

A Longitudinal Analysis of the Relationship Between Fertility Timing and Schooling

Kevin Stange

Published online: 30 June 2011
© Population Association of America 2011

Abstract This article quantifies the contribution of pre-treatment dynamic selection to the relationship between fertility timing and postsecondary attainment, after controlling for a rich set of predetermined characteristics. Eventual mothers and nonmothers are matched using their predicted birth hazard rate, which shares the desirable properties of a propensity score but in a multivalued treatment setting. I find that eventual mothers and matched nonmothers enter college at the same rate, but their educational paths diverge well before the former become pregnant. This pre-pregnancy divergence creates substantial differences in ultimate educational attainment that cannot possibly be due to the childbirth itself. Controls for predetermined characteristics and fixed effects do not address this form of dynamic selection bias. A dynamic model of the simultaneous childbirth-education sequencing decision is necessary to address it.

Keywords Fertility timing · Educational attainment · Matching

Introduction

Educational attainment and timing of entry into parenthood are intimately related. Individuals who have children early are less likely to graduate from high school, attend college, or receive a college degree. Teenage parents also have lower incomes, are more likely to be in poverty, and are more likely to receive welfare assistance as adults than their peers who delay parenthood. How much these correlations reflect causation or simply unobserved factors has been the subject of much research and debate, but is very important for policy.¹ This debate has focused almost exclusively on the relationship between teenage childbearing and high school graduation, with relatively little attention paid to fertility and schooling after high school. This is an

¹See Hofferth (1987) for a review of the early literature that interpreted this correlation as causal. See Abrahamse et al. (1988) for a discussion of how teenage mothers differ from other mothers.

Electronic supplementary material The online version of this article (doi:10.1007/s13524-011-0050-3) contains supplementary material, which is available to authorized users.

K. Stange (✉)

Gerald R. Ford School of Public Policy, University of Michigan, 5236 Weill Hall, 735 S. State Street,
Ann Arbor, MI 48109-3091, USA
e-mail: kstange@umich.edu

unfortunate gap in the literature, given the increased labor market importance of a college degree and dramatic changes in the timing of fertility over the past several decades.² From 1970 to 2000, the share of first births to mothers aged 23 or younger dropped from 76% to less than 50% (Fig. 1). For most women today, childbearing occurs after the typical college age. This article examines the relationship between fertility at college ages and postsecondary attainment.

To assess the causal impact of fertility timing on education, many researchers have used cross-sectional data with controls for observable characteristics, matching, or sibling fixed effects to overcome the omitted variable, selection, and endogeneity problems that plague identification in this context. Levine and Painter (2003) matched teen mothers to similar nonmothers in the same junior high school as a flexible way of controlling for the many environmental and background factors that may influence both education and fertility timing. Sanders et al. (2007) used a kernel matching estimator based on the estimated propensity score, which should be more efficient than the pair matching estimator used by Levine and Painter (2003). To account for unobserved differences in family background, Geronimus and Korenman (1992) compared sisters whose first births occurred at different ages. Chevalier and Viitanen (2003) and Lee (2010) used binary propensity score matching to form comparison groups. All these studies found that matching (either within schools or sibling pairs or via the propensity score) greatly diminishes but does not eliminate cross-sectional differences in socioeconomic outcomes between teenage mothers and nonmothers. The identifying assumption in these approaches is that any within-match unobserved determinants of fertility timing are uncorrelated with educational decisions, but this is inherently not testable.

Recognizing that fertility timing may not be exogenous even conditional on a rich set of observables, previous researchers have dealt with selection on unobservables in two ways. The most common approach is to use instrumental variables (IV)—for example, miscarriages, twin births, or abortion access—as a source of exogenous variation in fertility timing.³ These studies have found evidence of selection on unobservables, but the bottom-line conclusions have been mixed, with some (e.g., Hotz et al. 2005) finding negligible detrimental effects of teenage fertility and others (e.g., Angrist and Evans 1999) finding larger effects. A second strategy, taken by Sanders et al. (2007), is to quantify the extent of selection on unobservables that would need to be present in order to eliminate the estimated effect. Although useful as a bounding diagnostic, this approach does not allow one to quantify just how problematic selection on unobservables is in a particular setting. I propose a third approach, exploiting the fact that educational attainment results from the accumulation of educational investment made over many time periods. For instance, students typically must accumulate 120 course credits over several years to earn a

² See Katz and Autor (1999) for a review of trends in earnings differences by education.

³ Hotz et al. (2005), Ashcraft and Lang (2006), and Fletcher and Wolfe (2008) used miscarriages and Cristia (2008) used unsuccessful fertility counseling/treatment to identify women who intended to have children but did not or could not. See Bronars and Grogger (1994) and Angrist and Evans (1998) for studies on the use of twin births and child sex mix as determinants of number of children. Ribar (1994) used age at menarche and Olsen and Farkas (1989), Angrist and Evans (1999), and Klepinger et al. (1999) used abortion availability as instruments for teen fertility. Ribar (1999) attempted to reconcile the evidence from sibling matching and IV estimates.

Fig. 1 Change in distribution of age at first birth, 1970 to 2000. Tabulations include all first births occurring in the United States. Mothers younger than 14 or older than 42 are included in these end categories. Data come from author's tabulations from the 1970 and 2000 Natality Public Use Data files available from the National Center for Health Statistics



bachelor's degree. This feature of schooling outcomes permits an identification test of cross-sectional approaches when longitudinal data are also available: treatment (childbirth in this context) should have no effect on intermediate outcomes (such as credits taken in each semester) before the treatment actually occurs. Intermediate outcomes can also be used to quantify the fraction of cumulative outcome differences between treatment and control groups resulting from pre-treatment factors that were not adequately controlled for, whether these factors are fixed or time-varying. This identification test has a long history in labor economics, particularly in the study of job training and welfare programs, but has yet to be applied to the study of fertility and education.⁴

To implement this approach, I match women who had children within eight years of high school to similar women who did not, using the predicted birth hazard rate estimated from a Cox proportional hazard model. The hazard ratio is a sufficient statistic for the entire probability distribution of possible birth outcomes, so it is valid as a "balancing score" in the Rosenbaum and Rubin (1983) framework. This approach generalizes the propensity score to the situation in which treatment is temporal and may prove useful in other applications when the treatment has this characteristic.⁵

The longitudinal approach taken also illuminates the proximate mechanism through which early childbirth and ultimate educational attainment are related. As noted by Moffitt (2005:102), most studies of teenage childbirth "that have attempted to address the endogeneity problem have not been able to determine the mechanism by which postponement affects outcomes, which makes it difficult to interpret the results." This article determines whether mothers leave school earlier, attend college less intensively, or begin on a completely different educational trajectory than nonmothers. While this research cannot directly overcome the endogeneity problem solved by IV, it can show how early, on what dimensions, and with what magnitude the educational trajectories of eventual mothers and nonmothers diverge.

⁴ For an early application of this identification test to the analysis of job training programs, see Ashenfelter and Card (1985).

⁵ To my knowledge, no study has used a multivalued treatment generalization of the propensity score method to study the timing of fertility.

As shown in previous work, I find that cross-sectional estimates indicate that women with earlier first childbirths accumulate significantly fewer college credits and are much less likely to obtain a college degree than those who give birth later or not at all. This difference diminishes, though does not disappear, when a rich set of demographic, family background, fertility and educational expectations, and sexual behavior controls are included. However, longitudinal analysis reveals clear evidence of pre-childbirth dynamic selection even after I control for this rich set of characteristics. The college participation rates and credits taken by mothers and matched nonmothers diverge well before the former become pregnant, even though these two groups begin college at the same rate. No such divergence should exist if the causal effect of childbirth is the only reason educational attainment differs for the two groups. This finding holds for women entering parenthood in many different time periods, though the divergence is greatest for the earliest mothers. Overall, only 56% of the difference in college credits accumulated after high school occurs post-pregnancy and could plausibly be due to childbirth. This calls into question the validity of using nonmothers as a control group even when very rich predetermined matching variables are available. The strong relationship between timing of childbirth and educational attainment appears to be partially due to time-varying factors that cause women to reduce their educational investment and eventually enter parenthood; controls for predetermined characteristics and fixed effects cannot address this source of bias. A dynamic model of the simultaneous childbirth-education sequencing decision seems to be a necessary framework for exploring this topic.

This article is organized as follows. The next section introduces the data. Benchmark cross-sectional results are reported in the third section. The fourth section presents a descriptive analysis of the time profile of postsecondary educational investment and its relationship to fertility timing. In the fifth section, I discuss the parameter of interest and implement a semiparametric matching routine based on the estimated birth hazard rate; this section also quantifies how much of the fertility effect estimated in the cross-section occurs pre-childbirth and thus is not causal. Section six concludes.

Data

The data come from the fourth follow-up of the National Education Longitudinal Study of 1988 (NELS) and the Postsecondary Education Transcript Study (PETS). The NELS/PETS (published by the National Center for Education Statistics) surveyed a nationally representative sample of the high school class of 1992 several times and collected college transcripts for any postsecondary participants. My analysis sample consists of 2,955 women, including 751 eventual mothers. The sample excludes women who gave birth before August 1992 (or high school graduation), graduated before January 1992, or did not participate in either the 1992 or 2000 survey, as well as those for whom transcript data or key covariates were incomplete or missing. The analysis sample is slightly more advantaged on most measures than the full representative sample. Therefore, my results pertain to a slightly more advantaged group than

the actual 1992 cohort of high school seniors. Online Resource 1 provides further details about the sample and also evaluates its representativeness.

From the transcripts, I calculate the number of postsecondary credits taken by institution type in the 16 semesters after high school, as well as indicators for having earned an associate's or bachelor's degree by 2000. Postsecondary participation is indicated by enrolling in at least one credit in a semester. Semester of first childbirth is calculated using the birth date of each respondent's first child, as reported in the 2000 survey.

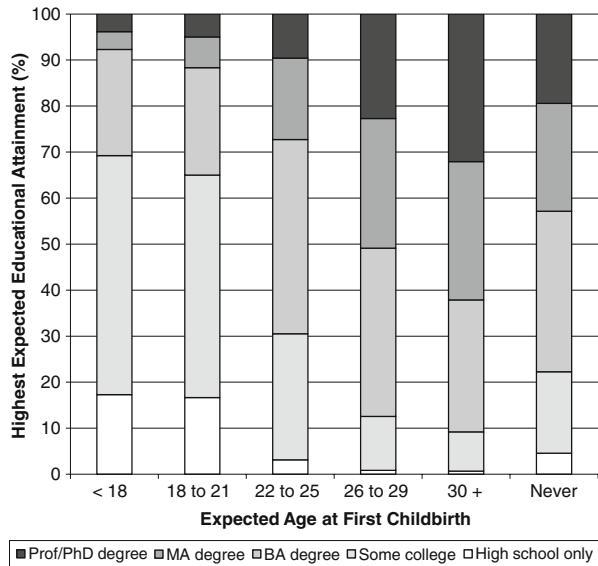
Tables 4 and 5 in Online Resource 1 contain basic summary statistics for the overall sample and separately by the timing of first childbirth. On average, women in the sample accumulated 111 units of college credit within eight years of high school graduation, and approximately half earned a bachelor's degree. Approximately one-quarter gave birth during this period. Women who had children within four years of high school had earned fewer college credits and were much less likely to earn a college degree than those who had children later or not at all. Differences in outcomes, however, may also reflect that young mothers are also more likely to come from disadvantaged backgrounds, as measured by parental education, household income, and sibling teen pregnancy. Young mothers are also less likely to plan to go to college immediately after high school and do take steps toward parenthood earlier: by the end of 1992, 79% of early mothers were sexually active, compared with only 67% and 59% for later mothers and nonmothers, respectively.

Much previous work on the effects of early fertility has been concerned with the likely presence of unobserved or difficult-to-measure factors that would lead some women to begin parenthood early and also obtain less education. Figure 2 documents a strong correlation between two such factors: educational aspirations and expected fertility timing. Women who plan to have children while they are young are also less likely to aspire to earn a bachelor's degree. Almost all women who plan to delay parenthood until their late 20s or later expect to obtain a college degree, and many of these also aspire to earn a graduate or professional degree. Failing to account for this correlation will overstate the adverse consequences of early childbirth on educational attainment.

Cross-Sectional Estimates

As a basis for comparison with previous work, Table 1 reports cross-sectional estimates of the effect of birth timing on postsecondary credit accumulation. Linear models were estimated using least squares. Each model includes eight birth-year indicator variables (one for each academic year of first childbirth) and the indicated control variables. Coefficients in column 1 are unadjusted for covariates; they calculate the unadjusted mean difference in accumulated college credits between mothers and nonmothers, by the academic year of first childbirth. Women who gave birth within the first year of graduating from high school accumulated 88 fewer college credits than women who had not yet had children seven years later. The accumulated credit difference between mothers and nonmothers generally increases with time from high school. Compared with the first cohort of mothers, women who

Fig. 2 Relationship between educational and childbirth expectations



gave birth in the eighth year after high school accumulated 62 more college credits, yet still accumulated 26 fewer than nonmothers.

Columns 2 and 3 include an increasingly rich set of control variables observed in respondent’s senior year of high school, including family background, educational expectations, expected timing of childbirth, and sexual behavior. Accumulated credit differences are reduced considerably—sometimes by more than 50%—when controls are included, but the differences are not eliminated entirely. Even after background, expectations, and sexual behavior are controlled for, the earliest mothers accumulated 30 fewer college credits than nonmothers. Controlling for a rich set of covariates cuts the 53 overall credit difference between mothers and nonmothers in half. All differences are significant at conventional levels.

While not directly comparable to previous work, these cross-sectional estimates are slightly smaller than those found in recent studies using abortion laws and the availability of family planning services as instrumental variables when studying fertility timing and total years of schooling. Kleplinger et al. (1999) found that teenage childbearing reduces years of schooling by about two and one-half years, which is roughly equivalent to 60 college credits.

Mothers and nonmothers may also differ in the type of postsecondary education they receive. Community colleges, which have more adult and part-time learners and flexible schedules, may be more accommodating to young mothers trying to blend parenthood and school. The second and third panels of Table 1 repeat the analysis, using cumulative credits at four-year and two-year institutions as the dependent variable. The relationship between fertility timing and educational attainment is stronger for four-year college credits than for credit accumulation overall. In fact, mothers actually accumulate slightly more community college credits than nonmothers. Fertility timing is thus related to both the amount and the type of education, with earlier mothers receiving less education overall and at four-year institutions in particular.

Table 1 Cross-sectional OLS estimates of fertility timing on credit accumulation

Academic Year of First Childbirth	Total Cumulative Credits			Cumulative Four-Year Credits			Cumulative Two-Year Credits		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1992–1993	-87.74 (10.46)	-58.94 (10.93)	-29.84 (11.91)	-91.49 (9.18)	-61.10 (9.99)	-33.49 (12.17)	1.83 (5.96)	0.50 (5.86)	2.10 (6.37)
1993–1994	-71.44 (6.10)	-52.26 (5.62)	-30.41 (5.94)	-80.55 (5.78)	-59.53 (5.63)	-37.18 (5.92)	9.47 (4.15)	7.52 (4.22)	7.05 (4.23)
1994–1995	-73.75 (5.04)	-54.51 (4.86)	-35.45 (4.70)	-72.14 (4.68)	-50.63 (4.70)	-30.12 (4.66)	-1.58 (2.78)	-3.95 (2.94)	-5.44 (2.84)
1995–1996	-63.23 (5.48)	-48.24 (4.91)	-36.24 (4.80)	-69.01 (5.26)	-51.68 (4.78)	-39.31 (4.75)	6.05 (3.04)	3.59 (3.06)	3.18 (3.09)
1996–1997	-57.94 (6.12)	-45.13 (5.66)	-31.83 (5.29)	-64.31 (6.00)	-49.63 (5.69)	-33.82 (5.41)	5.55 (3.06)	3.47 (3.09)	1.07 (3.17)
1997–1998	-37.96 (6.05)	-29.66 (5.63)	-19.85 (4.98)	-45.48 (6.76)	-35.79 (6.43)	-24.64 (5.74)	6.68 (3.49)	5.04 (3.49)	3.72 (3.44)
1998–1999	-32.14 (5.74)	-20.62 (5.43)	-17.68 (4.77)	-33.28 (6.30)	-20.54 (5.81)	-17.62 (5.08)	0.76 (2.99)	-0.54 (2.99)	-0.46 (2.93)
1999–2000	-25.70 (6.13)	-14.34 (5.70)	-12.32 (4.74)	-31.94 (6.83)	-19.06 (6.45)	-16.76 (5.44)	6.61 (3.24)	5.03 (3.21)	4.73 (3.16)
R^2	.130	.276	.465	.126	.290	.469	.007	.044	.083
1992–2000	-52.88 (2.61)	-38.20 (2.50)	-26.25 (2.33)	-57.31 (2.72)	-40.88 (2.64)	-28.10 (2.46)	4.22 (1.34)	2.35 (1.41)	1.55 (1.43)
R^2	.111	.264	.461	.111	.280	.466	.003	.040	.079
Control Variables	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Background									

Table 1 (continued)

	Total Cumulative Credits			Cumulative Four-Year Credits			Cumulative Two-Year Credits		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Academic Year of First Childbirth	No	No	Yes	No	No	Yes	No	No	Yes
Educational expectations	No	No	Yes	No	No	Yes	No	No	Yes
Expected age at first childbirth	No	No	Yes	No	No	Yes	No	No	Yes
Age became sexually active	No	No	Yes	No	No	Yes	No	No	Yes
Observations	2,955	2,955	2,955	2,955	2,955	2,955	2,955	2,955	2,955

Notes: Robust standard errors are in parentheses. Academic year is August through July. The omitted category is women who have not yet had children as of July 2000. Background variables include race, Hispanic ethnicity, parent's highest education, mother's and father's employment status, log of family income, and having a sister with a high school pregnancy. Educational expectations include own educational expectations and an indicator for planning to go directly to college after high school.

Table 2 repeats this analysis, using degree attainment as the dependent variable. Early mothers are 30% to 40% less likely than nonmothers to obtain an associate’s or bachelor’s degree within eight years of high school graduation, even after observable differences in background, expectations, and sexual behavior are accounted for. Degree attainment differences are not diminished as much as accumulated credit differences when these observable characteristics are accounted for.

The identifying assumption in this and previous cross-sectional analyses is that individuals’ timing of first childbirth is uncorrelated with unobservable determinants

Table 2 Cross-sectional probit estimates of fertility timing on degree attainment

Academic Year of First Childbirth	Associate’s Degree or Higher			Bachelor’s Degree or Higher		
	(1)	(2)	(3)	(4)	(5)	(6)
1992–1993	–0.516 (0.054)	–0.443 (0.087)	–0.366 (0.116)	–0.490 (0.028)	–0.459 (0.060)	–0.403 (0.072)
1993–1994	–0.455 (0.041)	–0.396 (0.054)	–0.303 (0.077)	–0.449 (0.029)	–0.413 (0.042)	–0.358 (0.048)
1994–1995	–0.498 (0.029)	–0.448 (0.038)	–0.393 (0.046)	–0.445 (0.025)	–0.399 (0.034)	–0.326 (0.040)
1995–1996	–0.434 (0.035)	–0.385 (0.044)	–0.353 (0.051)	–0.449 (0.025)	–0.432 (0.031)	–0.395 (0.031)
1996–1997	–0.353 (0.043)	–0.310 (0.047)	–0.254 (0.054)	–0.379 (0.033)	–0.350 (0.040)	–0.294 (0.044)
1997–1998	–0.273 (0.048)	–0.245 (0.052)	–0.203 (0.059)	–0.296 (0.041)	–0.277 (0.046)	–0.235 (0.050)
1998–1999	–0.211 (0.046)	–0.149 (0.050)	–0.149 (0.053)	–0.181 (0.044)	–0.114 (0.049)	–0.105 (0.053)
1999–2000	–0.113 (0.052)	–0.054 (0.054)	–0.059 (0.058)	–0.173 (0.048)	–0.111 (0.054)	–0.118 (0.056)
Pseudo- <i>R</i> ²	.089	.173	.285	.097	.203	.355
1992–2000	–0.350 (0.020)	–0.285 (0.022)	–0.242 (0.025)	–0.372 (0.019)	–0.312 (0.021)	–0.272 (0.024)
Pseudo- <i>R</i> ²	.071	.159	.277	.079	.187	.345
Control Variables						
Background	No	Yes	Yes	No	Yes	Yes
Educational expectations	No	No	Yes	No	No	Yes
Expected age at first childbirth	No	No	Yes	No	No	Yes
Age became sexually active	No	No	Yes	No	No	Yes
Observations	2,955	2,955	2,955	2,955	2,955	2,955

Notes: Reported coefficients are marginal effects. Robust standard errors are in parentheses. Academic year is August through July. The omitted category is women who have not yet had children as of July 2000. Background variables include race, Hispanic ethnicity, parent’s highest education, mother’s and father’s employment status, log of family income, and having a sister with a high school pregnancy. Educational expectations include own educational expectations and an indicator for planning to go directly to college after high school.

of educational attainment. This assumption is inherently not testable with cross-sectional data, although evidence from siblings models suggests it is likely violated (Geronimus and Korenman 1992). Longitudinal data can be used to partially address this shortcoming. The educational outcomes analyzed in Tables 1 and 2, and in most of the previous literature, are the cumulative result of educational investments made incrementally over many time periods. Receiving a bachelor's degree is not a decision made at one point in time. Rather, it is the end result of a sequence of decisions to enroll in and complete college courses over a span of four to eight or more years. The remainder of this article exploits this cumulative nature of educational attainment to partially test the identification assumption. Intuitively, the pre-birth course-taking of women who enter childbirth early and women in an appropriate control group should be similar. Cross-sectional approaches do not permit such a test.

Descriptive Longitudinal Analysis

This section documents the temporal relationship between childbirth and educational investment during the eight years following high school, differentiating between several possible mechanisms through which early childbirth is negatively related to educational attainment. Young mothers may leave school earlier, attend college less intensively, or begin on a completely different educational trajectory than nonmothers, but cross-sectional analyses cannot distinguish between these mechanisms.

Figure 3 plots the postsecondary participation rate and average number of postsecondary credits taken by women in the first 16 semesters (eight years) following high school, separately by semester of first childbirth.⁶ Women in the final panel had not yet had a child by the spring of 2000. Their enrollment and course-taking is as expected from traditional college students: on average, they take 12 to 14 credits per semester (approximately full-time) for four years, then reduce investment levels significantly thereafter. The drop-off in credit accumulation precisely after eight semesters is much smaller for women who had children, although smaller sample sizes make these series much noisier than for nonmothers.

The vertical line in each graph indicates the semester of first childbirth. In most cases, educational investment declines steadily in the semesters leading up to childbirth and flattens out afterward. There is no evidence of a precipitous drop in educational investment precisely at the time of childbirth; the investment decline is much more gradual. Women learn of their pregnancy either one or two semesters before their birth. There does not appear to be an acceleration of the decline during the semesters the conception is known. Looking across all panels, however, there is clear evidence that later childbirth is associated with later attrition from school: the earliest mothers begin dropping out before slightly later

⁶ Because more of the summer months were included in the spring, the number of credits taken in the spring tends to be greater than the number of credits taken in the fall.

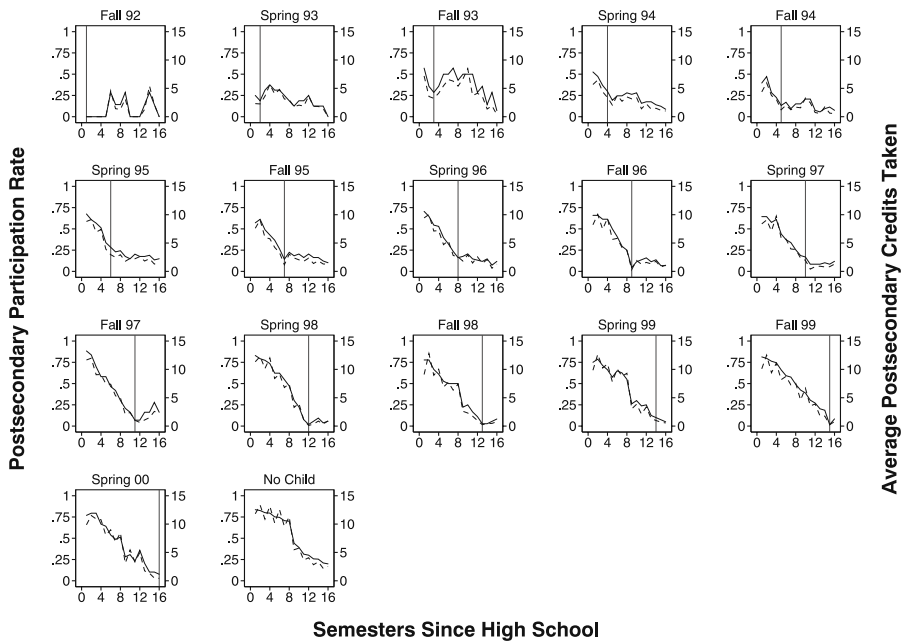


Fig. 3 Postsecondary participation rate and average number of credits taken, by semester of first childbirth. Solid lines denote participation rate, and dashed lines denote average number of postsecondary credits taken. Vertical lines indicate the semester of first childbirth. Postsecondary participation is indicated by taking at least one unit of credit in the semester. Spring includes the months January through July; fall includes August through December

mothers, who drop out before slightly later mothers, and so on. The precise timing of childbirth seems to matter.

The time profile of participation and total average credits taken are remarkably similar, save for the spring-fall seasonality in credits. This suggests that nearly all adjustment in postsecondary educational investment occurs on the participation margin, rather than in investment intensity. Figure 8 in Online Resource 1 plots the average credits taken, conditional on taking at least one college credit. There is very little discernible adjustment in the number of credits individuals take, either over time or with childbirth. On average, those participating in postsecondary education take 10 to 15 units of college credit both before and after childbirth and regardless of how much time has passed since high school.

Figure 3 demonstrates an important element of heterogeneity in the sample. Women who gave birth in 1994 or earlier began college with much lower participation rates; as a result, they accumulated far fewer credits than those who began parenthood later, even before the onset of parenthood. There is much more pre-birth similarity among women who began parenthood in 1995 or later, although women who delayed parenthood until at least 1997 appear to have started off even stronger, participating in college in large numbers. The next section addresses this initial condition heterogeneity by matching mothers with nonmothers who have similar estimated birth hazard rates. This approach is similar in spirit to propensity score methods used to deal with nonrandom treatment assignment, when assignment is assumed to be random conditional on observable covariates. The primary

limitation of this approach is that differences attributable to unobserved factors cannot be accounted for.

Longitudinal Matching Results

Evaluation Approach

This article is concerned with the causal effect of childbirth timing on the time path of educational investment. To clarify ideas, it is useful to characterize the problem using Rubin's (1974) potential outcomes framework. Suppose that women can have their first child during $(T+1)$ different time periods, corresponding to the first T semesters after graduating from high school plus the post-observation period. In what follows, I refer to first childbirth timing as $(T+1)$ mutually exclusive treatments, in the language of the treatment effect literature.⁷ Participation in a particular treatment s is indicated by the variable $S \in \{1, 2, \dots, T, T^+\}$. Treatment T^+ refers to first childbirth occurring some time after the observation period and also includes women who will never have children. The outcome of interest is the entire time profile of educational investment (e.g., participation or credits taken) in the semesters following high school graduation. The potential outcomes associated with the various treatments are denoted by $Y^{(s)} \equiv \{Y_t^{(s)}\}_{t=1}^T$.

The parameters of interest, $\Delta^{(s, T^+)}$, are the average causal effects of having a child s semesters after high school relative to not having a child during the observational period.

$$\Delta^{(s, T^+)} = E[Y^{(s)} - Y^{(T^+)} | S = s] = E[Y^{(s)} | S = s] - E[Y^{(T^+)} | S = s] \text{ for } s = 1, 2, \dots, T. \quad (1)$$

There are many different causal effects that could be examined, such as the effect of having a child during period s versus period $s + 1$ ($\Delta^{(s, s+1)}$) or during period s versus some period after s ($\Delta^{(s, [s+1, T^+])}$).⁸ The multitude of parameters that could be estimated may partially explain the range of findings from previous work. For instance, the ages between which childbearing is shifted by abortion availability may be different than those for which childbearing is affected by miscarriages. Each of these instruments (potentially) identifies different parameters. Similarly, some previous matching estimates contrasted fertility in high school with later fertility,

⁷ I focus here on first births because they are likely to have the biggest impact on educational choices. The impact of having additional children will be part of the treatment effect estimated here.

⁸ For example, Sianesi (2004) examined the causal effect of receiving job services after being unemployed for t periods versus not receiving services at (or before) time t . She argued that this is the treatment effect of greatest interest to policy-makers and individuals, and cautioned that using the "never-treated" as a comparison group is equivalent to conditioning on successful employment outcomes. This concern is less problematic in the current context because fertility timing is not as directly driven by educational failure as job service receipt is driven by job search failure.

while others focused on comparisons between specific age groups, such as 16/17 versus 18/19. The strong age pattern to educational outcomes makes the decision of which causal effect to focus on a nontrivial one. Although this article emphasizes $\Delta^{(s,T^+)}$, I also examine the effect of delaying childbearing until some later period $\mathcal{I}(\Delta^{(s,\mathcal{I},T)})$ for $s = 1, 2, \dots, \mathcal{I} - 1$

The parameters $\Delta^{(s,T^+)}$ are not identified without further assumptions because of the well-known causal inference problem: the counterfactual outcome $E[Y^{(T^+)}|S = s]$ is not observed in the data. We cannot observe what the educational profile of women who had children in period s would have been had they delayed childbirth until after period T . To identify the counterfactual, I assume that potential outcomes are independent of treatment, conditional on a set of observable attributes X . I assume that among individuals with a given set of baseline characteristics $X = x$, the unobserved distribution of $Y^{(s)}$ for individuals with $S \neq s$ is the same as the observed distribution of $Y^{(s)}$ for individuals with $S = s$ for all $s = 1, 2, \dots, T^+$.

$$Y^{(1)}, Y^{(2)}, \dots, Y^{(T^+)} \perp S | X = x. \tag{2}$$

With this conditional independence assumption (CIA), the counterfactual is identified from the observed outcomes of individuals in the comparison group:

$$E[Y^{(T^+)}|S = s] = E_{X|S=s} \left[E[Y^{(T^+)}|X, S = T^+] \right]. \tag{3}$$

The outer expectation is taken over the distribution of X in the treatment group s . In the case of a binary treatment ($S \in \{0,1\}$), Rosenbaum and Rubin (1983) showed that if the CIA holds conditional on X , then it also holds conditional on specific functions of X , referred to as balancing scores and denoted by $b(X)$. This property is very convenient when X is high dimensional and contains continuous covariates, as is the case here. The most commonly used balancing score is the propensity score $P(X) \equiv \Pr(S = 1|X)$, the conditional probability of treatment. Imbens (2000) and Lechner (2001) showed that this property also holds in the more general case in which treatments are multivalued. They showed that Eq. 3 can be replaced by:

$$E[Y^{(T^+)}|S = s] = E_{b(X)|S=s} \left[E[Y^{(T^+)}|b(X), S = T^+] \right] \tag{4}$$

if the balancing score satisfies $E[\Pr(S = s|X = x)|b(X) = b(x)] = \Pr[S = s|X = x] \equiv P^{(s)}(x)$ and $0 < P^{(s)}(x) < 1$ for all $s = 1, 2, \dots, T^+$. The entire vector of treatment probabilities $[P^{(1)}(x), P^{(2)}(x), \dots, P^{(T^+)}(x)]$ satisfies this condition. These probabilities can be estimated by a multinomial probit or logit model, for instance. One limitation is that the dimensionality is reduced only to the order of T , which may still pose common support problems if the number of treatments is large. Consequently, Lechner (2001 and 2002) concentrated on pairwise treatment effects (e.g., between s and s'), where one needs only to condition on the binary conditional probabilities $\Pr(S = s|X = x, S \in (s, s'))$. Here, I make use of the fact that the multivalued treatment is temporal in my setting and model the distribution of treatments using a Cox proportional hazard model, where time is given by months since high school:

$$h_i(t_m) = h_0(t_m) \cdot e^{\beta X_i + \epsilon_i} . \tag{5}$$

$h_i(t_m)$ is the probability that individual i has a child during the t_m month after high school, having not had a child by the end of the previous month.⁹ The baseline hazard rate $h_0(t_m)$ is allowed to vary over time, but it is left unspecified and is not directly estimated. Individual heterogeneity enters multiplicatively on this baseline hazard. In the estimation, I include a rich set of covariates observed at the end of high school, including race, Hispanic ethnicity, parent’s highest education level, mother’s and father’s labor force participation, log of family income, sister’s high school pregnancy, own educational expectations and plans to attend college immediately, expected age of first childbirth, year of first sexual intercourse, and use of birth control during first sexual intercourse.¹⁰ The predicted hazard rate from the maximum likelihood estimation of Eq. 5 parsimoniously summarizes the probability that an individual will have a child early, based on characteristics that are observable at the end of high school. I use the predicted hazard ratio as the balancing score:

$$b(X) = r(X) \equiv e^{\beta X} . \tag{6}$$

In the Cox proportional hazard model, this hazard ratio is a sufficient statistic for the full vector of treatment probabilities. Recall that each fall semester contains five months, and each spring semester contains seven. The cumulative hazard is defined as $H_0(t_m) \equiv \int_0^{t_m} h_0(\tau) d\tau$. The vector of treatment probabilities is given by

$$\begin{aligned} P^{(1)}(X) &= 1 - e^{-r(X)}(e^{-H_0(5)}) \\ P^{(2)}(X) &= e^{-r(X)}(e^{-H_0(5)} - e^{-H_0(12)}) \\ P^{(3)}(X) &= e^{-r(X)}(e^{-H_0(12)} - e^{-H_0(17)}) \\ &\vdots \\ P^{(T^+)}(X) &= e^{-r(X)}(e^{-H_0(96)}) . \end{aligned} \tag{7}$$

Based on the results of Imbens (2000) and Lechner (2001), the counterfactual outcome for eventual mothers $E[Y^{(T^+)}|S = s]$ can thus be estimated from the observed outcome of nonmothers with a similar hazard ratio:

$$E[Y^{(T^+)}|S = s] = E_{r(X)|S=s} [E[Y^{(T^+)}|r(X), S = T^+]] . \tag{8}$$

An alternative approach taken in the dynamic treatment participation literature is to match on the basis of covariates and pre-treatment outcomes directly (Fredriksson and Johansson 2008; Lechner 1999) or on a time-varying discrete hazard rate implemented using a sequence of probit regressions (Sianesi 2004). Since treatment timing is treated as stochastic, some eventual late-treated observations in this approach are used in the construction of the unobserved counterfactual for earlier-treated observations.

The estimated hazard rate at the end of high school does a decent job predicting the timing of first births. Figure 9 in Online Resource 1 plots the median and spread of predicted hazard rate by actual year of first childbirth. Women who gave birth earliest had the highest risk of childbirth, based on factors observable at the end of

⁹ In estimating the hazard rate, I specify months since July 1992 as the time an individual is at risk before the event (childbirth) occurs. Nonmothers are included in the hazard rate estimation, but their time at risk is censored at 96 months.

¹⁰ Results from estimation of the proportional hazard model are contained in Table 7 in Online Resource 1.

high school. The median hazard rate generally decreases, though not monotonically, as childbirth is delayed. Fortunately for the matching strategy described above, there is considerable overlap across childbirth years. For instance, most mothers who gave birth within a year of high school graduation can be matched to a comparable nonmother who was at a similar risk for early childbirth at the end of high school. I use these similarly risked nonmothers as a control group in the estimation of the counterfactual educational investment profile of eventual mothers. This approach is a specific form of matching whereby the weight that each control variable receives in the balancing score is related to its importance in predicting birth timing.

In the main analysis, I implement matching through stratification of the balancing score (also known as interval, blocking, or subclassification; see Rosenbaum and Rubin 1984).¹¹ All individuals were sorted by predicted hazard ratios and divided into 10 ordered groups on the basis of this ranking. As a robustness check, I also present results using kernel and nearest-neighbor matching. Table 8 in Online Resource 1 displays the mean difference for several important characteristics between mothers and nonmothers within each hazard group.¹² By construction, mothers and nonmothers with similar estimated hazard rates are very similar on observable characteristics. In no hazard rate group is there a significant difference between mothers and nonmothers in more than one of the characteristics examined. Based on this quality of balance across observable characteristics and the predictive power of the predicted hazard rate, I use similarly risked nonmothers to estimate the counterfactual for each mother in my sample.

Pre-treatment Test of the Identification Assumption

Strictly speaking, the conditional independence assumption is not testable. However, longitudinal data containing observations of intermediate outcomes (such as credits taken in each semester) before the treatment actually occurs can suggest whether it is likely to be violated. The pre-treatment effect is defined as the average difference in accumulated credits between eventual mothers and their unobserved counterfactual in the periods before their first childbirth actually occurs:

$$\Delta_{pre}^{(s,T^+)} \equiv E\left[Y_{pre}^{(s)}|S = s\right] - E\left[Y_{pre}^{(T^+)}|S=s\right] \text{ for } s = 1, 2, \dots, T,$$

where $E\left[Y_{pre}^{(s)}|S = s\right] \equiv E\left[\sum_{t=1}^{t=s-1} Y_t^{(s)}|S = s\right]$. As described above, the unobserved counterfactual is estimated using nonmothers with similar predicted hazard rates. If the process generating pre-birth outcomes and post-birth outcomes is the same in the absence of treatment and if the treatment has no effect on pre-birth outcomes, then the CIA implies that $\Delta_{pre}^{(s,T^+)} = 0$ for all s . Nonzero estimates of $\Delta_{pre}^{(s,T^+)}$ suggest that pre-treatment factors (either fixed or time-varying) were not adequately controlled for, violating the conditional independence assumption.

In many settings (and in most of the fertility-education literature) researchers have access only to cross-sectional post-treatment outcome data, so this pre-treatment test is not possible. Previous studies thus controlled for selection using pre-treatment

¹¹ For a practical overview of implementing propensity score matching, see Caliendo and Kopeinig (2005).

¹² Table 9 in Online Resource 1 contains the group means in addition to their difference.

baseline covariates only. With longitudinal data, however, one can use the pre-treatment outcomes to test the CIA of this cross-sectional work. Heckman and Hotz (1989) termed this a “pre-program specification test,” and it has been used extensively in the evaluation of job training and welfare programs but has yet to be applied to the study of fertility and education. In addition to quantifying the extent of pre-treatment selection on unobservables, pre-treatment outcomes can also be used in the matching process itself, as done in the program evaluation literature (Card and Sullivan 1988; Dolton et al. 2008; Heckman and Smith 1999). Flexibly matching on pre-program outcomes is one strategy for eliminating selection bias in non-experimental evaluations.¹³

Longitudinal Estimates

To construct a counterfactual educational investment profile for mothers, I regress the outcomes of nonmothers on a vector of 160 group \times time indicator variables (10 hazard rate groups \times 16 time periods).

$$y_{i,t} = \gamma_{j,t} \times 1(\text{Group}_i = j, \text{Time} = t) + \varepsilon_{i,t}. \quad (9)$$

The estimated parameters $\{\gamma_{j,t}\}$ are the average number of credits taken (or average participation rate) by nonmothers in hazard group j at time t . The coefficients obtained from estimating Eq. 9 are used to predict credits taken and participation for mothers in the sample.

Figure 4 plots the actual (solid lines) and predicted (dashed lines) postsecondary participation rate separately by semester of first childbirth.¹⁴ As before, the vertical line indicates the semester of first childbirth. Several features of Fig. 4 are striking. Most important, the actual participation rate deviates from its prediction well before the event of pregnancy for most groups. For example, women who gave birth in spring of 1996 had similar participation rates as nonmothers with similar birth hazards immediately following high school. By fall of 1994, however, the participation rate of eventual mothers was 20 percentage points lower, even though their future pregnancy was still unknown. Participation rates fall even further immediately preceding and during the semester of childbirth. This general pattern of significant pre-birth decline in relative participation rate holds for most birth cohorts prior to 1999.¹⁵ Women whose first childbirth occurred in the spring of 1999 or later had similar postsecondary participation rates as nonmothers with comparable birth hazards, presumably because most of those who would earn bachelor's degrees did so by this time and participation in graduate education is less deterred by parenthood.

¹³ If pre-treatment outcomes are used in the matching process, then one cannot test the CIA itself. In this article, I use pre-treatment outcomes to test the CIA associated with matching on the basis of baseline covariates exclusively.

¹⁴ An analogous analysis on average postsecondary credits taken (not reported here) produced nearly identical qualitative and quantitative results.

¹⁵ This does not apply to mothers in the first two birth cohorts for which pre-pregnancy educational experience does not exist.

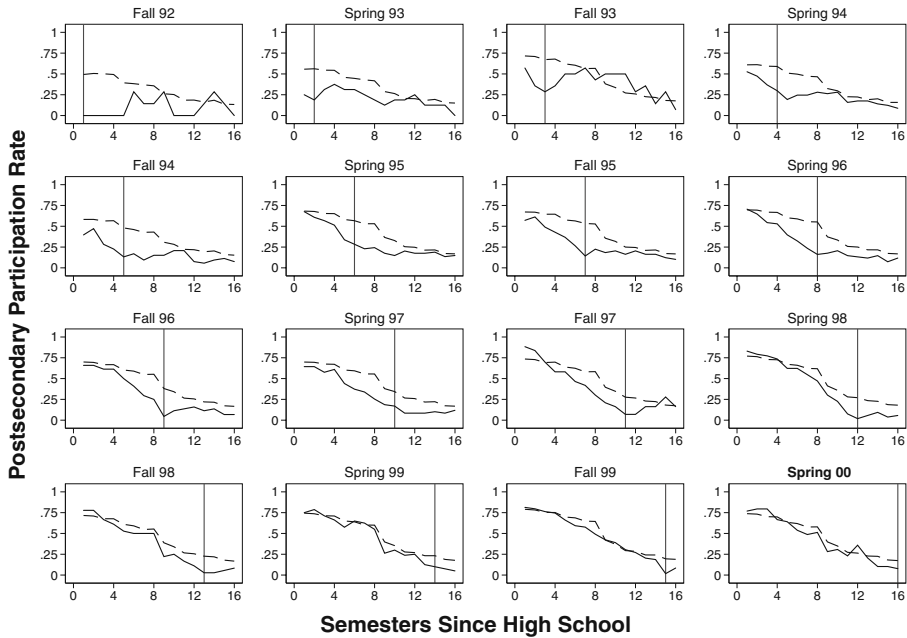


Fig. 4 Predicted and actual postsecondary participation rate, by semester of first childbirth. Solid lines indicated actual participation rate of mothers with the indicated birth year. Dashed lines are participation rates predicted using nonmothers with similar predetermined characteristics. Postsecondary participation is indicated by taking at least one unit of credit in the semester. Vertical lines indicate the semester of first childbirth. Spring includes the months January through July; fall includes August through December

Figure 5 plots estimates of $\Delta^{(s, T^+)}$, the actual minus predicted postsecondary participation rates following high school, separately by semester of first childbirth (s). If educational investments of mothers followed the same time path as those of nonmothers with comparable fertility risk at the end of high school, then each plot would lie precisely on the zero line. Alternatively, if mothers took a fixed number of fewer courses than observably similar nonmothers, each plot would be a horizontal line below zero. Both of these possibilities are clearly rejected by the data. The temporal pattern of educational investment for mothers deviates significantly from that for observably similar nonmothers over the eight years following high school. The deviation is most significant for the earliest mothers and appears to be approximately centered and symmetric around the time of first birth. Mothers reduce postsecondary participation in the years leading up to their first childbirth, then gradually catch up to their nonmother peers (albeit to a lower absolute level of participation) in the years following. The reduction in educational investment associated with childbirth is gradual and begins several semesters before the actual birth occurs, not sharply at the time of birth.

Figure 6 distinguishes between credits earned at four-year and two-year institutions. With the exception of the earliest cohort of mothers, four-year institutions account for the great majority of the credit difference between mothers and matched nonmothers. The initial credit accumulation rate at four-year colleges is similar for nonmothers and several cohorts of mothers. Therefore, some of the four-year credit difference between mothers and matched nonmothers is due to a failure of matching to control completely for pre-determined characteristics that predict college type. However, the general pattern

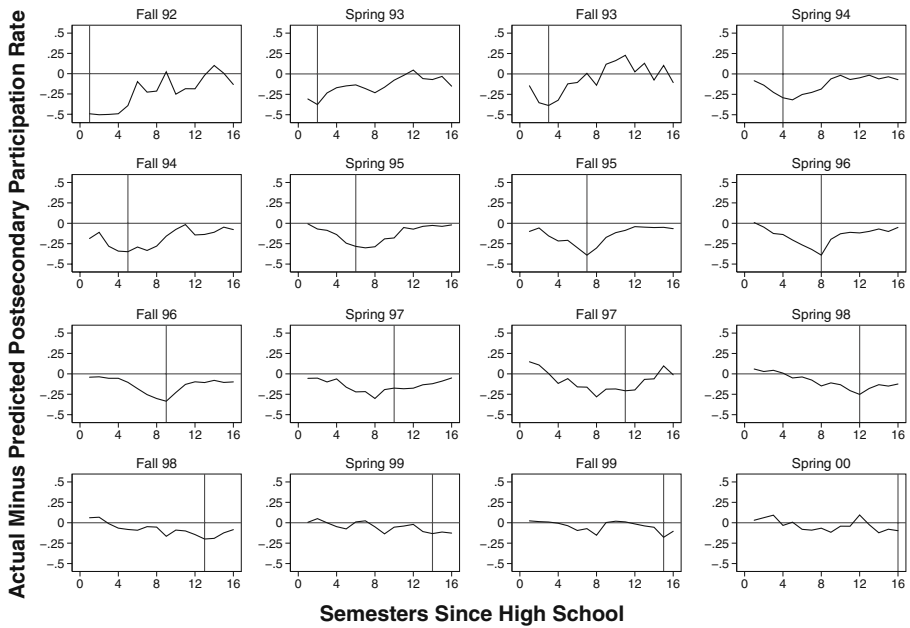


Fig. 5 Deviation from predicted average postsecondary participation rate, by semester of first childbirth. Vertical lines indicate the semester of first childbirth. Postsecondary participation is indicated by taking at least one unit of credit in the semester. Spring includes the months January through July; fall includes August through December

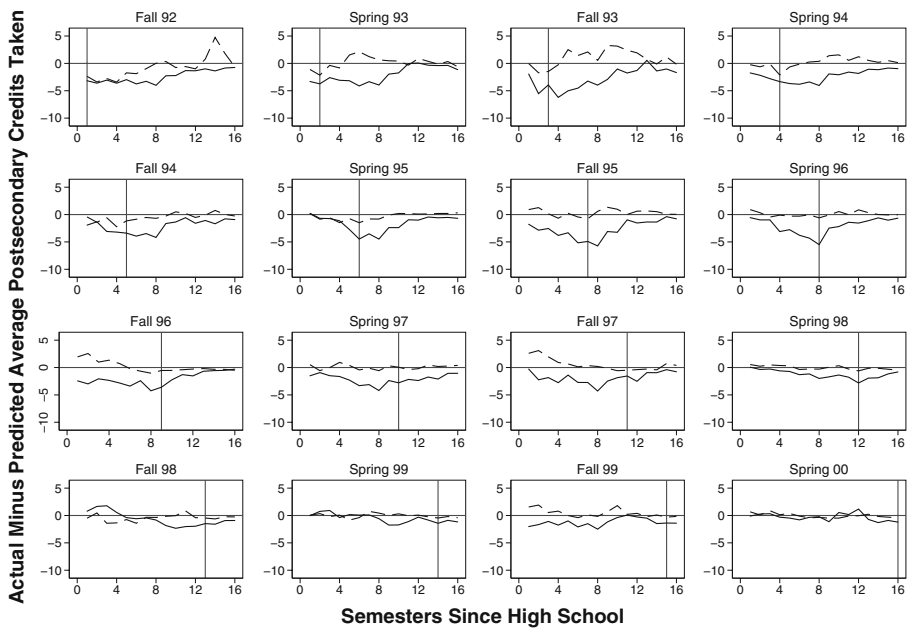


Fig. 6 Deviation from predicted average postsecondary four-year and two-year college credits, by semester of first childbirth. Solid lines represent four-year college credits. Dashed lines represent two-year college credits. Vertical lines indicate the semester of first childbirth. Spring includes the months January through July; fall includes August through December

of substantial pre-pregnancy divergence between mothers and matched nonmothers continues to hold for four-year credits as it does for credits overall.

The large differences in ultimate educational attainment presented in Tables 1 and 2 stem from the integral of this deviation over the full observational period, and clearly much of the accumulated difference arises well before the advent of pregnancy. The main benefit of a longitudinal approach is that this pre-pregnancy difference can be quantified. Table 3 decomposes the accumulated credit difference into differences arising pre- and post-childbirth. As a point of comparison, column 1 repeats the OLS cross-sectional estimates from Table 1. Total accumulated credit differences between mothers and nonmothers estimated with matching, presented in column 2, are very similar to these OLS estimates. Standard errors of the matching estimators were constructed using bootstrapping.¹⁶ The credit difference that arises before childbirth is presented in column 3, and the fraction of the total difference that occurs pre-birth is shown in column 4. Although estimates for some birth cohorts are imprecise, all the point estimates indicate that a large fraction of credit loss for mothers relative to similar nonmothers occurs before their first child is born. The condition that $\Delta_{pre}^{(s,T^+)} = 0$ is clearly rejected. Overall, only 56% of the accumulated credit difference between mothers and matched nonmothers arises post-childbirth, suggesting that the true causal effect is, at most, about half of what the cross-sectional estimates suggest, even after a rich set of covariates is controlled for.

The last four columns of Table 3 repeat this analysis, using late mothers as a control group for earlier mothers. Early mothers are expected to be more similar to late mothers than to nonmothers, so this comparison will reduce omitted variable bias. However, the treatment effect estimated by this comparison is that of having an early birth relative to a late one, so the results are not directly comparable to previous estimates. Early mothers accumulate 14 fewer credits than late mothers, and a large fraction of this difference occurs before the former has children, though estimates are much less precise because of a smaller number of control observations. Regardless, the conclusion that a large share of credit differences occur pre-childbirth still holds when early mothers are compared with late mothers.

The sensitivity of the main results in Table 3 to choice of matching method is examined in Table 10 in Online Resource 1. Although many matching estimators are consistent given the conditional independence assumption, they do behave differently in finite samples, and characterizing the finite sample properties of various types of matching estimators is an active area of research. For instance, Frölich (2004) presented Monte Carlo simulations suggesting that kernel and ridge matching estimators performed well (in an efficiency sense) relative to pair matching, k -nearest-neighbor matching (of which stratification is a variant), and inverse propensity score weighting, particularly when the number of controls was large.¹⁷ Based on this evidence, Sanders et al. (2007) employed a kernel matching estimator, which should be more efficient than the nearest-neighbor matching estimator used by Levine and

¹⁶ Though Abadie and Imbens (2008) cautioned that the standard bootstrap may not be valid for simple nearest-neighbor matching with a fixed number of matches, the bootstrap should provide valid inference for kernel-based matching estimators in which the number of matches increases with the sample size.

¹⁷ Busso et al. (2009) refuted this latter result, finding that properly constructed reweighting estimators outperform matching estimators in realistic empirical settings, rather than the unrealistic hypothetical setting examined by Frölich (2004).

Table 3 Matching estimates of fertility timing on total and pre-birth credit accumulation

Academic Year of First Childbirth	Dependent Variable: Cumulative Postsecondary Credits							
	Control Group: Nonmothers				Control Group: Late Mothers			
	OLS	Matching			OLS	Matching		
	Total (1)	Total (2)	Pre-birth (3)	% Pre-birth (4)	Total (5)	Total (6)	Pre-birth (7)	% Pre-birth (8)
1992–1993 [n = 23]	–29.84 (11.91)	–34.03 (11.86)	–3.25 (1.18)	0.096 (0.041)	–19.71 (11.25)	–9.65 (13.93)	–2.34 (1.28)	0.243 (5.434)
1993–1994 [n = 71]	–30.41 (5.94)	–32.41 (6.02)	–7.82 (1.94)	0.241 (0.048)	–14.72 (5.90)	–13.39 (7.46)	–8.05 (2.56)	0.602 (1.303)
1994–1995 [n = 127]	–35.45 (4.70)	–35.45 (4.92)	–10.66 (2.51)	0.301 (0.047)	–19.25 (5.14)	–15.23 (6.03)	–8.34 (3.21)	0.548 (0.265)
1995–1996 [n = 117]	–36.24 (4.80)	–34.40 (5.16)	–16.95 (3.14)	0.493 (0.051)	–18.41 (5.00)	–14.77 (5.77)	–11.53 (3.85)	0.781 (0.413)
1996–1997 [n = 103]	–31.83 (5.29)	–32.29 (5.20)	–18.93 (4.44)	0.586 (0.071)				
1997–1998 [n = 96]	–19.85 (4.98)	–19.59 (5.71)	–10.84 (5.44)	0.553 (0.262)				
1998–1999 [n = 116]	–17.68 (4.77)	–11.35 (5.29)	–6.03 (5.06)	0.532 (15.35)				
1999–2000 [n = 98]	–12.32 (4.74)	–10.14 (5.70)	–7.43 (5.63)	0.733 (1.977)				
R^2	.465				.399			
1992–2000 [n = 751]	–26.25 (2.33)	–25.47 (2.47)	–11.17 (1.80)	0.438 (0.041)				
1992–1996 [n = 338]					–18.00 (3.71)	–14.31 (4.46)	–8.98 (2.22)	0.628 (0.156)
R^2	.461				.399			
Control Observations	2,204		2,204		413		413	

Notes: OLS robust standard errors are in parentheses. For matching estimates, bootstrapped standard errors from 500 replications are in parentheses. Number of treated observations is in brackets. Academic year is August through July. Nonmothers are women who have not had a child as of July 2000. Late mothers are women whose first child was born between August 1996 and July 2000. Treated and control observations were matched based on their predicted birth hazard rate using stratification with 10 groups, as described in the text.

Painter (2003). The base results in this article use stratification matching with 10 groups. Stratification is easy to implement, computationally much faster than kernel and nearest-neighbor matching, and more efficient than the latter. In practice, the results are not much affected by the choice of matching method. Kernel estimators with a wide bandwidth and a Gaussian kernel produce larger estimated credit differences than the base case, but this is likely due to the bias caused by using

controls with more dissimilar birth hazards. Nearest-neighbor matching should be the least biased method because the matches are better, on average, but the estimates are less efficient. The total credit difference between mothers and matched nonmothers is very similar using stratification, nearest-neighbor, or kernel (with Gaussian kernel and small bandwidth or with Epanechnikov kernel) matching. The estimated fraction of the credit difference arising pre-birth is also very similar between the different matching estimators, ranging from 44% to 52%.

Another advantage of the longitudinal approach is that the comparability of mothers and similar-risk nonmothers can be directly tested in the periods before childbirth. Intuitively, the pre-birth course-taking and participation pattern of mothers and comparable-risk nonmothers should be similar. If this condition holds, the plots in Fig. 5 should lie on the zero horizontal axis in the years preceding childbirth. From the spring of 1994 onward, mothers and nonmothers do have similar participation rates immediately after high school. All the deviations are negligible in the first semester after high school, but they increase dramatically shortly thereafter. This indicates that characteristics that are observable at the end of high school—such as family background, educational aspirations, and fertility expectations—are useful predictors of immediate postsecondary education behavior, but this predictive power erodes quickly.

To quantify the average magnitude and timing of this dip across all cohorts of mothers, I estimate a pooled model that combines data from all cohorts. This formulation is an application of the event-study approach utilized by Jacobson et al. (1993) to estimate the earnings loss experienced by displaced workers. This method can be thought of as a generalized difference-in-difference, where a treatment-control difference is calculated at all points before and after treatment and averaged across treatments occurring at different calendar times. I estimate Eq. 10 using linear regression:

$$\begin{aligned}
 y_{i,t} = & \gamma_{j,t} \times (\text{Group}_i = j, \text{Time} = t) \\
 & + (\text{early}_i) \times \sum_{k=-15}^{k=15} (\delta_k^{\text{early}} \times D_{i,t}^k) \\
 & + (\text{late}_i) \times \sum_{k=-15}^{k=15} (\delta_k^{\text{late}} \times D_{i,t}^k) + \epsilon_{i,t} .
 \end{aligned}
 \tag{10}$$

The first component of the right-hand side estimates the counterfactual educational investment pattern for mothers using nonmothers with a similar fertility risk at the end of high school.¹⁸ The last two terms calculate the average deviation from this estimated counterfactual for the 15 periods before and after childbirth,

¹⁸ This equation could be estimated without matching by making assumptions on the functional form of the counterfactual. Replacing the first term in Eq. 10 with γ_t assumes that, absent childbirth, all individuals would have the same postsecondary investment pattern, an assumption refuted by the descriptive longitudinal evidence. Mothers who give birth early begin postsecondary schooling on a different trajectory than mothers who delay parenthood. Replacing the first term in Eq. 10 with $\gamma_t \times \beta \mathbf{X}_i$, where \mathbf{X}_i is a vector of individual and family characteristics, assumes that heterogeneity shifts the intercept of the time profile by a fixed amount that is proportional to individual and background characteristics. This specification has the undesirable property that predicted credits or participation may be negative. Replacing the first term with $\gamma_t \times \beta \mathbf{X}_i$ allows individual characteristics to scale the average profile by a fixed amount. This specification restricts all individuals to have the identical profile shape, other than a proportional scale factor. Matching permits individual characteristics to affect the counterfactual profile in a much more general way.

separately for births that occur early and those that occur late. The indicator variables $early_i$ and $late_i$ denote a first childbirth occurring before or after August 1, 1996, by which time a large number of students have obtained bachelor's degrees. $D_{i,t}^k$ is an indicator for time relative to childbirth. $D_{i,t}^k$ equals 1 if individual i had her first child in period $t - k$ (where k can be positive or negative), and 0 otherwise. $\varepsilon_{i,t}$ is an individual- and time-specific error component. Coefficients on the birth time indicators, δ_k , measure the credit or participation differences between actual and predicted behavior in the current period, having had a child k periods earlier (if $k > 0$). These are the parameters of interest. For instance, δ_0 is the credit or participation drop for mothers during the period of childbirth, relative to not having had a child. Because the deviation from predicted behavior around the time of childbirth is much smaller for later births than for earlier ones, I permit δ_k to vary with birth timing as well. Conceptually, this procedure realigns the graphs in Fig. 5 around a common vertical line at the time of childbirth to create a common "event time" relative to childbirth. The coefficients δ_k are the average deviation of each plot from the zero horizontal axis for each event-time period. No pre-pregnancy divergence would be indicated by $\delta_k = 0$ for all $k < 0$.

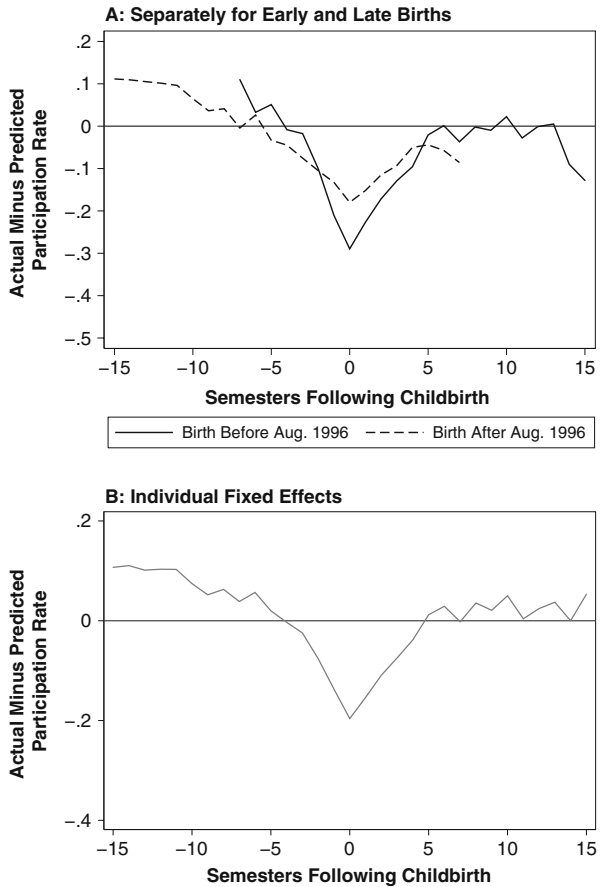
Panel A in Fig. 7 plots the estimated coefficients δ_k from Eq. 10.¹⁹ Childbirth is predicated by a sustained decline in participation for at least three years and is followed by a gradual recovery, albeit to a lower absolute level. By the time of childbirth, early mothers have a participation rate that is 30 percentage points lower (approximately half of the base) than comparable nonmothers. The participation deviation at childbirth is approximately 50% larger for early mothers than for mothers who delay parenthood until four years after high school. Panel B restricts $\delta_k^{early} = \delta_k^{late}$ but includes individual fixed effects. In addition to the counterfactual attendance pattern predicted from nonmothers, this specification permits each mother to have a different average participation rate. The basic results are robust to this alternative specification. Although mothers begin postsecondary education at a rate and intensity comparable to nonmothers, investment begins to fall immediately thereafter for early mothers and only a few years thereafter for late mothers.

Conclusion

Like previous studies, I found that women who enter parenthood earlier have much lower levels of postsecondary educational investment over the eight years following high school. This sustained lower level of investment results in much lower rates of college degree attainment. About half of this discrepancy is explained by factors such as background, educational aspirations, fertility expectations, and sexual activity that is observed or stated at the end of high school. Although these factors predict postsecondary behavior immediately after high school reasonably well, eventual mothers deviate sharply from this prediction shortly thereafter, even before

¹⁹ Because the educational data are restricted to August 1992 through July 2000, coefficients cannot be estimated for more than seven semesters before childbirth for early mothers or more than seven semesters after childbirth for late mothers. The plots in Panel A are therefore truncated at these points.

Fig. 7 Average deviation from predicted postsecondary participation rate. Because the educational data are restricted to August 1992 through July 2000, coefficients cannot be estimated more than seven semesters before childbirth for early mothers or more than seven semesters after childbirth for late mothers. The plots in Panel A are therefore truncated at these points



childbirth. The implication is that studies that control only for fixed individual effects or differences in predetermined observable characteristics do not address a major time-varying source of omitted variable bias. This bias can account for nearly half (or more) of the effect estimated using matching with cross-sectional data.

Eventual mothers reduce educational investment well before the actual occurrence of parenthood, primarily through nonparticipation rather than lower intensity. This suggests that parenthood *per se* is not the causal explanation for the strong link between fertility timing and educational outcomes, at least for women who give birth after high school. The presence of time-varying factors that compel women to halt postsecondary education and then eventually enter parenthood provides a better description of the data. Deliberate postponement of childbirth until after the completion of college is one such explanation. This finding is consistent with the research of Upchurch et al. (2002), who found strong evidence that women purposely sequence childbirth and school attendance. A dynamic model of the simultaneous childbirth-education sequencing decision seems to be a necessary framework for exploring the causal effect of fertility timing on educational attainment.

This article leaves several areas for future research. High school dropout is generally of greater social concern than college dropout, which is why the literature on fertility timing and schooling outcomes has focused on the effects of pregnancy during high school. The present approach could be applied to high school pregnancy and dropout using grade progression, credit accumulation, and course grades as indicators of educational investment, analogous to my use of college attendance. Such an analysis would be more directly comparable to previous studies and would also illuminate any differences in childbirth effects between high school and college educational decisions.

Within the context of postsecondary schooling, it would be fruitful to examine the temporal relationship between women's labor supply, educational investment, and timing of first birth. Childbirth-induced adjustments in work hours could not be accounted for in the preceding analysis. There also appear to be differences in the pre-birth rate of decline in education investment by fertility timing. Understanding the sources of these differences may further illuminate why women choose to discontinue postsecondary education and how this choice is related to fertility decisions. Other life events—notably marriage or cohabitation—are also absent from this analysis. These events likely change the risk of pregnancy during an individual's lifetime. A time-dependent hazard model that relaxes the constant proportional hazard assumption by incorporating these life events would probably predict pre-birth behavior better than the static one I employed. I leave these tasks for future work.

This article also demonstrates the use of the estimated hazard ratio as a balancing score when the causal effect of interest concerns the timing of treatment. The dimension-reduction properties of the propensity score can be directly applied in this case because the hazard ratio is a sufficient statistic for the full vector of treatment probabilities. This approach may prove useful in other applications in which treatment has an important temporal dimension, such as the timing of health care, job services, or school dropout.

Acknowledgments Financial support for this work was provided by the National Institute for Child Health and Human Development (Interdisciplinary Training Grant No. T32-HD007275). I am especially grateful to Ron Lee for extensive guidance on this paper. I also thank Mia Bird, David Card, Ken Chay, Avi Ebenstein, Jane Herr, David S. Lee, Robert D. Mare, Rachel Polimeni, Marit Rehavi, Lawrence Wu, and participants at the UC Berkeley Demography Department Brown Bag and the 2006 annual meeting of the Population Association of America for numerous useful suggestions. All errors are, of course, my own.

References

- Abadie, A., & Imbens, G. (2008). On the failure of the bootstrap for matching estimators. *Econometrica*, *76*, 1537–1557.
- Abrahamse, A. F., Morrison, P. A., & Waite, L. J. (1988). *Beyond stereotypes: Who becomes a single teenage mother?* Santa Monica, CA: The Rand Corporation.
- Angrist, J., & Evans, W. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, *88*, 450–477.
- Angrist, J., & Evans, W. (1999). Schooling and labor-market consequences of the 1970 state abortion reforms. In S. Polachek and J. Robst (Eds.), *Research in Labor Economics* (Vol. 18, pp. 75–114).

- Ashenfelter, O., & Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *Review of Economics and Statistics*, 67, 648–660.
- Ashcraft, A., & Lang, K. (2006). *The consequences of teenage childbearing* (NBER Working Paper No. 12485). Cambridge, MA: National Bureau for Economic Research.
- Bronars, S., & Grogger, J. (1994). The economic consequences of unwed motherhood: using twin births as a natural experiment. *American Economic Review*, 84, 1141–1156.
- Busso, M., DiNardo, J., & McCrary, J. (2009). *New evidence on the finite sample properties of propensity score matching and reweighting estimators*. Unpublished manuscript, Ford School of Public Policy, University of Michigan, Ann Arbor.
- Caliendo, M., & Kopeinig, S. (2005). *Some practical guidance for the implementation of propensity score matching* (IZA Discussion Paper No. 1588). Bonn, Germany: Institute for the Study of Labor.
- Card, D., & Sullivan, D. (1988). Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica*, 56, 497–530.
- Chevalier, A., & Viitanen, T. (2003). The long-run labour market consequences of teenage motherhood in Britain. *Journal of Population Economics*, 16, 323–343.
- Cristia, J. (2008). The effect of a first child on female labor supply: Evidence from women seeking fertility services. *Journal of Human Resources*, 43, 487–510.
- Dolton, P., Smith, J., & Azevedo, J. (2008). The impact of the UK new deal for lone parents on benefit receipt. Unpublished manuscript, Department of Economics, University of Michigan, Ann Arbor.
- Fletcher, J. M., & Wolfe, B. L. (2008). *Education and labor market consequences of teenage childbearing: Evidence using the timing of pregnancy outcomes and community fixed effects* (NBER Working Paper No. 13847). Cambridge, MA: National Bureau of Economic Research.
- Fredriksson, P., & Johansson, P. (2008). Dynamic treatment assignment: the consequences for evaluations using observational data. *Journal of Business and Economic Statistics*, 26, 435–445.
- Frölich, M. (2004). Finite-sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics*, 86, 77–90.
- Geronimus, A., & Korenman, S. (1992). The socioeconomic consequences of teen childbearing reconsidered. *Quarterly Journal of Economics*, 107, 1187–1214.
- Heckman, J., & Hotz, J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association*, 84, 862–874.
- Heckman, J., & Smith, J. (1999). The pre-programme earnings dip and the determinants of participation in a social programme. Implications for simple programme evaluation strategies. *The Economic Journal*, 109, 313–348.
- Hofferth, S. (1987). The social and economic consequences of teenage childbearing. In C. Hayes & S. Hofferth (Eds.), *Risking the future: Adolescent sexuality, pregnancy, and childbearing* (Vol. II, pp. 123–144). Washington, DC: National Academy Press.
- Hotz, J., McElroy, S., & Sanders, S. (2005). Teenage childbearing and its life cycle consequences: exploiting a natural experiment. *Journal of Human Resources*, 40, 683–715.
- Imbens, G. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87, 706–710.
- Jacobson, L., LaLonde, R., & Sullivan, D. (1993). Earnings losses of displaced workers. *American Economic Review*, 84, 685–709.
- Katz, L., & Autor, D. (1999). Changes in the wage structure and earnings inequality. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3A, pp. 1463–1555). New York: Elsevier.
- Kleplinger, D., Lundberg, S., & Plotnick, R. (1999). How does adolescent fertility affect the human capital and wages of young women? *Journal of Human Resources*, 34, 421–448.
- Lechner, M. (1999). Earnings and employment effects of continuous off-the-job training in East Germany after unification. *Journal of Business and Economic Statistics*, 17, 74–90.
- Lechner, M. (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In M. Lechner & F. Pfeiffer (Eds.), *Econometric evaluations of active labour market policies in Europe* (pp. 43–58). Heidelberg, Germany: Physica-Verlag.
- Lechner, M. (2002). Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics*, 84, 205–220.
- Lee, D. (2010). The early socioeconomic effects of teenage childbearing: A propensity score matching approach. *Demographic Research*, 23, 697–736. doi:10.4054/DemRes.2010.23.25
- Levine, D., & Painter, G. (2003). The schooling costs of teenage out-of-wedlock childbearing: Analysis with a within-school propensity-score-matching estimator. *Review of Economics and Statistics*, 85, 884–900.

- Moffitt, R. (2005). Remarks on the analysis of causal relationships in population research. *Demography*, *42*, 91–108.
- Olsen, R., & Farkas, G. (1989). Endogenous covariates in duration models and the effect of adolescent childbirth on schooling. *Journal of Human Resources*, *24*, 39–53.
- Ribar, D. (1994). Teenage fertility and high school completion. *Review of Economics and Statistics*, *76*, 413–424.
- Ribar, D. (1999). The socioeconomic consequences of young women's childbearing: Reconciling disparate evidence. *Journal of Population Economics*, *12*, 547–565.
- Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–50.
- Rosenbaum, P., & Rubin, D. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*, 516–524.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.
- Sanders, S., Smith, J., & Zhang, Y. (2007). Teenage childbearing and maternal schooling outcomes: Evidence from matching. Unpublished manuscript, Department of Economics, University of Maryland, College Park.
- Sianesi, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics*, *86*, 133–155.
- Upchurch, D. M., Lillard, L. A., & Panis, C. W. A. (2002). Nonmarital childbearing: Influences of education, marriage, and fertility. *Demography*, *39*, 311–329.