a sister	—	-2.83 (1.11)	—	-31.4 (11.8)	—
Children ever born to mother	—	—	—	—	-25.9 (3.89)
Mother is Black	-16.3 (0.953)	-16.7 (1.33)	-175 (9.90)	-171 (14.1)	-172 (9.82)
Mother is Hispanic	-13.2 (1.11)	-13.8 (1.51)	-152 (11.5)	-145 (16.1)	-147 (11.5)
R ²	.133	.155	.147	.148	.162
N	2,311	1,135	2,311	1,135	2,311

^a Standard errors in parentheses.

The problem that the any-sisters instrument is correlated with "ability" via fertility stems from the fact that the instrument incorporates information on the gender of births subsequent to the first. Recognizing the source of the problem suggests that a remedy would be to use as an instrument the gender of the first birth, which our results show conforms to the randomness criterion.¹⁶ Thus, assuming that all women have at least two births, the gender of the first birth can be used as a valid instrument for the schooling of the second child.¹⁷ In particular, the gender of

the first child cannot be correlated with the innate ability of the second child, but would be correlated with the schooling of the second child through the avenues discussed by BC.

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 experiment, we incorporate the BC assumption that the cost of schooling to a second-born girl is higher if the first born is a boy as opposed to a girl, specifically c_b and c_g respectively with $c_b > c_g$. We also adopt the assumption that ability is independent of sex. Now, suppose that with these costs, it is optimal for the girl with a firstborn brother not to attend school regardless of her ability, so that she completes only the minimum, S_0 , years of schooling. On the other hand, suppose that among girls with a first-born sister, who face a lower cost of schooling, type 1's optimally choose to

¹⁶ The same remedy, as will be shown, applies to the "twins" instrument used by Rosenzweig and Wolpin (1980a). We show below that this instrument has some limited power to explain schooling variation for siblings.

¹⁷ Of course, sample selection based on having at least two births, given that fertility is a choice, may be a source of bias. Angrist and William Evans' (1998) natural natural experiment estimate of the effects of fertility on labor supply also must assume that the two-birth minimum does not result in a choice-based sample.

attend school, thus completing $S_0 + 1$ years, while the type 2's still do not attend. In this case, the difference in expected (\ln) incomes of girls with firstborn sisters vs. firstborn brothers (experience constant) divided by the difference in their expected schooling levels, that is, the Wald estimator, is simply $f(S_0 + 1, \mu_1) - f(S_0, \mu_1)$, the return to schooling for the more able group. Thus the variation in completed schooling that arises from the sex of the first born sibling, variation that is independent of ability, provides identification of the schooling return, but only for the more able group.

3.1.3 Specification Matters: Schooling, Age and Experience

We have shown in one model that the AK and BC (suitably modified) natural experiments yield estimates of returns to schooling for different ability groups in the population, exemplifying that the instrumental approach identifies treatment effects only for the treated. Other models may yield different results for each instrument when ability and schooling returns are correlated. Again, the interpretation of what has been estimated and for whom depends on how one models behavior. If the schooling return is independent of ability, however, in principle the two experiments yield the same estimate of the schooling return in (3). But AK and BC do not estimate equation (3). Both the BC and AK specifications of the earnings function control for age rather than for work experience. Our model suggests that this seemingly innocuous specification choice leads to bias in both the AK and BC estimates and, perhaps surprisingly, 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.

Suppose that, as suggested in Jacob Mincer (1974), time spent in the labor force after completing schooling—post-school work experience—is the appropriate proxy for post-school human capital investments, as in (3). As Mincer first demonstrated, in this case controlling for age rather than experience will underestimate the return to an additional year of schooling by the earnings loss associated with having one less year of experience. Under the assumption that all post-schooling periods are spent at work, the assumption used in the schooling choice model, experience at any age X_a in equation (3) is simply $a - S - a_e$. It is easily seen that the Wald estimator for the BC experiment is given by $[f(S_0 + 1, \mu_1) - f(S_0, \mu_1)] - [g(a - a_\kappa - 1, \mu_1) - g(a - a_\kappa, \mu_1)]$. Because the additional year of schooling comes at the expense of an additional year of experience, the consequence of omitting experience from the BC specification is exactly as Mincer showed—what is identified is the difference between the returns to schooling and experience (for the more able group).

In contrast, the Wald estimator based on the quarter-of-birth experiment of AK overstates the schooling return when age is used to proxy experience. The reason is that this experimental treatment is based on an exogenous delay in schooling, and school delay not only affects schooling attainment, but must also affect post-school experience for members of the same year-of-birth cohort. To see this, we rewrite the expected (\ln) earnings at age a for the population with two ability types, where we do not control for work experience, with school entry age a_e and minimum school leaving age a_κ :

$$\begin{aligned} E(\ln y_a) = & \pi_1[f(S_0 + 1, \mu_1) + g(a - a_\kappa - 1, \mu_1)] \\ & + (1 - \pi_1)[f(S_0, \mu_2) + g(a - a_\kappa, \mu_2)]. \end{aligned} \quad (6)$$

Type 1's attend school an extra year and so have one year less of accumulated work experience at age a . Under the alternative scenario with the school entry age reduced by one year, type 1's achieve their optimal level of schooling exactly at the school leaving age, rather than one year later, and thus have accumulated one more year of work experience, $a - a_K$ years, at age a . Type 2's, as before, will be forced to obtain one year of schooling beyond their optimal level, will leave school at a_K and will have accumulated $a - a_K$ years of work experience at age a . Expected (\ln) earnings under this scenario are thus

$$E(\ln y_a) = \pi_1[f(S_0 + 1, \mu_1) + g(a - a_K, \mu_1)] + (1 - \pi_1)[f(S_0 + 1, \mu_2) + g(a - a_K, \mu_2)]. \quad (7)$$

The difference in expected (\ln) earnings divided by the difference in expected schooling ($0 \cdot \pi_1 + 1 \cdot (1 - \pi_1) = 1 - \pi_1$), the Wald estimator, is thus

$$\frac{\Delta E(\ln y_a)}{\Delta S} = \frac{\pi_1}{1 - \pi_1}[g(a - a_K, \mu_1) - g(a - a_K - 1, \mu_1)] + [f(S_0 + 1, \mu_2) - f(S_0, \mu_2)]. \quad (8)$$

Compared to the experience-constant Wald estimator, the return to schooling is overstated by the return to an additional year of experience weighted by the type one odds ratio in the population. Although reducing the age at entry forces type 2's to obtain an additional year of schooling, an additional consequence is that type 1's gain a year of work experience.

3.2 The Returns to Civilian Work Experience: The Angrist Natural Lottery Experiment

Obtaining additional schooling is not the only factor reducing "Mincer" experience for members of the same birth cohort, if that definition is restricted to civilian labor force experience. One component of the cost of military service to an individual is the loss in lifetime earnings that is incurred that

arises in part by reducing time in the civilian labor force, given that the two forms of experience are imperfect substitutes. Measuring the loss from military service is complicated by the fact that those who enter military service are not usually randomly selected. Even in a regime of military conscription draftees are not necessarily random with respect to traits that are related to earnings because some individuals choose voluntarily to enter military service and some individuals are exempted. Angrist (1990) uses the theoretical natural randomness of the Vietnam War draft lottery, based on birth dates, to assess both the total costs of military experience on earnings and to estimate the returns to civilian experience. Angrist exploits the fact that the draft lottery randomly assigns the pool of those eligible for the draft.¹⁸

The suitability of the draft lottery as an instrument requires that a random process choose who serves and, equally, who does not. However, although the draft lottery randomly chose who must enter military service, subject to some qualifying criteria, some who were not draft-eligible as determined by the lottery entered the military. There were no barriers to volunteer one's service even if not chosen, and some incentives to volunteer prior to knowing whether one would be drafted (e.g., choice of service). This "unnatural" military draft lottery thus does not strictly conform to the randomness criteria and, as Heckman (1997) has pointed out, if those who enter the military voluntarily do so in anticipation of high earnings gains from military service or have traits that

¹⁸The randomness of the draft lottery was also exploited in articles in the biomedical literature to estimate the effects of military service on post-service mortality (N. Hearst, T. B. Newman, and S. B. Hulley 1986) and IV drug use (Hearst, J. W. Buehler, Newman, and G. W. Rutherford 1991) (a true IV estimator).

are related to unobservables that affect earnings, then the lottery cannot strictly serve as a valid instrument for military service. Throughout our discussion, we will assume this problem away, as does Angrist. Even if the draft lottery were a valid instrument, its use to identify the returns to civilian experience requires still more assumptions about behavior.

Angrist obtains an estimate of the impact of military service on earnings using a Wald estimator given by $[\bar{y}^e - \bar{y}^n]/[\hat{p}^e - \hat{p}^n]$, where \bar{y}^e is mean earnings of draft eligibles, \bar{y}^n mean earnings of draft ineligibles, \hat{p}^e the proportion of draft eligibles who enter the military and \hat{p}^n the proportion of ineligibles who enter the military (voluntarily). The Wald estimates (over different birth cohorts) imply that military service reduced annual earnings of Vietnam War veterans by 15 percent. Conditional on the draft lottery randomly assigning individuals to the military, the gross earnings impact (gross of all other investments influenced by being draft-eligible) of military service is thus identified.

To quantify the importance of the earnings loss from the reduction in civilian experience, Angrist uses the same instrumental variables approach. Unlike in the later AK study, however, he adopts the Mincer earnings specification in which post-school years of experience, and not age, matter for earnings, modified such that potential experience is net of time spent in the military service, m . Because Angrist uses Social Security-based data that does not provide schooling, the earnings function for an individual at age a who spends time m in the military is written as:

$$\ln y_a = \alpha_0 + \alpha_1 S_a + \alpha_2 X_a + u_a \quad (9)$$

$$= \alpha'_0 + \alpha_2(a - m) + u'_a,$$

where the second line is obtained by substituting for potential civilian experi-

ence, $X_a = a - S_a - m - a_e$, and where the composite error contains both the unobserved earnings endowment and the omitted schooling variable, $u'_a = (\alpha_1 - \alpha_2)S_a + u_a$.¹⁹

As seen in (9), if military service is correlated with unobservables that affect earnings (the composite error), then an OLS estimator of the return to civilian experience will be biased. Now, the original motivation for using draft eligibility derived from the lottery as an instrument for military service was that m and (some permanent component of) u , ability, might be correlated. However, given the non-observability of schooling in the data set used, in order to get a valid estimate of the experience return, it is also necessary to assume that the variation in military service induced by the lottery system is uncorrelated with completed schooling. Angrist notes that such a correlation might exist because eligibility for G.I. benefits would increase the schooling of veterans relative to nonveterans. This correlation would lead to an understatement of the return to civilian experience if the return to schooling exceeds the return to experience, i.e., $\alpha_1 - \alpha_2 > 0$.²⁰ Alternatively, as Angrist notes, one can interpret the Wald estimator as a composite of the effect of military service on experience and schooling.

However, Angrist ignores the possibility that schooling and ability may be correlated due to the fact that there is another effect of military service on completed schooling that arises from the interruption of schooling (for those

¹⁹ For simplicity, the quadratic in potential experience is ignored. Angrist's specification differs slightly as well in that the length of military service, conditional on service, is estimated as a separate parameter.

²⁰ In this case, where veterans obtain more schooling than nonveterans, $a - m$ and S are negatively correlated and S increases earnings, conditional on $a - m$.

who had not completed their schooling). The effect is analogous to the effect of a delay in school entry associated with the AK natural experiment.²¹ Suppose in the model of school choice we presented a_e is redefined not as the age of school entry but as the age at which military service is completed, for those who were inducted. Assume that all individuals had completed S_0 years of schooling at the time of induction, say at age $a_e - 1$, and that it was optimal for those not inducted to have completed an additional year of schooling regardless of their ability by age a_e . Thus, those who actually served in the military have one less year of schooling at a_e than those who had not served. Now, given either a finite schooling or age horizon, as in the model presented in the previous section, those who experienced the (exogenous) interruption due to military service will optimally obtain less schooling overall. Moreover, regardless of military service, completed schooling will be higher for the more able and therefore the more able will always have less civilian work experience.²² This means that unless the schooling subsidy in the form of the G.I. Bill offset these factors the Wald estimator cannot be interpreted as a composite of experience and schooling effects; that composite effect is biased upward because of the schooling-ability correlation. The single draft-lottery in-

strument cannot be used to identify the returns to experience when both schooling and military service are correlated with ability, as in this very simple model, and schooling is unobserved.

3.3 Estimating the Returns to Schooling When Work Experience Is a Choice

The AK, BC, and Angrist studies assume that individuals work "full-time" after completing school or military experience. Given this assumption, consistent estimates of the schooling return can be obtained based on the first-birth gender instrumental variable or the quarter of birth instrumental variable when schooling is a choice and that choice is influenced by an unobservable, as depicted by the simple model, holding Mincer work experience X_a constant.²³ However, suppose that individuals, after leaving school, make labor supply decisions. Would the Wald estimators proposed by BC and AK, or suitably modified estimators that account for actual work experience, consistently estimate the schooling return? If not, how will that affect the estimates of the schooling returns for the instruments used? Note that the issue of post-school work choice is not relevant for methods that attempt to directly control for unmeasured ability using test scores or fixed-effects.

Before turning to these issues, it is useful to consider the empirical plausibility of the post-schooling "full-time" work assumption. Table 3, based on the 1979 National Longitudinal Survey of Youth (NLSY79), depicts the amount of accumulated work experience, separately for men and women, over the ten-year period from age 25 through 34

²¹ The lottery experiment differs from the school entry age experiment in that there is no analog to the mandatory minimum school leaving age.

²² The condition for obtaining an additional year of schooling for those who served in the military is given by (5). Individuals who did not serve choose between having $S_0 + 1$ and $S_0 + 2$ years of schooling. If the marginal return to an additional year of schooling is increasing in ability, one would expect that at least the more able among those who did not serve will obtain the additional year. One would also expect that a greater proportion of those who served would obtain the additional year if $f(S)$ is concave.

²³ The differential effect of military experience on earnings found by Angrist appears to have been ignored in the subsequent AK study. The 1990 Census indicates that in the male cohorts studied in AK, over 30 percent were veterans.

TABLE 3
MEAN ACCUMULATED HOURS OF WORK
BETWEEN THE AGES OF 25 AND 34, BY
SCHOOL ATTAINMENT AND SEX^a
(WEIGHTED NLSY79)

	Males	Females
HS dropout	15,741 (11,645)	6,793 (9,446)
HS graduate	18,971 (10,993)	11,892 (10,202)
Some college	18,164 (10,797)	13,929 (8,789)
College graduate	19,908 (7,662)	15,958 (7,127)

^a Standard deviation in parentheses.

for four levels of completed schooling.²⁴ For men, the largest difference in accumulated work experience is between high school dropouts and all others. Specifically, high school graduates accumulated 3,000 more hours of work over the period than dropouts; dropouts accumulated on average about 1600 hours per year. Males with some college actually accumulated 800 fewer hours of work experience than high school graduates, while those with college degrees accumulated 1,000 more hours than high school graduates. For females, accumulated work experience increases monotonically with school completion levels. High school graduates accumulated over 4,000 more hours than high school dropouts (who averaged only about 700 hours per year over the period), those with some college 2,000 more than high school graduates and those with college degrees 2,000 more than those with some college. The evidence suggests that the full-time work assumption is certainly not universally valid and that there is systematic

variation in work intensity among schooling groups.

To illustrate the consequences of introducing a labor supply decision for estimates of the schooling return based on the BC and AK natural experiments, consider an extension of the schooling model as presented above that allows for a work participation decision in each post-schooling period. For simplicity, assume that utility is additively separable in income (=consumption) and the monetary-equivalent value of leisure (home production) and that the individual maximizes the present value of lifetime utility. In each period the individual receives a wage offer that depends on schooling, experience, and ability as before, and additionally depends on a period-specific shock that is iid over periods, i.e., $\ln y_a = f(S, \mu) + g(X_a, \mu) + \epsilon_a$. The value of leisure, $b(S, X_a, \mu)$, may also depend on schooling, experience, and ability, but, for simplicity, is assumed to be deterministic.

In the terminal period, age A, the individual will work, $p_A = 1$, if the current wage offer exceeds the current value of leisure. That decision thus depends on whether the wage shock is larger than some cut-off value that itself depends on the individual's schooling, experience, and ability, namely,

$$\begin{aligned} p_A &= 1 \text{ iff } \epsilon_A > \epsilon_A^*(S, X_A, \mu) \\ &= 0 \text{ otherwise,} \end{aligned} \quad (10)$$

where $\epsilon_A^*(S, X_A, \mu) = \ln[b(S, X_A, \mu)] - f(S, \mu) - g(X_A, \mu)$. Clearly, the effects of schooling, work experience and ability on the propensity to work will depend on their relative effects on the value of leisure and on wage offers. For concreteness, suppose that neither schooling nor work experience affects the value of leisure, but that ability increases the value of leisure. In that case increasing schooling or experience will increase the propensity to work at age A. However, ability may either

²⁴ To maintain a large enough sample, the final age is the last age, between 32 and 34, that we observe the individual.

increase or decrease the propensity to work (given schooling and experience).

The participation decision in period $A - 1$ depends on a comparison of the expected present values of working and not working, namely,

$$\begin{aligned} V_{A-1}(p_{A-1} = 1 | \Omega_{A-1}) &= \exp[f(S, \mu) \\ &\quad + g(X_{A-1}, \mu) + \varepsilon_{A-1}] \\ &\quad + \delta E\max[y_A(S, X_{A-1} + 1, \mu, \varepsilon_A), b(\mu)] \quad (11) \\ V_{A-1}(p_{A-1} = 0 | \Omega_{A-1}) &= b(\mu) \\ &\quad + \delta E\max[y_A(S, X_{A-1}, \mu, \varepsilon_A), b(\mu)], \end{aligned}$$

where Ω_{A-1} , the state space at $A - 1$, consists of the elements of S , X_{A-1}, μ , and ε_{A-1} , and where the expectation is taken over the wage shock at A . The decision rule, as in the last period, is to work if the wage draw at $A - 1$ is larger than a cut-off value that depends on the non-stochastic elements of the state space at $A - 1$, that is,

$$\begin{aligned} p_{A-1} &= 1 \text{ iff } \varepsilon_{A-1} > \varepsilon_{A-1}^*(S, X_{A-1}, \mu) \\ &= 0 \text{ otherwise.} \end{aligned} \quad (12)$$

The extent to which the AK and BC natural-experiment instrumental estimates provide consistent estimates of schooling returns depends on whether participation and schooling are uncorrelated, on whether schooling attainment affects participation choices beyond the mechanical effect on Mincer-type experience. It is clear that as long as the wage offer is monotonically increasing in S , the cut-off value of the wage shock is declining in completed schooling. Because an increase in S increases the current wage, the future wage, and the propensity to work in the future, an increase in S also increases the propensity to work in the current period. Unlike the effect of schooling, an increase in the effect of ability on the propensity to work is ambiguous; however, if ability increases the (positive) effect of experience on wages, the propensity to work will be more likely to increase with abil-

ity in period $A - 1$ than in period A .²⁵ If we continue to solve backwards, by induction the effect of schooling on the propensity to work will be positive at all ages. Thus, because experience at any age is simply the cumulation of past participation decisions, the deficit in work experience of the more schooled that results from their delayed entry into the labor market will diminish once they do enter and continue to diminish with age. And, it is possible that work experience of the more schooled may exceed that of the less schooled after some age, as is consistent with the data in Table 3.

One effect of this endogenous accumulation pattern is to make the bias in the BC and AK instrumental variables estimators that control for age but not for actual work experience depend on the age at which wages are measured. To see that, consider our previous example of the AK experiment with two ability types. It is easy to show that the Wald estimator based on the difference in earnings at age a for the population with entry age $a_e - 1$ vs. the population with entry age a_e (assuming that everybody works at age a) is:

$$\begin{aligned} \frac{\Delta E(\ln y_a)}{\Delta S} &= \frac{\pi_1}{1 - \pi_1} \{g[X_a(S_0 + 1, a \\ &\quad - a_\kappa, \mu_1, \varepsilon_{a-1}), \mu_1] - g[X_a(S_0 + 1, a \\ &\quad - a_\kappa - 1, \mu_1, \varepsilon_{a-1}), \mu_1]\} \\ &\quad + \{g[X_a(S_0 + 1, a - a_\kappa, \mu_2, \varepsilon_{a-1}), \mu_2] \\ &\quad - g[X_a(S_0, a - a_\kappa, \mu_2, \varepsilon_{a-1}), \mu_2]\} \\ &\quad + [f(S_0 + 1, \mu_2) - f(S_0, \mu_2)], \end{aligned} \quad (13)$$

where ε_{a-1} represents the vector of wage shocks through age $a - 1$, all of which are known at a . The first term reflects the fact that the more able, and thus more

²⁵ The effect of an increase in experience is more complicated. Even if the wage is monotonically increasing in experience the effect on participation of an increase in experience depends on the second derivative of the g function. Because of the finite horizon, for given experience the propensity to work is lower at $A - 1$ than at A .

schooled, have one more year of potential experience when they are permitted to enter school a year earlier. It is the same term as in the model imposing the exogenous full-time work assumption. The second term reflects the fact that when the less able are forced to attend school for an additional year, they now have the option of choosing a different sequence of participation. The simple model of participation presented above implies that they will be more likely to participate at each age, and that the difference in work experience will be negative at early ages and possibly positive at later ages. Thus, the magnitude and even the direction of the bias in the Wald estimator will be age-dependent. Notice that controlling for potential (Mincer) experience rather than age as in AK when participation is a choice corrects for the bias associated with the first term to the extent that an additional year of potential experience leads to an additional year of actual experience. Replacing age by potential experience cannot correct for the bias from the second term because the second term only arises because potential and actual experience differ.

Replacing age by actual work experience also does not eliminate the bias in the natural natural experiment-based Wald estimator of the estimated return to schooling (the third term in (9)) if participation decisions are dependent on ability, as suggested in the model. It is possible to develop a test for the exogeneity of experience using the BC and/or AK natural experiments based on linearized versions of the decision rules from the model. Specifically, consider the statistical framework that corresponds to the AK experiment (where all equations are in deviation from mean form):

$$\begin{aligned} \ln y_a &= \alpha_1 S + \alpha_2 X_a + \mu + \varepsilon_a, \\ S &= \beta_1 a_e(q) + \beta_2 \mu + \varepsilon_s, \\ X_a &= \gamma_1 S + \gamma_2 \mu + \varepsilon_x, \end{aligned} \quad (14)$$

where q is quarter of birth. Notice first that as long as γ_2 is not zero, the return to schooling is not identified; there is essentially one instrument, the age of school entry (which is not related to ability) and two endogenous variables, schooling and experience. Although we cannot identify the earnings function parameters in (10), it is still possible to test whether experience is correlated with ability (assuming that schooling is correlated with ability), that is, whether γ_2 is zero, using a modified Wu-Hausman test. First, note that estimating the schooling equation provides a consistent estimate of β_1 . Next, form the residuals from the estimated schooling equation, $S - \hat{S}$, which must be correlated with ability under the maintained assumption that β_2 is not zero. Finally, estimate the experience equation, including S and $S - \hat{S}$. If the coefficient on the residuals is non-zero, then γ_2 must be non-zero. An analogous set of equations substituting a firstborn sex dummy for age of school entry describes the BC experiment.

Table 4 presents regression estimates of the p -values for the coefficient on the schooling residual for men and for women using data from the NLSY79. Experience is defined as actual total hours worked from age 18 until the last year the individual is observed in the data (for all individuals observed at least through age 25). The first column for each sex presents the estimates using the AK instrument, quarter of birth interactions with state of birth. The second column presents estimates using the modified BC instrument, the sex of the firstborn sibling and the third column presents estimates based on using the two instruments together. The results indicate that for males the exogeneity of experience is rejected based on either the AK or the BC instrumental variables, but not on the combination of the two sets of instruments. For females,

TABLE 4
TESTS OF THE ENDOGENEITY OF WORK EXPERIENCE USING NATURAL INSTRUMENTS
(NLSY79)

	Males			Females		
	AK ^a	BC ^b	AK + BC ^c	AK ^a	BC ^b	AK + BC ^c
P-value of schooling residual coefficient	.008	.031	.277	.666	.733	.097
No. of observations	2,592	1,250	1,245	2,793	1,341	1,336
First-stage R ²	.161	.172	.271	.120	.110	.187

^a Instrumental variables are quarter of birth interactions with state of birth. Additional regressors are state of birth, state of residence at age 14, a black dummy, a Hispanic dummy and age.

^b Instrumental variable is sex of firstborn sibling. Additional regressors as above.

^c Instrumental variables are quarters of birth interactions with state of birth and sex of firstborn sibling. Additional regressors as above.

however, the exogeneity of experience is rejected only with the combined set of instruments. Although the lack of robustness of these results to the choice of instruments does not lead us to a firm conclusion, the results are certainly not inconsistent with the hypothesis that work intensity, for both males and females, is subject to choice.²⁶ In sum, in a world in which schooling and work choices are made optimally by individuals heterogeneous in ability, as characterized even in a very simple model and for which there is some evidence, neither the AK nor the BC natural natural instrument as implemented identifies the returns to schooling. And the model suggests that the estimates obtained using the different instruments but on the same sample will result in different estimates of schooling returns if such returns vary by ability and, even if not, because of the use of age as a proxy for work experience. Due to the latter, the estimates will further differ if there are age differences in the samples used.

²⁶ One of the reasons for the difference in results may be the fact that the date of birth of a child does not strictly conform to the criterion of randomness, as noted.

3.4 *The Returns to Schooling and Work Experience When Both Are Choices: The Natural Human Cloning Experiment*

The natural natural experiment approach exemplified by Angrist, AK, and BC exploited random events occurring in nature that affect schooling or (military) experience decisions but are unrelated to ability. The use of these natural natural experiments is an attempt to mimic the conceptual experiment of forcing otherwise identical people to invest differently in human capital. A major problem with this approach, as we have seen, is the limited number of instruments, for if both schooling and post-school investments, as proxied by actual work time, are correlated with ability (and with each other), then at least two instruments are required. Butcher and Case and Angrist and Krueger must assume that work experience is exogenous in their studies, an assumption often made in earnings studies, because they only have one instrument. The exogeneity of work experience is thus a necessary assumption for their methodology in addition to the randomness of the instrument. The

military draft lottery, birth date, and the sex of a firstborn child could potentially be used together as instruments to estimate the returns to schooling, and military and civilian work experience from a more realistic specification of the earnings function, but this has not been carried out. However, there is evidence that other dimensions of labor force experience than cumulated years matter for earnings, such as the timing and length of work interruptions. If these work characteristics also reflect choices that are related to ability and schooling, additional natural instruments would be needed to obtain consistent estimates of schooling returns.

An alternative experimental approach is to identify people who are actually identical but who, for whatever reason, obtain different levels of human capital. Of course, the underlying problem is that we do not observe abilities and so cannot *a priori* identify people who are the same. This fundamental observational problem has led researchers to exploit within-pair differences of "identical" twins starting in the late 1970's (Jere Behrman, Z. Hrubec, Paul Taubman, and Terrence Wales 1980; Orley Ashenfelter and Krueger 1994; Behrman, Rosenzweig, and Taubman 1994; Paul Miller, Charles Mulvey, and Nick Martin 1995; Ashenfelter and Cecilia Rouse 1998; Behrman and Rosenzweig 1999; and Rouse 1999).²⁷ Monozygotic (MZ) twins are identical at conception. It has been argued that the differential levels of schooling obtained by monozygotic twins, which cannot be due to differences in genetic endowments, corresponds exactly to the conceptual experiment described above.

A major advantage of natural human

cloning, in contrast to the Angrist, AK, and BC natural experiments, is that the twins experiment is robust to the introduction of endogenous labor supply choices, or to the addition of any number of endogenous choices that may affect outcomes of interest. To see why, note that the cloning experiment can be placed within an instrumental variables framework. The difference in the schooling of twins, because it is independent of genetic endowments, can serve as an instrument for the school attainment of either twin. Similarly, the difference in the work experience of twins can serve as an instrument for either twin's work experience. Thus, unlike the previous experiments, the cloning experiment supplies as many instruments as there are endogenous variables.

A number of assumptions are needed to achieve identification of the parameters of interest in the earnings equation given by (3) based on the cloning experiment. To see how identification is achieved, consider (3) rewritten in linear form for a pair of identical twins 1 and 2 born to family j :

$$\ln y_{ij}^M = \alpha + \beta S_{ij}^M + \delta X_{ij}^M + \mu_j^M + v_{ij}^M, \quad i = 1, 2 \quad (15)$$

where the superscript M refers to MZ twins, X is actual work experience, and the returns to schooling and experience are β and δ , respectively. We can also write in linear form the relationships between schooling, work experience, and the unmeasured common ability endowment of the twins μ_j^M , other common family factors given by f and twin-specific factors u_{Sij} and u_{Xij} .

$$S_{ij}^M = b_s \mu_j^M + f_{Sj} + u_{Sij}, \quad (16)$$

$$X_{ij}^M = b_x \mu_j^M + f_{Xj} + u_{Xij}. \quad (17)$$

Estimation of (15) by OLS for any individual, a twin or not, given (16) and (17), results in biased estimates of β and δ . Put another way, there are more

²⁷ The first published study of earnings determinants based on twins appears to be Behrman and Taubman (1976), which presents preliminary results to the study Behrman, Hrubec, Taubman, and Wales 1980.

parameters than there are empirical moments in the data. Within-MZ-twin estimates are obtained by differencing (15) across the twins, which sweeps out the common unobserved endowment component μ_j^M . Given the assumption that the ability endowment is common across the twins, that the twin-specific factors determining schooling and experience u_{Sij} and u_{Xij} are orthogonal to the errors v_{ij}^M in the earnings equation, and that these "random" factors are drawn from the same distribution for each twin, the within-family variances and covariances based on the twin differences in terms of the parameters of the model are:

$$\text{Var}(\Delta \ln y^M) = 2\beta^2 \sigma_{Su}^2 + 2\delta^2 \sigma_{Xu}^2 + 2\sigma_v^2, \quad (18)$$

$$\text{Var}(\Delta S^M) = 2\sigma_{Su}^2, \quad (19)$$

$$\text{Var}(\Delta X^M) = 2\sigma_{Xu}^2, \quad (20)$$

$$\text{Cov}(\Delta S^M, \Delta X^M) = 2\text{cov}(u_S, u_X), \quad (21)$$

$$\begin{aligned} \text{Cov}(\Delta \ln y^M, \Delta S^M) &= 2\beta \sigma_{Su}^2 \\ &\quad + 2\delta \text{cov}(u_S, u_X) \text{ and} \end{aligned} \quad (22)$$

$$\begin{aligned} \text{Cov}(\Delta \ln y^M, \Delta X^M) &= 2\delta \sigma_{Xu}^2 \\ &\quad + 2\beta \text{cov}(u_S, u_X). \end{aligned} \quad (23)$$

As can be seen, there are now as many empirical moments as there are parameters and the equations are linearly independent. There is exact identification. Adding more determinants of earnings potentially correlated with the endowment to (15), given the same set of assumptions for these additional determinants, adds as many additional empirical moments as additional parameters. Thus, as noted, under the maintained assumptions, there is no limit to the number of observed earnings factors, potentially correlated with ability, that can be added for which returns can be identified.

Only one of the twins human cloning studies so far has employed information on actual work experience. As a consequence, under the assumption that all workers work full-time after school and

that post-school work experience X_a is the appropriate measure of post-school human capital, all but one of the set of estimates from the cloning studies are subject to the downward "Mincer" bias. Twins are always the same age, so the twin with the greater (lesser) amount of schooling must have less (more) work experience. All but one of the twins-based "schooling returns" estimates thus actually identify the *difference* between the returns to schooling and work experience, not the returns to schooling given work experience. Relatedly, the "ability" bias identified from such studies is not the bias in the returns to schooling but the bias in the difference in schooling-experience returns.

Ironically, then, only if post-school work choices vary among twins such that schooling and age do not completely determine work experience can twins-based studies be used to identify the returns to schooling, as long as information on actual work experience is collected and used in the earnings specification. In fact, actual work experience appears to differ significantly across identical twins. In data collected in 1994 by Behrman, Rosenzweig, and Taubman based on all twins born in Minnesota from 1936–55, 35 percent of (215) male-twin pairs had differences of three or more years of actual full-time work experience beyond that accounted for by differences in completed schooling. The comparable figure for the 339 female-twin pairs in the data is 64 percent, with more than half of those exhibiting differences of eight or more years of actual full-time work experience net of schooling attainment differences.

With appropriate information on work experience, there are two remaining problems with the cloning experiment approach, however. The first is that the existence of errors in measurement in any of the earnings determinants is not

only a barrier to identification, as one or more error variances are introduced but no more empirical moments, but the resulting biases due to measurement error are larger when within-twin estimators are used compared with estimates based on individuals (John Bishop 1976; Griliches 1979). This is because the bias from measurement error is greater the higher the within-pair correlation of the "true" variables. These are especially high among identical twins. Behrman and Rosenzweig (1999) report for their sample of Minnesota-born twins that the within-pair correlation in schooling is .74 for identical twins. This compares with a correlation of .52 for nonidentical twin pairs.²⁸

With respect to the measurement-error problem, Ashenfelter and Krueger (1994) proposed using reports by each twin of his or her twin's schooling as an instrument to eliminate the bias due to measurement error in schooling. With a sufficient set of restrictions on the orthogonality of the measurement errors in these reports, these "instruments" add more empirical moments than theoretical parameters so that, with measurement error confined to schooling, the returns to true schooling are identified. All recent studies using twins to identify the returns to schooling have used a variant of this approach to correct for measurement errors in schooling.²⁹ Of course, there is nothing "natural" about the instruments employed in twins studies to correct for measurement error. Without knowledge of true schooling attainment, it is not

²⁸ They estimate that if the proportion of the total variance in schooling that was measurement error was ten percent, the bias towards zero in estimates based on individuals would be nine percent but would be 35 percent for the within-MZ estimator.

²⁹ In Behrman, Rosenzweig, and Taubman (1994), the report by each twin's son of the twin's schooling attainment is used as an instrument to correct for measurement error.

possible to test the orthogonality assumptions about the correlation in measurement errors across the twins or across own and cross-twin reports for a twin that are necessary for identification.

The second potential problem with the cloning experiment, and it is a major one, is that the twin-specific errors u_{ij} may not be orthogonal to earnings, as assumed. One reason is that the unobservables that affect schooling decisions and wages (productivity) are not necessarily only genetic in origin, i.e., monozygotic twins may be identical at conception, but not identical at birth or as children. For example, monozygotic twins differ in birth weight, which may indicate that such twins do not face identical environments within the womb. In a sample of 1534 monozygotic twin pairs from the Minnesota Twin Registry, described in Behrman, Rosenzweig, and Taubman (1994), the average absolute value of the difference in birth weights between "identical" twins is 10.5 ounces. The corresponding figure for same-sex nonidentical twin pairs ($N = 1357$) is 11.2 ounces. Thus, monozygotic twins may have unequal mental and/or physical capacities that manifest themselves at birth and beyond. In fact, Behrman, Rosenzweig, and Taubman (1994) report a significant positive correlation between the difference in the birth weights of "identical" twins and subsequent differences in their schooling attainment and thus earnings.³⁰ Additionally, during childhood

³⁰ There are few studies of the relationship between birth weight and adult outcomes. One recent exception is Richard Strauss (2000), who finds that low birth weight predicts lower academic achievement and income at age 26 based on the 1970 British Birth Cohort. The problem with this study, and all prior studies of the consequences of birth weight variation, however, is that the separate effects of increasing the nutrition received by a fetus, which accounts for birth weight variation within MZ twin-pairs, and possible genetic influences are not distinguished.

seemingly random events may occur to one twin that have long-lasting, if not permanent, effects on characteristics that affect both schooling and earnings. For example, one twin may suffer an accidental physical or mental impairment that hinders schooling attainment and earnings capacity. The permanent unobservable in these cases is not fully captured by a genetic endowment. Such health shocks or in-womb nutritional differences are likely to result in a positive bias in human capital returns that is not eliminated by differencing across MZ twins.

Even if monozygotic twins were identical at birth and substantial health shocks are relatively rare, economic theory that recognizes that investment in human capital is a decision suggests that unobservables specific to twins will be systematically related to their investment decisions and to their earnings. It is straightforward to provide an example in which optimizing behavior induces a downward-biased estimate from twins-based estimators. Recall that the original schooling choice models assumed that either wages were deterministic, with exogenous post-schooling full-time work, or that wage offers were received only after completing schooling. Suppose, instead, the original formulation is changed so that a random wage offer is received at the beginning of the first period, when the individual is deciding on whether or not to attend school. Also, assume that this wage, if accepted, is fixed over the working life.³¹ If the individual rejects the offer and attends school for that period, another wage is drawn (independently of the first period wage draw) at the beginning of the next period that is also permanent. Denoting the stochastic

component of the wage draw as ε , the expected present values of earnings for the school attendance choices are:

$$V_1(s_1 = 1 | S_0) = \exp[f(S_0 + 1, \mu)] E_1\{\exp[\varepsilon_2]\} \\ \cdot \sum_{a=0}^A \beta^{a+1} \exp[g(a, \mu)] - c, \quad (24)$$

$$V_1(s_1 = 0 | S_0) = \exp[f(S_0, \mu)] \exp[\varepsilon_1] \\ \cdot \sum_{a=0}^A \beta^a \exp[g(a, \mu)].$$

Notice that these values differ from the corresponding ones in (3) by the actual (permanent) wage shock if the individual does not attend school in period one and by the expected wage shock if the individual attends school. The decision rule, as before, is to attend school if $V_1(s_1 = 1 | S_0) \geq V_1(s_0 = 1 | S_0)$. However, because in this case the value of not attending school depends on the actual wage draw, the individual will attend school only if the wage draw is below a cut-off value, which depends on the rate of interest, the cost of schooling, the parameters of the wage function (f and g), and the expectation of the (exponential of the) wage draw, i.e.,

$$s_1 = 1 \text{ if } \varepsilon_1 < \varepsilon_1^* \{r, c, f, g, E(\exp(\varepsilon_2))\}, \quad (25)$$

$$s_1 = 0 \text{ otherwise.}$$

Thus, if one of a twin pair does not attend school and the other does attend, for the former we only observe wages that satisfy the condition that $\varepsilon_1 > \varepsilon_1^*$, while for the latter we observe wages that span the entire wage distribution. The average (accepted) wage for twins with lower schooling will therefore overstate the average (offered) wage and thus the return to schooling will be understated.

Thus, despite the potential advantage of natural human cloning in providing as many instruments as there are measured earnings determinants, it is difficult to conclude from the human cloning

³¹The assumption of a permanent wage is only illustrative, although it is necessary for the argument that there be a persistent component.

experiment, as implemented to date, what the true returns to schooling are or even to define their bounds. This is because of (i) the possibility that schooling decisions are influenced by persistent wage shocks, which biases schooling return estimates based on within-twin estimators downward, (ii) the omission of work experience, which evidently differs across twins and is correlated with schooling, and (iii) the existence of at least early environmentally-determined differences in endowments across clones that may lead to upward biases in schooling. And, given that monozygotic twins differ neither in birth date nor gender, twins-based estimates cannot be used in conjunction with the AK, BC, or Angrist natural instruments to circumvent some of these sources of bias.

4. Estimating Determinants of Consumption and Labor Supply Using Weather Variation

4.1 Permanent and Transitory Income Elasticities of Consumption and Savings

One of the most widely employed and tested models of life-cycle savings and consumption is the "permanent income" model. In recent years, researchers have been interested in providing evidence on the extent to which capital markets are incomplete, as is indicated if consumption is "excessively" sensitive to contemporaneous income. Life-cycle models of consumption which admit to uncertainty about future income flows have the feature that the responsiveness of consumption (and savings) to income depends on the extent to which fluctuations in income are unanticipated and transitory. Distinguishing between transitory and permanent income components and identifying their effects are the principal challenges of this literature. Data sets

do not provide measures of income that conveniently identify incomes by these theoretical concepts. Moreover, fluctuations in incomes may reflect the choices of agents—because income at any given point in the life-cycle may reflect past investment and savings decisions, correlations between income changes and consumption may not provide much insight into these models. For example, in agricultural populations farmers plant seeds or invest in equipment that affects the level and variability of incomes (Rosenzweig and Hans Binswanger 1993), and these may reflect preferences (e.g., for risk) that also affect consumption choices.

Five studies have used weather variables, in the context of farm households, as a means of identifying the effects of transitory and permanent components of income (Wolpin 1982; Christina Paxson 1992; Hanan Jacoby and Emmanuel Skofias 1998; Anjini Kocchar 1999; and Elaina Rose 1999). Weather has desirable features for the analysis of income effects: weather events have significant effects on farm income; weather events cannot be affected by the behavior of the farmers themselves and satisfy the criterion of randomness; weather distributions are characterized by stationarity over periods of time relevant to the study of income effects on consumption so that the distinctions between permanent and transitory are meaningful; and long time-series of data describing daily rainfall, which are available in many countries of the world, enable the precise estimation of the permanent parameters describing the moments of the weather distributions.³² Despite these

³² Note that there can be more than one measure of weather (e.g., daily rainfall, the timing of the onset of the rainy season, number and length of dry spells) that affects output and multiple moments describing the distribution of each measure.

advantages, however, the estimates of income effects based on weather events have been based on theoretical frameworks, implicit or explicit, that have employed strong and untested assumptions about the operation of rural labor markets, preferences, and farm technology, assumptions that appear to be necessary for identification of the effects of interest. Indeed, recent work has used weather information to look at the direct relationship between income risk, income shocks, and labor supply (Rose forthcoming; Kocherlakota 1999).

4.1.1 *Income and Consumption without Labor Supply*

To illustrate what weather-based instruments identify and theoretical issues that weather-based instrumental estimates must confront, we construct a simple farm model that incorporates transitory and permanent income effects. We begin with the simplest two-period case in which there are no inputs to production other than weather and only consumption provides utility. Thus we initially abstract from the choice of labor supply and non-farm labor earnings. Each farm household maximizes its expected present discounted utility flow from consumption, $E_1[U(c_1) + \beta U(c_2)]$, where $\beta = \frac{1}{1+r}$ is the subjective discount factor, subject to income constraints. In each period the farmer receives farm income from his output and an exogenous amount of "assured" income y_0 that is known in advance.³³ Farm income is stochastic, such that farm output $y_t = f(\omega_t)$, with ω_t a measure of weather at t (e.g., the amount of rainfall). Weather is random and iid over time. A weather realization is drawn

³³We assume the existence of a permanent assured income flow as a means of contrasting its impact on savings to that of income effects generated by differences in permanent weather. Extending the model to more than two periods would not change any essential points of the argument.

each period independently from a known distribution with finite moments. To distinguish permanent and transitory weather, let $\omega_t = \bar{\omega} + \varepsilon_t$, where $\bar{\omega}$ is mean weather and ε_t is an unforecastable weather shock (iid over time).

Farmers can save and borrow, but they cannot purchase insurance and do not make bequests. With a unit output price, the budget constraints in the first and second periods are thus:

$$\begin{aligned} c_1 + a_2 &= y_1 + y_0 \\ c_2 &= a_2(1+r) + y_2 + y_0 \end{aligned} \quad (26)$$

where a_2 is savings in period 1 and r is the interest rate.³⁴ In the second period, because all income is consumed, having an extra dollar of income from any source—whether due to living in an area in which mean rainfall is higher, getting a favorable weather shock, or having a higher amount of assured income—will increase second-period consumption by one dollar. In period one, however, consumption effects will differ depending on whether the assured income level is increased (a permanent income change) or whether only farm income increases in the period. In the first period the farm household's problem is to choose the level of assets to carry over to period two that maximizes expected discounted lifetime utility, $U[f(\omega_1) + y_0 - a_2] + \beta E_{\omega} U[(1+r)a_2 + f(\omega_2) + y_0]$.

The first-order condition is the standard Euler equation,

$$U'_1 = \beta(1+r)E_{\omega}U'_2. \quad (27)$$

It is straightforward to show that the effect of a change in the assured income flow on first-period savings is

$$\frac{\partial a_2}{\partial y_0} = \frac{U''_1 - \beta(1+r)E_{\omega}U''_2}{U''_1 + \beta(1+r)^2E_{\omega}U''_2}. \quad (28a)$$

³⁴Given the assumption of no bequests, initial assets (a_1), other than land, is zero. Thus, assets carried into period two is identical to savings from period one.

³⁵We are assuming the Inada conditions hold.

In contrast, the instrumental variables estimator of the effect of permanent income on savings that uses variation (say, over space) in permanent weather characteristics, in this case the mean ($\bar{\omega}$) of the weather distribution, is

$$\begin{aligned} & \frac{\partial a_2 / \partial \bar{\omega}}{\partial y_1 / \partial \bar{\omega}} \\ &= \frac{U''_1 \cdot \partial y_1 / \partial \bar{\omega} - \beta(1+r)E_{\omega}[U''_2 \cdot \partial y_2 / \partial \bar{\omega}]}{U''_1 + \beta(1+r)^2E_{\omega}U''_2] \cdot \partial y_1 / \partial \bar{\omega}}. \end{aligned} \quad (28b)$$

Comparing (28a) and (28b), it is clear that the instrumental variables estimator of the permanent income effect on savings is equivalent to the true permanent income effect only if weather affects output linearly, i.e., only if $y_t = \gamma(\bar{\omega} + \varepsilon_t)$.³⁶

Unlike permanent weather, the instrumental variables estimator based on a change in transitory weather in period one does identify the effect on savings of a one-dollar increase in transitory income, say from a one-dollar increase in income that occurred in period one only, namely

$$\begin{aligned} \frac{\partial a_2 / \partial y_1}{\partial y_1 / \partial \varepsilon_1} &= \frac{\partial a_2 / \partial \varepsilon_1}{\partial y_1 / \partial \varepsilon_1} \\ &= \frac{U''_1}{U''_1 + \beta(1+r)^2E_{\omega}U''_2}. \end{aligned} \quad (28c)$$

If the model is correct, as seen in (28a) and (28c), the estimated effect on savings of a one-dollar increase in transitory income in period one should exceed the effect of an anticipated permanent one-dollar increase in income.³⁷

³⁶ Using second period output in the first stage of the IV estimator, or any fixed-weighted average of output in the two periods, would not change the general result, although the bias in the estimator would be different for different measures of output.

³⁷ In the special case of quadratic utility, and where the rate of interest and the rate of time preference are both zero, one-half of any increase in transitory income is saved (and consumed in the next period) while none of any increase in permanent income is saved.

Wolpin used differences across Indian districts in the mean of the rainfall distribution in order to estimate the permanent income effect for rural farmers.³⁸ He did not use an explicit optimizing model, instead assuming both a consumption function that was linear in permanent income and an income-generating function that was linear in the rainfall mean. His finding of a permanent income elasticity close to unity could, in the context of the optimizing model presented above, be artifactual.³⁹

Wolpin used contemporaneous village-level indicators of weather deviations as instruments to estimate the transitory income effect. He found those effects to be statistically indistinguishable from zero, although point estimates were quite imprecise. Paxson used household survey data on farmers in rural Thailand and time-series of regional rainfall to estimate the effects of transitory income on savings. In particular, Paxson jointly estimated the effects on savings and on income of deviations in contemporaneous rainfall from regional means for each of four crop seasons. She then tested what she called a "strong version" of the permanent income hypothesis—that the four season-specific

³⁸ Note that spatial differences in permanent characteristics of the weather distribution could be poor instruments for differences in permanent income levels if property values reflect the returns to weather and each generation pays for these returns. In the worst case, each generation purchases land at the market price with those in better-climate areas paying a higher price. In that case, permanent climate characteristics would have no explanatory power in the determination of income variation across space. Wolpin finds, however, that these variables produce precise estimates of income effects on consumption in India.

³⁹ Paxson used measures of land holdings to predict permanent income. This latter instrument does not appear to qualify as a natural natural experiment, as, for example, households with larger accumulated land holdings may be more likely to be savers.

estimated weather-variable effects on savings and on income were identical, equivalent to the IV estimator of the transitory income effect on savings being equal to one. Her estimates could not reject this hypothesis.⁴⁰

4.1.2 Endogenous Labor Supply, Labor Markets, and Savings

Both Paxson and Wolpin used gross farm income net of the cost of paid inputs as the measure of farm income. However, farm households often use "unpaid" family labor instead of or in addition to hired labor. It is obvious that if the household uses only hired labor inputs, the IV estimators in (28b, 28c) would have to be based on net revenues, i.e., the value of output net of the cost of the hired labor.⁴¹ When family labor is also used, however, the IV estimator based on the income measure used in the Paxson and Wolpin studies yields biased estimates of income effects on consumption. The direction of the bias and the appropriate measure of income to use in order to retrieve parameters of interest depend on what is assumed about family labor, and more generally on the extent to which labor markets are complete.

We now augment the simple farm consumption model to allow the household a choice of how to allocate the labor of its family members to show that the properties of the IV estimator using random weather as an instrument for income depend importantly on assumptions about the heterogeneity and marketability of farm family labor. We fo-

⁴⁰ Interestingly, the IV estimates using landholdings to test the strong version of the permanent income hypothesis led to rejection, because larger land holders appeared to save too much. This is consistent with wealth and savings both being related to unobserved propensities to save.

⁴¹ Wolpin notes that profits should be defined as net of the implicit cost of family labor, but, as with Paxson, did not have data on family labor inputs.

cus on the identification of transitory income effects both because expressions are simpler and because the permanent income effect cannot be correctly estimated using permanent weather except under strong simplifying assumptions about its effect on production. For convenience, we assume that the household's utility in each period depends on the total amount of leisure consumed by its household members and total household consumption, $U_t(C_t, l_t)$. With respect to production, we assume that there are two types of labor inputs—manual labor and supervisory labor. Family labor and hired labor supplied to manual tasks are assumed to be perfect substitutes in production, while supervisory labor must be supplied by household members alone. Output in period t is thus $y_t = f(h_t + m_t, s_t, \omega_t)$, where h_t is the amount of hired manual labor, m_t is the amount of family manual labor, and s_t is the amount of family supervisory labor. Hired labor can be purchased competitively at wage w . The total amount of household labor is normalized to one. We assume that labor utilization decisions take place after the realization of the weather shock and that the wage is neither affected by these decisions nor by the weather shock.

Assuming no bequests (other than land and other fixed inputs) as before, the household's problem in period two is to choose the amounts of hired and family manual labor and the amount of supervisory labor to use in farm production. Defining maximized utility in period two as

$$\begin{aligned} V_2(a_2, \varepsilon_2) = \max_{h_2, m_2, s_2} & U[(1+r)a_2 \\ & + f(h_2 + m_2, s_2, \bar{\omega} + \varepsilon_2) + y_0 \\ & - wh_2, 1 - m_2 - s_2], \end{aligned} \quad (29)$$

the Kuhn-Tucker first-order conditions are:

$$\begin{aligned}
 h_2: f_1 &= w, \\
 m_2: \frac{\partial U / \partial l_2}{\partial U / \partial c_2} &\geq f_1, \\
 s_2: \frac{\partial U / \partial l_2}{\partial U / \partial c_2} &= f_2,
 \end{aligned} \tag{30}$$

There are two important cases to consider: (i) when the marginal rate of substitution between leisure and consumption equals the market wage for manual labor ($m_2 > 0$), which in this case means that family labor and hired labor are both marketed, and (ii) family labor is only used to supervise and supervisors cannot be hired ($m_2 = 0$), in which case the relevant first-order condition is an inequality and no market wage exists for family labor.

Case I: In this case, the marginal product of (total) manual labor and the marginal product of supervisory labor are both equal to the market wage of hired labor. Thus, their input demand functions depend only on the market wage and the weather realization, i.e., $h_2 + m_2 = hm(\omega_2, w)$ and $s_2 = s(\omega_2, w)$, as well as technology parameters.⁴² Unlike family-supplied supervisory labor, however, family manual labor depends not only on weather and the market wage, but also on the level of assured income, i.e., $m_2 = m(\omega_2, w, y_0)$. As long as family leisure is a normal good, an increase in assured income will reduce the amount of family-supplied manual labor.⁴³ Weather has two effects on family manual

⁴² The effects of the weather realization on input demands are:

$$\frac{\partial hm_2}{\partial \omega_2} \propto y_{23}y_{12} - y_{13}y_{22} \text{ and } \frac{\partial s_2}{\partial \omega_2} \propto y_{13}y_{21} - y_{23}y_{11}.$$

⁴³ The pure income effect on family manual labor,

$$\frac{\partial m_2}{\partial y_0} \propto u_{2u11} - u_{1u21},$$

is negative as long as leisure is a normal good. Because supervisory labor is independent of assured income, an increase in leisure due to an increase in assured income must be exactly offset by a reduction in family manual labor.

labor. First, a favorable weather outcome that increases output reduces m_2 through the income effect. Second, weather affects the use of supervisory labor and thus leisure, which changes the leisure-consumption marginal rate of substitution. However, because the latter is equal to the (unchanged) wage rate for hired labor, family manual labor must change one-for-one in the opposite direction. Specifically,

$$\frac{\partial m_2}{\partial \omega_2} = -\frac{\partial s_2}{\partial \omega_2} + f_3 \frac{\partial m_2}{\partial y_0}.$$

Consumption increases with assured income, but less than dollar for dollar, as some of the increased income is taken in leisure; specifically,

$$\frac{\partial c_2}{\partial y_0} = w \frac{\partial m_2}{\partial y_0} + 1.$$

The effect of the weather realization on consumption is proportional to this pure income effect,

$$\frac{\partial c_2}{\partial \omega_2} = f_3 \frac{\partial c_2}{\partial y_0}.$$

Given this relationship, the IV estimator of the pure income effect on consumption must be based not on revenues net of only the cost of hired labor inputs, but on revenues net of the cost of hired labor and family labor inputs, where the latter are valued at the market wage for hired manual labor. Thus, the effect of a change in last period income on consumption (the same as the effect of assured income) is

$$\frac{\partial c_2}{\partial y_2} = \frac{\partial c_2 / \partial \varepsilon_2}{\partial(y_2 - wh_2 - wm_2 - ws_2) / \partial \varepsilon_2}. \tag{31}$$

It is tedious but straightforward to demonstrate that the IV estimator of the effect on first-period savings of a transitory change in first-period income also requires that this same income measure be used (as in period two, we are assuming that the family supplies a positive amount of labor to the manual

task in period one). For example, the effect on savings of a transitory change in income in period one is

$$\frac{\partial a_2}{\partial y_1} = \frac{\partial a_2 / \partial \varepsilon_1}{\partial (y_1 - wh_1 - wm_1 - ws_1) / \partial \varepsilon_1}. \quad (32)$$

These results imply that in the Wolpin (1982) and Paxson (1992) studies which used crop revenues net of only hired inputs as the measure of income, the weather-based estimates of savings effects are biased upward. If leisure is a normal good, then favorable weather must reduce the total amount of family supplied labor and, therefore, the effect of weather on profits net of the cost of hired labor will underestimate its effect on profits net of the cost of hired plus family labor. Thus, the IV estimator of the impact on savings of a transitory income change will be overstated. Conversely, the impact on consumption will be understated. The study by Kochhar, based on longitudinal data from forty households in each of three villages in South India, focuses explicitly on how labor supply by family members is responsive to income shocks and, indeed, can serve as a consumption-smoothing mechanism. She employs a model in which labor supply is a choice, and, in particular, pays attention to the fact that there are different cases or corner solutions; in particular, that in some households family members do not work in the wage labor market. Kochhar estimates a household consumption equation only for those households with labor market participants, taking into account the selectivity of that sample. However, transitory income, instrumented by monthly rainfall levels, is measured, as in Paxson's and Wolpin's studies, gross of the value of family labor. Kochhar's finding of a small effect of the income shock on consumption could therefore also be due to the pres-

ence of on-farm labor-supply responses, highlighted in her model, that are embedded in her measure of farm profits.

Jacoby and Skoufias (1998) assume, in contrast to Kochhar, that family and hired labor are perfect substitutes and net out the cost of both hired and family labor, valued at the market wage, as is required to obtain consistent estimates of income effects in the model under that assumption. Jacoby and Skoufias used the same data set as Kochhar. Profits are decomposed into anticipated and unanticipated components. Estimates of anticipated profits are based on household and farm characteristics considered to be predetermined and their interactions with rainfall characteristics that are realized prior to planting, and thus known to the farmer at the time planting decisions are made. Estimates of unanticipated profits are based on interactions of the predetermined variables with rainfall characteristics realized after planting and just prior to harvesting, which are unknown at the time of planting. Aggregate (village) level shocks are thus assumed to affect profits of individual farmers idiosyncratically. The results of the instrumental variables estimation using rainfall show that neither the effect of anticipated nor unanticipated profits on consumption expenditures within a season is statistically or economically significant. This result implies that farm households are able to smooth consumption over seasons as if there were perfect markets.

Case II. $m_2 = 0$. We now show that the validity of the Jacoby and Skoufias findings, and the inferences about the biases in the Wolpin, Paxson, and Kochhar studies, rest importantly on the assumption that hired and family labor are perfect substitutes for at least one labor task and that family labor is actually employed in that task. In particular,

we consider the case in which family labor is not employed in the manual task, e.g., the value of leisure is sufficiently high and/or the marginal product of supervisory labor is sufficiently high. This case is equivalent to one in which there are no production tasks in which family labor and hired labor are perfect substitutes. As a consequence of this, the supply of farm labor to the supervisory task as well as the demand for hired manual labor will not be independent of the level of income; that is, the production and consumption sectors are no longer separable. In particular, if leisure is a normal good, then the supply of farm supervisory labor will decline as income in any period increases and the demand for hired labor will also decline if manual labor and supervisory labor are complements (and increase if they are substitutes).

As in the previous case, a dollar increase in last-period income (assured income) increases consumption by less than a dollar as leisure is increased; specifically,

$$\frac{\partial c_2}{\partial y_2} = f_2 \frac{\partial s_2}{\partial y_2} + 1.$$

This expression differs from the previous case in that the marginal product of the input that is affected by the change in income (supervisory labor in this case and manual labor in the previous case) cannot be valued at a market wage. However, the lack of a market for tasks performed by family labor implies that the effect of weather on consumption is no longer proportional to the (assured) income effect. The expression for the effect of weather on consumption is given by:

$$\frac{\partial c_2}{\partial \varepsilon_2} = f_3 \frac{\partial c_2}{\partial y_2} + f_2 \left[\frac{f_{13}f_{21} - f_{11}f_{23}}{\Delta} \right], \quad (33)$$

where $\Delta > 0$ is the appropriate bordered Hessian for the maximization problem. As seen in (33), the effect of weather on

consumption has two components, one that arises from the direct effect of weather on income (the effect that is proportional to the income effect on consumption) and one that arises from the effect of weather on net profits through its effect on the supply of supervisory labor (an effect that does not arise in equilibrium in the previous case because the marginal product of supervisory labor is equal to the fixed market wage).⁴⁴ Dividing (33), the weather effect on consumption, by the input-constant effect of weather on revenues, i.e., by

$$f_3 = f(h_2, s_2, \omega_2) - w \frac{\partial h_2}{\partial \omega_2} - f_2 \frac{\partial s_2}{\partial \omega_2},$$

will not provide a consistent estimate of the transitory income effect on consumption because of the existence of the second term in (33). Moreover, the direction of the bias in the IV estimator based on such an income measure can be of either sign. Similar results can be demonstrated for the effect of transitory income effects on period-one savings. Thus, in this case, there is no measure of net income that can be used as the basis of a weather-based IV estimator of transitory income effects that would be consistent.

The conclusions made by Rose (1999) about Indian households' ability to smooth consumption, in her innovative study of the effects of weather shocks on sex-specific child survival, based on the same national probability sample of rural households used by Wolpin, also rests importantly on these labor-market and technology assumptions. In that study, Rose looks at the reduced-form relationship between weather shocks—district-specific deviations in annual

⁴⁴This is just an application of the envelope theorem in the complete markets case. In fact, the bracketed component of the second term in (33) is exactly the effect of weather on supervisory labor in the complete markets case.

rainfall from a 21-year average—occurring at three life-cycle stages for a child: at birth, in its first year of life, and in its second year. She finds that girls are less likely to survive than are boys if an adverse weather shock occurs in the first year after birth. She interprets this as indicating that households cannot successfully smooth consumption and that when income levels are low, resource allocations are such as to discriminate against girl infants. However, while the inferences about discrimination appear to be appropriate, the absence of information on the relationship between the weather shocks and income in the study means that it is not possible to assess whether the effects of the weather shocks on child survival indicate that income shocks on resource allocations are small or large. Moreover, if there are no perfect substitutes for the mother's time, for example, in the caring of children and in farm production, then the effect of the weather shock may reflect optimal time-allocation responses (and preferences for boys) to weather-induced land productivity changes and not just the effect of income shocks, consistent with Rose's work (Rose forthcoming) showing that an adverse weather shock leads to more off-farm work.

It would appear that a necessary condition for weather variation to be useful for estimating income effects and the degree of consumption smoothing, given the appropriate measurement of net income, is that family and hired labor are perfect substitutes in at least one task. If, for example, it is not possible for reasons of moral hazard to hire supervisory farm labor, then weather-based income effects obtained from farm populations, where this instrument has the most power, will be biased. To validate studies of income effects based on the natural natural weather experi-

ment thus requires tests of separability. Such tests themselves, however, often require additional identifying restrictions.⁴⁵ We can carry out a simple test of the plausibility of the labor-market assumptions used in the Wolpin, Paxson, and Jacoby and Skoufias studies using a weather instrument. We use a sample of 2,567 farm households from a national sample survey of all rural Indian households, numbering 4,896, carried out in 1982, the Rural Economic and Demographic Survey (Vashishtha 1978), supplemented by rainfall data. A unique feature of these data is that they distinguish between supervisory and crop-labor time (days worked in the crop year 1981–82) among family members. The data indicate that in over 12 percent of farms at least one family member devotes time to supervisory tasks. These are the larger farm households in which the labor force is sufficiently large to have family members specialize and therefore where it is possible to readily distinguish supervisory activities from manual labor tasks.

We would expect that if there are hired substitutes for supervisory labor, as there appear likely to be for crop or manual labor, exogenous variations in permanent attributes of weather should have the same effects on the amounts of both types of family labor. In particular, given that leisure is a normal good, in the Indian context higher average levels of weather should be associated with lower days worked in both supervisory and crop tasks among family workers. Based on an annual time-series of average daily rainfall from 1961 through 1980 at the district level we constructed the mean rainfall per day in the district in which each farm household resided.

⁴⁵ For example, a critical assumption in Benjamin's (1992) test of separability among Indonesian farm households is that household size variation conforms to the randomness criterion.

TABLE 5
RURAL INDIAN FARM HOUSEHOLDS: AVERAGE DAILY RAINFALL, ANNUAL CROP-LABOR DAYS
PER FAMILY WORKER AND ANNUAL FAMILY SUPERVISORY DAYS
(NCAER REDS)

Variable	Crop-Labor Days per Family Worker (OLS)	Family Supervisory Days (ML Tobit)	Ratio of Supervisory to Crop-Labor Days (ML Tobit)
Mean (S.D.)	41.0 (39.6)	2.73 (13.5)	.356 (5.18)
Mean rainfall (mm per day)	-.00825 (.0016) ^a	.0208 (.0030)	.00647 (.0010)
Constant	49.3 (1.79)	-83.6 (5.36)	-27.8 (21.4)
S.E.	—	52.5 (2.38)	16.6 (.712)
N	2,567	2,567	2,567

^a Standard errors in parentheses.

We then regressed number of crop days per family worker (adults aged 15–59) on this rainfall measure. Table 5 presents the estimates. As can be seen, in areas in which rainfall levels are higher on average, adult family members work significantly fewer days per year. This is consistent with leisure being a normal good, given that higher rainfall levels are associated with higher output, and with family crop labor being substitutable with hired labor. The result also demonstrates why it is necessary to net out the cost of family labor from income in assessing income effects based on weather variation. However, family time in supervisory activities is actually higher in such areas, as is the ratio of supervisory to crop-labor family days, indicating the difficulty of obtaining market substitutes for supervisory labor. This suggests that netting out total family labor time valued at the hired labor wage may not be sufficient to obtain identification of either permanent or transitory income effects using weather events as instrumental variables.

Finally, assumptions about the local

marketability of labor inputs are not sufficient to pin down income effects using weather variables. Weather events and labor supply decisions can affect not only income but also relative commodity prices. In the agricultural model incorporating labor supply choice, identification of income effects using spatial and intertemporal variation in weather events also requires that the local weather events not affect the locale-specific price of the consumption good relative to the wage. One sufficient condition is that either all inputs, including labor, or all goods are perfectly spatially mobile. Another is that leisure and consumption are strongly separable. That the permanent income studies do not include relative prices, inclusive of wages, in the consumption function estimated reflects one or another of these additional assumptions, although the assumptions are never stated. In the absence of these restrictions, more instruments would be needed than weather events to account for both endogenously-determined price and income effects, as relative prices would likely be correlated

with weather-driven income variation and consequent consumption and labor supply choices. In that case too, the effects on consumption choices of uncertainty about prices and wages would also have to be considered.⁴⁶

5. Estimating the Effect of Children on Female Labor Supply

A major challenge in labor economics is to explain the secular increase in the labor force participation of married women in the United States and most industrialized countries. One of the candidate factors in the rise in female labor-force participation is the decline in fertility. Given that both labor supply and fertility decisions are endogenously chosen, instrumental variables have been employed in many studies (e.g., Belton Fleisher and George Rhodes 1982; T. Paul Schultz 1980; and see Martin Browning 1992) to assess the contribution of changing fertility on maternal participation and work time.⁴⁷ The first to use a random natural event—twins on the first birth—to estimate how fertility change affects maternal labor supply was Rosenzweig and Wolpin (1980b), followed by Stephen Bronars and Jeff Grogger (1994), Jaisri

⁴⁶ At least in the Indian context, spatial wage differences are significant and sensitive to weather events (Rosenzweig 1986). Given that consumption and leisure are substitutes, then if wage rates and income both co-move positively with favorable weather shocks and the wage is excluded from the specification of the consumption equation, the estimated transitory income effect is positively biased because it also contains a substitution term. In the case in which family (supervisory) labor is not marketed, an additional omitted variable is the "shadow" wage of this labor, which is endogenously determined and varies with local weather.

⁴⁷ Almost all of these restrictions and assumptions appear to be arbitrary. For example, in Fleisher and Rhodes' study (1982) of the effects of fertility on labor supply the authors assume that parents' schooling is exogenous (there are no common unobservable elements that affect both decisions), and that father's schooling affects fertility but not the mother's labor force decisions.

Gangadharan and Joshua Rosenbloom (1996), and Joyce Jacobsen, James Pearce and Rosenbloom (1999). Recently Angrist and Evans employed the gender of the first two births, specifically sex-sameness, as a natural instrument to estimate fertility effects on married women's labor supply. In this section we show that the specific identification strategies used in both the set of "twins-first" studies and that by Angrist and Evans place strong restrictions on both preferences and on household technology and we present empirical evidence based on unique data from India that calls into question one of these important restrictions. Before turning to the issue of identification, we first review the main empirical results of the studies that make use of the initial twin-births and child gender natural experiments.

Rosenzweig and Wolpin (1980b), using data from the 1965 and 1973 National Fertility Surveys, found that among women who had their first births between the ages of 15 and 24, those who had twins on their first birth had 0.65 more children on average ten years later.⁴⁸ The same comparison for women who had their first births between 25 and 34 yielded a difference of 0.31 births. However, for the younger age at first birth group of women, completed fertility for those with twins, as measured twenty years later, was only

⁴⁸ As noted, analogous to the "any sisters" instrument of BC, the existence of any twin births in a household would be an inappropriate instrument because its probability is increasing in the number of pregnancies, even if the probability of a twin birth is independent of the number of prior pregnancies (actually, the probability of twinning is increasing in the number of prior births). Using twinning on the first birth only avoids this problem. However, the probability of having a twin at any parity is increasing in age at birth. RW show that holding age at first birth constant, even though the timing of the first birth may itself be a choice, maintains the validity of the twins-first instrument.

0.15 greater than for those without twins.⁴⁹ This rather small difference in completed family size leads to two important conclusions: (i) that contraceptive adjustment costs are small and (ii) that the twins first experiment would seem to correspond mainly to a difference in the timing of births, i.e., having one additional child at an earlier age offset by having one less child over the remaining fecund period. Bronars and Grogger (1994) and Jacobsen, Pearce, and Rosenbloom (1999) use 1970 and 1980 Census data and thus have larger first-birth twins samples. Using true cohorts, they find a similar attenuation in total births over the ten-year period subsequent to the birth.⁵⁰

With respect to labor supply, RW found that among women who had their first births between the ages of 15 and 24, those who had twins had a .37 percentage-point lower labor force participation rate in that age interval than did women without twins. Between the ages of 25 and 34, the labor force participation rate differential fell to .10. However, between 35 and 44, twins mothers actually had a .14 percentage-point higher labor force participation rate.⁵¹ Dividing the labor supply effects by the fertility effects of twins-first yields the Wald estimates of the effect of an additional child on labor force participation rates.

RW interpret the twins first experiment within the context of a two-period model of fertility and labor supply. Parents also derive utility from the quality of their children, which depends directly on their number and on the amount of time the woman spends at

⁴⁹ These figures are based on synthetic cohorts.

⁵⁰ See also Gangadharan and Rosenbloom (1994).

⁵¹ This reversal also occurred using true cohort comparisons, although the sample sizes were quite small. BG find that labor force participation rates do not differ after ten years.

home. In addition, the market wage in the second period depends on work experience gained in period one. There is a fixed price of having an additional child, e.g., birth and contraceptive costs, independent of child quality. In the first period, women decide on how many children to have and on hours worked (and implicitly on child quality) and in the second period on hours worked. RW's main result is that the effect of an exogenous additional child on hours worked in either period is the ratio of the compensated effect of the fixed price of a child on hours worked to the same compensated effect on fertility.⁵² Note that the Wald estimator that used the price of a child, if it were observed, as an instrumental variable would correspond to the ratio of uncompensated price effects, which differs from the effect of an exogenous additional child on labor supply that correctly is derived from the conditional demand function.⁵³

Angrist and Evans (1998) showed that, as is the case for twin-first births, the sex ratio of the first two children affects subsequent fertility decisions. Using data from the 1980 and 1990 U.S. Census, they found that among parents who have two or more children the proportion who have a third birth is greater by .06 if the first two children were of the same sex than if they were of the opposite sex. In terms of labor supply effects, AE estimate that the additional (third) child reduces female labor force participation rates by about .12 (1980).

⁵² This result is a direct application of rationing theory (James Tobin and Hendrik Houthakker 1950–51). The resulting relationship between labor supply and (exogenous) fertility is a conditional demand function (Robert Pollak 1969).

⁵³ Of course, full estimation of the ordinary demand functions for labor supply and fertility would, in the context of the two-period model, provide all of the information necessary to calculate the conditional demand function.

In contrast, they also find that using as an instrument having twins on the *second* birth reduces labor force participation rates by .08.

AE's published paper does not present a behavioral model within which to interpret their estimates. However, they do refer to a specific model set out in an earlier version of the published article (Angrist and Evans 1996). This highly restrictive model, however, fails to deliver the result that the Wald estimator of the labor supply effect of exogenous additional children based on the same-sex instrument is identical to the direct effect of manipulating fertility except in the case of having a perfect measure of and controlling for lifetime wealth.

The twins-based and child-gender-based Wald estimators both require restrictions on preferences and (household) technology in order to go from the identification of an exogenous variable beyond the control of economic agents to the identification of how fertility affects labor supply. Both the studies that use the twins-first and sex-sameness natural experiments impose the (exclusion) restriction that the natural event—having twins on the first birth or having children of the same sex in the first two births—does not directly affect the subsequent labor supply of either parent except through its effect on having an additional birth. What assumptions about behavior or technology yield this restriction? What does the Wald estimate reveal about behavior and/or technology? We set out a simple model of fertility and labor supply choices to elucidate the interpretation of the Wald estimators from these studies and to provide a framework for garnering additional empirical evidence that sheds light on the realism of the identification restrictions. The model shows that the restrictions required for identifying the fertility effect on labor supply

using either natural experiment involve strong (and similar) assumptions about preferences and household technology.

The model we set out incorporates features highlighted in the twins-first and sex-sameness studies—parental preferences and child rearing costs depend on the spacing of children and on the sex composition of children. Specifically, in each period of the model a woman decides on whether or not to have a child, $n = \{0,1\}$, and whether or not to work, $h = \{0,1\}$, up to a last period when she can have no more children. In that period, the woman thus makes a labor supply decision only. The woman's utility in each period depends on her consumption, c , her labor supply, the stock of children at the beginning of the period, N , the stock of children at the beginning of all prior periods, and the sex composition of her children, where we denote r as the fraction of female children. The woman receives a wage, w , in each period that she works, has an exogenous income flow each period, y , and bears a per-child rearing cost, e . The child rearing cost is assumed to depend on the spacing of births and the sex composition of children.

All of our points can be illustrated by considering the final-period labor supply decision. To obtain analytical solutions, we adopt a quadratic utility function. We also assume that the "bliss-point" sex-ratio is one-half (parents prefer to have mixed-sex children) but that the child rearing cost, reflecting the savings from sex-specific hand-me-downs, is minimized when all children are the same sex, i.e., $r = 0$ or 1 . Both the sex-ratio and the spacing of children are allowed to affect the marginal utility of leisure. Assuming that the terminal period of the model is the fourth,⁵⁴ the

⁵⁴ A minimum of four periods is needed to capture the possibility of different sex ratios, which require at least two births.

final-period optimization problem is then to maximize

$$\begin{aligned} u_4 = & c_4 + \alpha_1 N_3 - \alpha_2 N_3^2 \\ & - \alpha_3(r_3 - .5)^2 + \alpha_4(1 - h_4) \\ & + \alpha_5 c_4 N_3 + \alpha_6 c_4 (1 - h_4) \\ & + \alpha_7 N_3 (1 - h_4) \\ & + \alpha_{71} N_1 (1 - h_4) + \alpha_{72} N_2 (1 - h_4) \\ & + \alpha_8 (1 - h_4)(r_3 - .5)^2 \end{aligned} \quad (34)$$

with respect to the choice of h_4 subject to the budget constraint

$$\begin{aligned} c_4 = & y + w h_4 - e_1 N_3 - e_{11} N_1 - e_{12} N_2 \\ & - e_2 (r_3 - .5)^2. \end{aligned} \quad (35)$$

The final-period work decision is made by comparing utility over the two work alternatives, which will depend on both the stock and spacing of children and the sex-ratio. Specifically, suppose that the woman has two non-twin children of the same sex by the end of the third period, then the difference between the utility of working and not working in the final period is given by

$$\begin{aligned} u_4[h_4 = 1 | N_3 = N_2 = 2, N_1 = 1, r_3 \\ = \{0, 1\}] - u_4[h_4 = 0 | N_3 = N_2 \\ = 2, N_1 = 1, r_3 = \{0, 1\}] \\ = w(1 + 2\alpha_5) - \alpha_4 - \alpha_6[y - 2e_1 - 2e_{12} \\ - e_{11} - .5e_2] - 2\alpha_7 - 2\alpha_{72} - \alpha_{71} - .25\alpha_8. \end{aligned} \quad (36)$$

If we assume that some of these parameters are stochastic from the researcher's perspective, then (36) can serve as the basis for probability statements, namely that the probability that a woman with a given wage and with two non-twin children of the same sex works is the probability that (36) is positive.

To determine the impact of the sex-ratio on the probability of working net of the effect of varying family size and spacing, (36) can be compared to the same utility difference when the sex-ratio is .5 (the only other possibility with two children), namely

$$\begin{aligned} u_4[h_4 = 1 | N_3 = N_2 = 2, N_1 = 1, r_3 \\ = .5] - u_4[h_4 = 0 | N_3 = N_2 = 2, N_1 = 1, r_3 \\ = .5] \\ = w(1 + 2\alpha_5) - \alpha_4 - \alpha_6[y - 2e_1 - 2e_{12} - e_{11}] - 2\alpha_7 - 2\alpha_{72} - \alpha_{71}. \end{aligned} \quad (37)$$

Subtracting (36) from (37) determines the extent to which the utility gain to (or loss from) working is affected by having two non-twin children of the same sex as opposed to two non-twin children of different sexes—the sex-composition effect holding fertility constant. Denoting the left hand side of (37) as $u_4(1, 0)$, this difference is given by

$$\begin{aligned} \frac{\Delta u_4(1, 0)}{\Delta r_3^*} &|_{N_3 = N_2 = 2, N_1 = 1} \\ &= .5\alpha_6 e_2 + .25\alpha_8, \end{aligned} \quad (38)$$

where $r_t^* = |r_t - 0.5|$ is the absolute deviation of the period- t sex ratio from 0.5.⁵⁵ As seen from (38), three parameters determine the magnitude of the sex-sameness effect, α_6 , α_8 and e_2 . Thus, the deviation from sex-sameness will affect the labor supply decision, net of family size effects, as long as either (i) the deviation from sex-sameness directly affects the marginal utility of leisure ($\alpha_8 \neq 0$) or (ii) changes in sex-sameness affect the cost of child rearing ($e_2 \neq 0$) and consumption and leisure are not strongly separable. Conversely, in this model to assume that sex-sameness does not affect the labor supply decision directly as in AE requires that the child sex-composition and leisure be strongly separable in the utility function and that either sex-sameness not affect child costs or consumption and leisure are strongly separable.⁵⁶

⁵⁵ Notice that this expression is independent of the spacing of children, i.e., of whether there were twins on the first birth, given the separability assumed between the sex-ratio and birth spacing in utility and child-rearing cost functions.

⁵⁶ As noted, birth spacing is ignored in the model presented by RW. In the model referenced in AE (1998) contained in their prior unpublished paper (AE 1996), leisure and the sex-composition of children are separable, child costs are independent of the sex composition of children and parental

Analogously, the effect on labor supply of having twins in the first birth as opposed to two single births in the first two periods, given the same sex-ratio, is

$$\frac{\Delta u_4(1,0)}{\Delta N_1} \Big|_{N_3=2,r_3} = \alpha_6 e_{11} - \alpha_{71}, \quad (39)$$

Thus, to assume that having twins in the first birth does not affect the labor supply decision directly as in RW requires that birth spacing and leisure be strongly separable in the utility function and that either having twins does not affect child costs or consumption and leisure are strongly separable.

In this model, the impact of having had an additional child on labor supply for a given sex-ratio can be similarly assessed by comparing, say, the utility in the final period associated with the work decision when the woman has had three children that are all of the same sex and all single births with having only two (non-twin) children of the same sex. The fertility effect on this utility difference is

$$\begin{aligned} \frac{\Delta u_4(1,0)}{\Delta N_3} \Big|_{N_2=N_1=1,r_3=\{0,1\}} \\ = w\alpha_5 + \alpha_6(e_1 - .25e_2) - \alpha_7. \end{aligned} \quad (40)$$

Fertility thus affects labor supply in the model, for a given sex composition and birth spacing, as long as either consumption and family size are non-separable, consumption and leisure are non-separable, or family size and leisure are non-separable.⁵⁷

Comparing (40) with (38) and (39), it can be seen that it is possible, by ignoring the effect of twins and sex-sameness

consumption is fixed at zero. Sex-sameness, the identifying instrument, does not appear as a specific measure in the model separate from sex-composition.

⁵⁷In the model in AE (1996), fertility affects labor supply directly by affecting child costs. This is equivalent, as in the model here, to fertility and leisure being non-separable in the utility function (Pollak and Michael Wachter 1975).

on child costs and imposing separability between the spacing and sex-composition of children and leisure, for twins in the first birth and for the sex-composition of children to affect fertility but not labor supply net of fertility while exogenous variations in fertility affect labor supply decisions. These assumptions are required for using twins in the first birth or the sex-composition of initial births as instruments to obtain an estimate that is equivalent to that of the effect of exogenous variation in N on labor supply.⁵⁸

The Wald estimate of the labor supply response to an increase in the number of children based on an instrumental variable x is

$$\frac{\Delta \bar{h}_4 / \Delta x}{\Delta \bar{N}_3 / \Delta x},$$

where the numerator is the change in the fraction of women who work due to the change in x and the denominator is the change in the mean number of children ever born due to the change in x . In the case of studies using twins in the first birth, $\Delta x = \Delta N_1$ represents the one additional child that is born. In AE, $\Delta x = \Delta r_2$ represents the sex ratio at the beginning of the penultimate period, r_2 . The sex-sameness and twins first instrumental variables estimates, assuming that (39) and (40) are non-zero, are

$$\begin{aligned} \frac{\Delta \bar{h}_4 / \Delta r_2^*}{\Delta \bar{N}_3 / \Delta r_2^*} &= \frac{\Delta \bar{h}_4}{\Delta \bar{N}_3} \Big|_{r,N_1,N_2} \\ &+ \frac{\Delta \bar{h}_4}{\Delta \bar{r}_3^*} \Big|_{N_3,N_2,N_2} \times \frac{\Delta \bar{r}_3^*}{\Delta \bar{N}_3} \end{aligned} \quad (41)$$

⁵⁸The use of instrumental variables to estimate the labor supply response to an additional child, $\Delta h_4 / \Delta N_3$, is predicated on the assumption that direct estimation, i.e., comparing the fraction of working women among those with twins in the first birth to those with single births or those with two same-sex children to those with three same-sex children, will give a biased estimate. This would be the case, for example, if women who had two children differed systematically from women who had three children in any parameters that determine their labor supply.

and

$$\frac{\Delta \bar{h}_4 / \Delta N_1}{\Delta \bar{N}_3 / \Delta \bar{N}_1} = \frac{\Delta \bar{h}_4}{\Delta \bar{N}_3} \Big|_{r, N_1, N_2} + \frac{\Delta \bar{h}_4}{\Delta N_1} \Big|_{N_3, r} \div \frac{\Delta \bar{N}_3}{\Delta N_1}. \quad (42)$$

The first term in (41) and (42) is the direct effect of a change in family size on labor supply, the object of interest. The sex-sameness and twins first instrumental variable estimates, however, each contain a second term if no additional restrictions are imposed on preferences and child costs. In the case of sex-sameness, this second term arises because having a third birth will in almost all cases also change the subsequent sex-ratio, and this may directly affect both the marginal utility of leisure and child costs and thus labor supply.⁵⁹ In the case of twins, having two children simultaneously may directly affect both the marginal utility of leisure and child costs and thus labor supply.⁶⁰

Besides clarifying the multiplicity of structural assumptions that are required to justify the interpretation of the instrumental-variable estimates based

on these natural experiments, the model indicates the information that would be useful in gauging their plausibility. Although providing credible estimates that would reveal whether specific commodities are separable or not in the preference function is likely not feasible, it is possible to examine the relationship between the sex-sameness of children and child costs. We use unique information from the same Indian data set we used to examine the relationship between permanent weather and family labor time. These data provide for 4,896 rural households in 1982 household expenditures on clothing and footwear for every child in the household, likely candidate commodities for which sex-specific hand-me-down savings are possible.

India is a country well-known for child sex-bias among parents. Parental preferences for boys in India probably dominate their concern for achieving a balance in the sex ratio that Angrist and Evans highlight for U.S. parents. This preference at the household level reflects in part cost differentials by sex, as in most parts of India a dowry must be paid when daughters are married (Behrman, Andrew Foster, Rosenzweig, and Prem Vashishtha 1999). Thus, the sex of children affects child-rearing costs directly, as allowed for in the Angrist and Evans' two-stage least-squares estimates, and we would expect due to parental sex-bias that measures of the sex-ratio of the first two births will also affect fertility in India, although perhaps in a different way than in the United States. However, in both countries the economies associated with sex-sameness from the use of same-sex hand-me-downs should be similar in kind if not in magnitude. The empirical issue is whether child-rearing costs are also affected by sameness, the identifying instrument.

As part of the sample survey, all mothers aged 15 through 50 residing in

⁵⁹ For example, in the extreme case in which only families with children of the same sex have an additional child, approximately half could not exploit same-sex hand-me-down savings for this additional child. Unless consumption and labor supply are separable and ignoring the issue of whether or not sex composition and leisure are separable, as indicated by (41), the sex-sameness instrumental-variable estimator confounds the direct effect of changing family size on labor supply with the effect of also changing child costs for a considerable proportion of women.

⁶⁰ AE also compare the sex-sameness instrumental-variable estimate to that based on twins (in the second birth) to draw inferences about the effects of birth spacing. However, without the special restrictions on costs and on the utility function that make expression (41) vanish, the comparison of the sex-sameness estimate with that obtained using twins as an instrument does not identify the effect of birth spacing. If (41) and (42) were by coincidence identical and non-zero, one would get the same "biased" estimate of fertility effects on labor supply.

TABLE 6
AVERAGE SAMPLE CHARACTERISTICS:
RURAL INDIAN MOTHERS AGED 30–50 AND THEIR CHILDREN AGED 6–7
(NCAER REDS)

	Mothers	Girls	Boys
Number of live children	4.51 (1.79) ^a	—	—
Average hours worked per day	4.67 (3.69)	—	—
Age	39.8 (5.37)	11.8 (3.27)	11.8 (3.29)
Per-child clothing expenditures (1982 rupees)	—	95.2 (175.9)	100.2 (121.8)
Per-child educational expenditures (1982 rupees)	—	76.6 (371.9)	107.0 (230.1)
Number in sample	2,356	2,703	3,040

^a Standard deviation in parentheses.

the sampled households were interviewed to obtain a complete fertility roster providing the birth dates, order, sex, and survival status of each child born, along with information on the average hours that the mothers worked in a day in each of three crop seasons. Each mother was also asked to provide information on expenditures in the past year for each child on clothing and footwear. Expenditures on education (books, school fees, writing instruments) were also provided. Unlike for clothes, however, as most Indian children attend mixed-sex schools, the sex-sameness of births should not affect cost savings from hand-me-downs with respect to educational expenditures. We should not therefore expect to see sex-sameness affect the patterns of these expenditures.

Table 6 provides descriptive statistics for the sample of mothers aged 30–50 and their children of potential school age. As can be seen, compared to countries like the United States, fertility is very high in India—mothers aged 30 and above in 1982 had 4.5 living children. Sex

discrimination is apparent in schooling expenditures, which are on average almost 40 percent higher for boys in the age range 6 through 17 than for similarly aged girls, a difference that is statistically significant at the .001 level.⁶¹ Expenditures on clothing do not, however, differ significantly across boys and girls in that age range. More importantly, for the purposes of examining the relationship between changes in child costs, fertility, and labor supply, it can be seen that expenditures on children's clothing are a non-trivial part of the household budget in rural India. At the time of the survey, the average wage for rural male workers was 8.5 rupees per day, and that for women 6.7 rupees per day. For the average household in the sample, earnings are about 3000 rupees. For women aged 30 through 50 in the sample, there are on

⁶¹ Discrimination may as well be reflected in the ratio of boys to girls—girls represent 47 percent of all surviving children. There is evidence that this disparity in part represents selective mortality (Rosenzweig and Schultz 1982; Rose 1999).

TABLE 7

NONPARAMETRIC, OLS, AND WALD ESTIMATES: SEX COMPOSITION OF FIRST TWO CHILDREN, LABOR SUPPLY, AND FERTILITY, RURAL INDIAN WOMEN AGED 30-50^a
(NCAER REDS)

Dependent variable	Sample Means	Nonparametric Reduced-Form Estimates		OLS Reduced-Form Estimates, with Restrictions		Wald Estimate
		Total children	Hours worked	Total children	Hours worked	
Girl/girl	.208	.386 (.114) ^b	-.169 (.228)	—	—	—
Boy/boy	.324	-.112 (.100)	-.301 (.206)	—	—	—
Boy/girl	.243	-.198 (.103)	.368 (.223)	—	—	—
Girl/girl or Boy/boy = 1	.532	—	—	.185 (.073)	-.441 (.154)	—
Total children	4.51	—	—	—	—	-2.38 (1.27)
Constant	—	4.51 (.076)	4.72 (.158)	4.41 (.052)	4.91 (.115)	15.4 (5.75)
F (d.f., d.f.)	—	10.6 (3,2250)	3.68 (3,2250)	6.36 (1,2250)	8.24 (1,2250)	3.49 (1,2250)

^a Number of women = 2,356; number of households = 2,251.

^b Standard errors in parentheses corrected for household clustering.

average 2.4 children aged 6 through 17 residing in the household. Annual clothing expenditures on these school-aged children thus represent 8 percent of income, and about 11 percent including all children younger than 18.

The high fertility of the Indian women means that the decision to have a third child, examined by AE in the context of the United States, is not as critical a fertility decision. Indeed, in the sample of mothers aged 30 and above, 80 percent had at least three living children. However, the Indian data indicate that, consistent with both sex-bias and sex-sameness interpretations, the sex-composition of the first two (surviving) children has a significant effect on total fertility, as in the United States. The first and second column of

estimates in Table 7 report nonparametric reduced-form estimates of the relationship between the exhaustive set of variables characterizing the sex-composition of the first and second living children and total children and average hours worked per day by the Indian sample mothers. The set of sex-composition variables explains a statistically significant proportion of the variation of both variables, with mothers with two girls as their first two children having on average almost ½ more children than women who have both first and second-born boys. Interestingly, mothers with two boys have both lower fertility and lower hours worked compared with women who have two girls. However, imposing the restriction as in Angrist and Evans' initial set of Wald

TABLE 8
 OLS ESTIMATES OF SEX AND SAME-SEX "HAND-ME-DOWN" EFFECTS,
 BY BIRTH ORDER AND EXPENDITURE TYPE: RURAL INDIAN CHILDREN AGED 6-17
 (NCAER REDS)

Expenditure Type	Per-Child Clothing Expenditures		Per-Child Educational Expenditures	
	1	3	1	3
Birth order of child				
Has same-sex older sibling	—	-19.9 (9.34)	—	4.88 (13.6)
Child is a girl	0.729 (5.21) ^a	-10.1 (6.39)	-27.0 (10.6)	-43.0 (18.6)
Age	4.31 (.807)	4.72 (1.40)	8.84 (1.72)	8.14 (1.89)
Constant	42.8 (24.3)	92.9 (33.1)	116.9 (47.7)	177.7 (87.9)
F-statistic (d.f., d.f.)	14.4 (2, 1718)	5.12 (3, 982)	15.4 (2, 1718)	8.88 (3, 982)
Number of mothers	1,878	1,023	1,878	1,023
Number of households	1,719	993	1,719	993

^a Standard errors in parentheses corrected for household clustering.

estimates that only sex-sameness matters for fertility yields results qualitatively similar to those from U.S. data: Indian women with same-sex first and second births have higher subsequent fertility and lower hours worked (columns three and four), and the restrictive Wald estimate of the effect of total children on labor supply, which makes use of the same-sex exclusion restriction, is negative and statistically significant (final column).

The finding that the Wald estimate is negative for sex-sameness, as we have shown, does not indicate anything about the structure of child costs, in particular of the absence of same-sex effects on costs. A direct test of the existence of same-sex hand-me-down cost savings that is highlighted in the model is to assess whether there are lower clothing expenditures among higher-parity children who have an older sibling of the same sex. The illustrative model that we

have set out omits, however, the possibility that parents care about a child's sex per se or exogenous differences in costs by sex. Identification of sex-sameness effects from sex effects on child costs is potentially a key issue here, as it is in AE's study of fertility effects on labor supply. This is particularly so if in general girls impose higher costs than boys, as is plausibly the case in India. However, it is possible to separate sex preference and general sex-specific cost differential effects from sex-sameness effects on clothing costs by looking at the relationship between the child's sex and child-specific expenditures on clothing for first children. Because same-sex hand-me-down cost savings are irrelevant for first births, the effect of the sex of the first child on clothing costs only reflects either preferences, e.g., parents may prefer to spend more on male children, or income effects, e.g., the birth of the male

child reduces lifetime income less than does that of a female child.

Table 8 reports in the first column the relationship between a child's sex and the child's clothing costs for school-age first children. As can be seen, there is no statistically or economically significant effect of the child's sex for these children—families who have a boy first compared with families who have a girl first do not spend any more or less on clothing for that first child. This is consistent with the absence of important sex-differentials in clothing costs or income effects on clothing expenditures related to the sex-composition of children.⁶² In the second column of Table 8, we report for order-three children the relationship between the child's clothing costs and whether or not that child has an older sibling of the same sex, controlling again for that child's sex and age. The estimates indicate that sex-sameness has a statistically significant and substantial negative effect on child costs for children aged 6–17; the point estimate indicates among order-three children clothing costs are 20 percent less for children with same-sex older siblings compared with those children without same-sex younger siblings.⁶³

⁶² Incorporating cost differentials in the technology of the model is straightforward and would imply that sex-sameness effects on costs for higher-order children would differ by sex. Allowing for sex-composition induced income effects on child expenditures would entail enlarging the model choice set. Consistent with the results for first children, however, we could find no statistically significant difference in cost savings from sex-sameness for boys or girls ($p = .21$). The data thus do not suggest that either of these explorations would be empirically fruitful.

⁶³ If there is heterogeneity in preferences for girls and boys among households and this is reflected in differential survival, then these estimates are likely to underestimate savings from same-sex births. This is because those households that prefer more strongly boys or girls are more likely to have surviving same-sex children compared with households with a more mixed-sex composition. In that case the third child in a same-sex

What happens to the sex composition of children when the third child is born evidently matters for child costs. Given that on average Indian families would have two same-sex children, the existence of these hand-me-down economies would evidently save 40 rupees a year, or slightly more than one week of full-time work for women,⁶⁴ for a considerable segment of the women's life-cycle.

To assess whether the same-sex effect for clothing is merely spurious and possibly reflective of complex but general sex-composition effects not related to hand-me-down savings, we estimate the same specification for educational expenditures, for which, in India, the scope for cost savings related to sex-sameness would be relatively limited. The educational expenditure estimates for first children in column three of Table 8 indicate that, unlike for clothing, having a boy first rather than a girl results in higher expenditures for the first child, so that for this component of child costs, sex composition evidently matters.⁶⁵ However, consistent with the absence of a cost advantage for school expenditures that are related to sex sameness, school expenditures on third children, controlling for their sex and age, are no different for those children who have same-sex older siblings and those who do not.

household will be on average in the preferred group and thus may be provided with more resources. Note that the existence of preference heterogeneity for sex-ratios, even if there were no selective mortality as is presumably the case in the United States, implies that the fertility variation induced by exogenous variation in sex ratios would reflect preference variation. In that case additional restrictions on preferences would be required to use sex-ratio variation as an instrument to identify how additional children affect labor supply.

⁶⁴ The average (8-hour) daily wage for women is 6.5 rupees as noted, and a standard work week is six days.

⁶⁵ This differential mainly reflects the fact that boys attend school for more years than girls on average in India.

That sex-sameness is related to expenditures on children's clothing and to educational expenditures in ways consistent with the existence of same-sex hand-me-down cost economies suggests that instrumental variable estimates of the effect of fertility on labor supply based on measures of sex sameness, as used in Angrist and Evans, will in the Indian context confound the effect of an exogenous increase in children on labor supply with direct child-rearing cost effects on labor supply—the sex composition of children plausibly alters labor supply through mechanisms other than through fertility change alone. Of course, it is not possible to infer from this evidence whether hand-me-down economies associated with the sex-mix of births are an important phenomenon in the United States, the context in which Angrist and Evans carried out their empirical work. Even if child costs were not importantly related to sex-composition, one still has to make strong assumptions about preferences to obtain identification of fertility effects as interpreted in AE. These results suggest, however, that additional data and theoretically grounded empirical work are required to better understand the relationship between fertility choices and labor supply.

6. Conclusion

The pioneering studies that have exploited the natural experiments provided by nature clearly demonstrate the potential value of these newly discovered tools. The contribution of these studies, however, is not in providing dramatically different results than obtained using more conventional approaches—most of the estimates are not very different—nor in providing estimates that are conclusive. It is evident that natural events used as instruments do not provide estimates that can be

unambiguously interpreted, although the range of possible alternative interpretations may have been reduced. We have seen, for example, that the choice of which natural events to use as an instrument matters—the use, for example, of date-of-birth as an instrument for schooling attainment provides an estimate of schooling returns for a different segment of the population than does the estimate based on gender as an instrument for schooling, and each is subject to a bias from the same misspecification that is of opposite sign. Nor is there a breakthrough here that obviates the need to specify behavioral models and market structure, to take care in measurement, or to pay attention to matters of specification.

What then is the contribution and the future for the natural natural experimental approach? The contribution is in both focusing attention on the important matter of identification in a world in which measurement of all relevant characteristics of agents is impossible and in drawing our attention to a set of valuable tools, provided by nature, that reduce the number of untestable assumptions required to obtain useful estimates. And, such gifts of nature may also limit the range of possible alternative interpretations that could be given to estimates obtained with them. Evidently, however, the potential for the strictly natural natural experimental approach, which relies exclusively on natural events as instruments, is constrained by the small number of random events provided by nature and by the fact that most outcomes of interest are the result of many factors associated with preferences, technologies, and markets. And the prospect of the discovery of new and useful natural events is limited. This is not to say that all natural events have been exploited—the contrasts of twin (first) birth effects that depend on

whether the twins are both girls or boys or mixed gender or identical have not been used in any empirical studies to our knowledge, and information based on lightning strikes that destroy specific facilities or stores could be employed to draw inferences about optimal store or facility location.⁶⁶

But it is clear that the number of natural instruments will never be sufficient to eliminate the necessity of imposing auxiliary assumptions or of obtaining supplementary empirical information relevant to the assumptions needed for identification. In combining information on natural events within the context of a coherent model that describes the behavior under study and with supplementary empirical information on the model structure, such as on technology or the extent of market completeness, future work can provide a more solid foundation for further advancement of empirical knowledge. Measurement without theory, however, is not significantly more valuable than it ever was before the use of natural natural experiments.

REFERENCES

- Angrist, Joshua D. 1990a. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *Amer. Econ. Rev.* 80:3, pp. 313-36.
- . 1990b. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records: Errata," *Amer. Econ. Rev.* 80:5, pp. 1284-86.
- Angrist, Joshua D. and William N. Evans. 1996. "Children and Their Parents' Labor Supply: Evidence From Exogenous Variation in Family Size," NBER (Cambridge MA) working paper 5778.
- . 1998. "Children and Their Parent's Labor Supply: Evidence From Exogenous Variation in Family Size," *Amer. Econ. Rev.* 88:3, pp. 450-77.
- ⁶⁶ Sean Ennis (1996) interprets the structural damage from the Northridge earthquake of 1994 as an exogenous shock to hospital capacity in the Los Angeles area. However, given the broad-based effects of this natural event, it is difficult to believe that the effect of reduced hospital capacity on prices is identified as is claimed in this study.
- Angrist, Joshua and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quart. J. Econ.* 106:4, pp. 979-1014.
- Ashenfelter, Orley and Alan Krueger. 1994. "Estimates of the Economic Return to Schooling from a New Sample of Twins," *Amer. Econ. Rev.* 84:5, pp. 1157-74.
- Ashenfelter, Orley and Cecilia Rouse. 1998. "Income, Schooling and Ability: Evidence from a New Sample of Identical Twins," *Quart. J. Econ.* 113:1, pp. 253-84.
- Becker, Gary S. and H. Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children," *J. Polit. Econ.* 81:2 pt. II, pp. S279-88.
- Becker, Gary S. and Nigel Tomes. 1976. "Child Endowments and the Quantity and Quality of Children," *J. Polit. Econ.* 84:4, pt. 2, pp. S143-62.
- Behrman, Jere R.; Andrew D. Foster, Mark R. Rosenzweig, and Prem Vashishtha. 1999. "Female Schooling, Home Teaching, and Economic Growth," *J. Polit. Econ.* 107:4.
- Behrman, Jere R.; Z. Hrubec, Paul Taubman, and T. J. Wales. 1980. *Socioeconomic Success: A Study of the Effects of Genetic Endowments, Family Environment and Schooling*. Amsterdam: North-Holland.
- Behrman, Jere R.; Mark R. Rosenzweig, and Paul Taubman. 1994. "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment," *J. Polit. Econ.* 102:6, pp. 1131-74.
- Behrman, Jere R. and Mark R. Rosenzweig. 1999. "'Ability' Biases in Schooling Returns and Twins: A Test and New Estimates," *Econ. Educ. Rev.* 18:2, pp. 159-67.
- Benjamin, Dwayne. 1992. "Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models," *Econometrica*, 60:2, pp. 287-322.
- Bishop, John. 1976. "Reporting Errors and the True Return to Schooling," U. Wisconsin mimeo.
- Bound, John and David Jaeger. 1996. "On the Validity of Season of Birth as an Instrument in Wage Equations: A Comment on Angrist & Krueger's 'Does Compulsory School Attendance Affect Schooling and Earnings,'" NBER working paper W5835.
- Bound, John and Gary Solon. 1988. "Double Trouble: On the Value of Twins-Based Estimation of the Returns to Schooling," NBER working paper W6721.
- Bronars, Stephen G. and Jeff Grogger. 1994. "The Economic Consequences of Unwed Motherhood: Using Twins as a Natural Experiment," *Amer. Econ. Rev.* 84:5, pp. 1141-56.
- Browning, Martin. 1992. "Children and Household Economic Behavior," *J. Econ. Lit.* 30:3, pp. 1434-75.
- Butcher, Kristin F. and Anne Case. 1994. "The Effect of Sibling Composition on Women's

- Education and Earnings," *Quart. J. Econ.* 109:3, pp. 531–63.
- Card, David. 1994. "Earnings, Schooling and Ability Revisited," NBER (Cambridge) working paper 4832.
- Chamberlain, Gary. 1977. "Education, Income, and Ability Revisited," *J. Econometrics*, 5:2, pp. 241–57.
- Ennis, Sean. 1995. "The Northridge Earthquake: A Natural Experiment in Market Structure," U. California Berkeley econ. dept. working paper 95/244.
- Fleisher, Belton M. and George F. Rhodes. 1982. "Fertility, Women's Wage Rates, and Labor Supply," *Amer. Econ. Rev.* 69:1, pp. 14–24.
- Gangadharan, Jaisri and Joshua Rosenbloom. 1996. "The Effects of Child-Bearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment," NBER working paper 5647.
- Griliches, Z. 1979. "Sibling Models and Data in Economics: Beginning of a Survey," *J. Polit. Econ.* 87:3 pt. II, pp. S37–S64.
- Griliches, Zvi and William Mason. 1972. "Education, Income, and Ability," *J. Polit. Econ.* 80:3 pt. II, pp. S74–S103.
- Hearst, N.; J. W. Buehler, T. B. Newman, and G. W. Rutherford. 1991. "The Draft Lottery and AIDS—Evidence against Increased Intra-venous Drug-Use by Vietnam-Era Veterans," *Amer. J. Epidemiology*, 134:5, pp. 522–25.
- Hearst, N.; T. B. Newman, and S. B. Hulley. 1986. "Delayed Effects of the Military Draft on Mortality—A Randomized Natural Experiment," *New England J. Medicine*, 314:10, pp. 620–24.
- Heckman, James. 1997. "Instrumental Variables," *J. Human Resources*, 32:3, pp. 441–62.
- Imbens, Guido and Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62:4, pp. 467–76.
- Jacobsen, Joyce P.; James Wishart Pearce III, and Joshua L. Rosenbloom. 1999. "The Effects of Child-Bearing on Married Women's Labor Supply and Earnings," *J. Human Resources*, 34:3, pp. 449–74.
- Jacoby, Hanan and Emmanuel Skoufias. 1998. "Testing Theories of Consumption Behavior Using Information on Aggregate Shocks: Income Seasonality and Rainfall in Rural India," *Amer. J. Agr. Econ.* 80:1, pp. 1–14.
- Kocherlakota, N. 1999. "Smoothing Consumption by Smoothing Income: Hours-of-Work Responses to Idiosyncratic Agricultural Shocks in Rural India," *Rev. Econ. Statist.* 81:1, pp. 50–61.
- McCourt K., T. J. Bouchard, David T. Lykken, M. A. Tellegen, and M. Keyes. 1999. "Authoritarianism Revisited: Genetic and Environmental Influences Examined in Twins Reared Apart and Together," *Personality Individual Differences*, 27:5, pp. 985–1014.
- Miller, Paul; Charles Mulvey, and Nick Martin. 1995. "What Do Twins Studies Tell Us about the Economic Returns to Education? A Comparison of US and Australian Findings," *Amer. Econ. Rev.*, 85:3, pp. 586–99.
- Mincer, Jacob. 1970. "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," *J. Econ. Lit.* 8:1, pp. 1–26.
- Paxson, Christina H. 1992. "Using Weather Variability to Estimate the Response of Savings to Transitory Income in India," *Amer. Econ. Rev.* 82:1, pp. 15–33.
- Pollak, Robert A. 1969. "Conditional Demand Functions and Consumption Theory," *Quart. J. Econ.* 83:1, pp. 60–78.
- Pollak, Robert A. and Michael L. Wachter. 1975. "The Relevance of the Household Production Function and Its Implications for the Allocation of Time," *J. Polit. Econ.* 83:2, pp. 255–77.
- Rose, Elaina. Forthcoming. "Ex Ante and Ex Post Labor Supply Response to Risk in a Low Income Area," *J. Devel. Econ.*
- _____. 1999. "Consumption Smoothing and Excess Female Mortality in Rural India," *Rev. Econ. Statist.* 81:1, pp. 41–49.
- Rosenzweig, Mark R. 1986. "Labor Markets in Low Income Countries," in *Handbook of Development Economics*, Vol. 1. Hollis Chenery and T. N. Srinivasan, eds. Amsterdam; New York and Oxford: Elsevier Science, North Holland.
- Rosenzweig, Mark R. and Kenneth I Wolpin. 1980a. "Testing the Quality-Quantity Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48:1, pp. 227–40.
- _____. 1980b. "Life Cycle Labor Supply and Fertility: Causal Inferences from Household Models," *J. Polit. Econ.* 88:2, pp. 328–48.
- Rosenzweig, Mark R. and T. Paul Schultz. 1982. "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India," *Amer. Econ. Rev.* 72:4, pp. 803–15.
- Rouse, Cecilia Elena. 1999. "Further Estimates of the Economic Return to Schooling from a New Sample of Twins," *Econ. Educ. Rev.* 18:2, pp. 149–57.
- Schultz, T. Paul. 1980. "Estimating Labor Supply Functions for Married Women," in *Female Labor Supply: Theory and Estimation*. James P. Smith, ed. Princeton: Princeton U. Press, pp. 25–89.
- Strauss, Richard S. 2000. "Adult Functional Outcomes of Those Born Small for Gestational Age," *J. Amer. Med. Assoc.* 283:5, pp. 625–32.
- Wald, Abraham. 1940. "The Fitting of Straight Lines If Both Variables Are Subject to Error," *Annals Math. Statist.* pp. 284–300.
- Wolpin, Kenneth I. 1982. "A New Test of the Permanent Income Hypothesis: The Impact of Weather on the Income and Consumption of Farm Households in India," *Int. Econ. Rev.* 23:4, pp. 583–94.