INSIGHTS FROM A SEQUENTIAL HAZARD MODEL OF SEXUAL INITIATION AND PREMARITAL FIRST BIRTHS

Lawrence L. Wu
Department of Sociology
New York University

Steven P. Martin
Department of Sociology
University of Maryland, College Park

April 2008

This research was funded from NICHD (HD 29550), which is gratefully acknowledged. Direct all correspondence to Lawrence L. Wu, Department of Sociology, Puck Building, 295 Lafayette Street, 4th floor, New York University, New York, NY 10012-9605, lawrence.wu@nyu.edu.

ABSTRACT

Demographers have long acknowledged that sexual activity is a key proximate determinant of fertility. Implicit in such accounts is the presumption of fertility within marriage, yet sexual activity is plausibly of even greater relevance for nonmarital births. In this paper, we model premarital first birth risks in terms of a woman's sequential risks of entry into sexual activity and her risk of a premarital first birth *conditional* on entry into sexual activity. We note that: (1) never-married women have an identifiable period during which their premarital first birth risks are negligible—the period prior to the initiation of sexual activity; (2) never-married women will vary considerably in their ages at onset of sexual activity; (3) and age at onset of sexual activity will vary systematically with observed factors, with these factors also typically influencing the risk of a premarital birth. We exploit this rich empirical structure using techniques developed by Wu and Martin (2008), decomposing the effects of covariates into direct and indirect components. Our empirical results suggest that the direct effects of covariates typically outweigh indirect effects. Exceptions to this pattern provide additional insight into premarital first births.

Since Davis and Moore (1956), demographers have acknowledged that sexual activity is a key proximate determinant of fertility. This insight has assumed even greater importance in the context of current U.S. fertility, in which a substantial proportion of births occur outside of formal marriage. That is, the sexual activity of unmarried women is likely to exhibit even greater variability than that for married women, with nonmarital sexual activity varying not only across unmarried women, but over time for a given unmarried woman. These sources of variation have received scant attention in existing empirical studies of nonmarital fertility, despite the considerable body of empirical research on nonmarital fertility.

In this paper, we employ a sequential hazard model of a woman's entry into sexual activity and her subsequent risk of a premarital first birth conditional on entry into sexual activity. We argue that this approach provides a more realistic model of the premarital first birth process by acknowledging that: (1) never-married women have an identifiable period during which premarital birth risks are negligible, i.e., the period prior to initiation of sexual activity; (2) never-married women will vary considerably in their age at first sexual intercourse; (3) the age at onset of sexual activity for never-married women will vary systematically with observed factors, with these factors also potentially influencing a woman's subsequent risk of a premarital birth; and (4) a woman's risk of a premarital first birth conditional on onset of sexual activity will be influenced by covariates and vary with both age and duration since onset.

In formal derivations below, we show that positing a sequential structure for women's onset of sexual activity and her ensuing premarital first birth risks implies both direct and indirect effects of covariates on these risks. That is, an effect of a covariate may affect premarital first birth risks by, for example, hastening age at onset of sexual activity, thus increasing the duration of exposure to risk, or by affecting premarital birth risks after onset. The first is analogous to an indirect effect in a structural equation setting, while the second is analogous to a direct effect.

It is important to note that the presence of both direct and indirect effects—a seemingly methodological observation—carries substantive implications for policy inasmuch as much current U.S. policy with respect to teen and nonmarital fertility focuses on issues such as sexual abstinence. In particular, our framework allows us to decompose the probability of a premarital first birth into components corresponding to: (a) indirect effects stemming from covariate-induced variation in women's age at onset of sexual activity, and (b) direct effects corresponding to covariate-induced variation in premarital first birth risks in the period following onset of sexual activity. Our empirical results suggest that the direct effects outweigh indirect effects for most of the covariate we examine. Exceptions to this pattern provide additional insights into the processes underlying premarital first births.

THEORY

Let T_1 and T_2 denote random variables for a woman's age at first sexual intercourse and a premarital first birth, respectively, with t_1 denoting the realized (observed) age at first intercourse for women who have initiated sexual activity, and $u = t - t_1$ denoting the duration since initiation of sexual activity conditional on initiation of sexual activity. Then under a standard proportional hazard specification, we consider

$$r_1(t|\mathbf{x}) = q_1(t)\exp(\mathbf{ax})\,, \tag{1}$$

and

$$r_2(t,u|\mathbf{x},t_1) = q_{21}(t|t_1)\,q_{22}(u)\exp(\mathbf{bx})\,, \tag{2}$$

with $r_1(t|\mathbf{x})$ denoting the age-graded risk of first sexual intercourse, $r_2(t,u|\mathbf{x},t_1)$ denoting the age- and duration-graded risk of a premarital first birth conditional on entry into sexual activity, $q_1(t)$ denoting the baseline hazard function for first sexual intercourse, and $q_{21}(t|t_1)$ and $q_{22}(u)$ denoting the baseline hazard functions for age and duration, respectively, for a premarital first birth. Without loss of generality, we assume that **x** is a vector of (exogenous) covariates that is the same in (1) and

(2). The corresponding left-truncated survivor function is given by:

$$S_2(t|\mathbf{x}, t_1) = \exp\Big[-\int_0^u \int_{t_1}^t r_2(s, v|\mathbf{x})\, ds\, dv\Big]\,, \tag{3}$$

with the staggered entry of women into risk reflected by the lower limit of integration t_1 in (3).

In addition, when modeling T_2, one can condition on any relevant aspect of an individual's history (Aalen 1978), including the timing of the event T_1; consequently, one can treat the observed value t_1 as an ordinary right-hand-side covariate in (1). As a result, we also consider the following modification of (2):

$$r_2(t, u|t_1, \mathbf{x}) = q_{21}(t|t_1)\, q_{22}(u)\, \exp(\alpha t_1 + \mathbf{bx})\,. \tag{4}$$

Estimates of **a** and **b** in (1) and (2) can be obtained using a Cox proportional hazard model; however, direct estimation of $S_2(t|\mathbf{x}, t_1)$ in (3) is most easily obtained using parametric proportional hazard models.

The underlying sequential hazard model, in which the occurrence of a first event is necessary for entry into risk of a second event, can be seen as equivalent to specifying the first event as a time-dependent dummy variable in the second event process. Kalbfleisch and Prentice (1980) discuss such a model in the context of the Stanford Heart Transplant study, in which a heart transplant not only affects subsequent mortality risks, but also modifies the effects of other covariates on the risk of mortality. The underlying model presented in this paper can thus be viewed as a special case of such a model, in which risks of the second event are identically zero prior to the occurrence of the first event. Similarly, Stolzenberg (1979) provides an elegant extension of the classical framework of Duncan (1975) to discrete outcomes, including calculation of direct and indirect effects of covariates (see also Winship and Mare 1984).

The expressions in (1) and (2) also carry implications for the interpretation of **a** and **b** in (1) and (4). A first implication is that because the timing of T_1 determines entry into the risk of the second event, prior to the occurrence of T_1, individuals are, by assumption, not at risk of T_2; hence, there can be no direct effect of **x** on the T_2 process prior to the occurrence of T_1. A second is that

when the occurrence of a first event determines entry into the risk of a second event, then variation in the timing of the first event will affect the prevalence of the second event within a population even if the population is otherwise homogeneous.[1] To see this informally, consider a population with two homogeneous subgroups, A and B,m differing only in the timing of T_1. If T_1 occurs later in group A than in B, then the second event will occur less frequently in group A than in B because members in group A spend less time exposed to the risk of the T_2 event than members in group B. Because groups A and B are identical in all respects save for the timing of T_1, differences in the prevalence of the T_3 will be generated only from differences in exposure.

The Transition to the T_1 Event

To formalize ideas, we return to (1) and note that two fundamental quantities related to (1) are the cumulative risk function $H_1(t|\mathbf{x})$ and the survivor function $S_1(t|\mathbf{x})$ given by:

$$H_1(t|\mathbf{x}) = \int_0^t r_1(u|\mathbf{x})du = \exp(\mathbf{ax})Q_1(t) \tag{5}$$

where

$$Q_1(t) = \int_0^t q_1(s)ds\,, \tag{6}$$

and

$$S_1(t|\mathbf{x}) = \Pr(T_1 > t|\mathbf{x}) = \exp[-H_1(t|\mathbf{x})] = \exp[-\exp(\mathbf{ax})Q_1(t)]\,. \tag{7}$$

Note that unlike the case of a linear regression, where a homogeneous subgroup will have a single predicted value of the outcome Y, under (7), a homogeneous subgroup will have a predicted distribution for the event times T_1 given by $S_1(t|\mathbf{x})$. Because of this, because some persons may be censored, and because the distribution of T_1 may be defective, it is more natural to compare

[1] We use the term "prevalence" to refer to $1 - S_j(t|\mathbf{x})$, i.e., the probability that individual i (or group of individuals with characteristics **x**) is in a given demographic state j at time t, with transitions among the mutually exclusive and exhaustive states $j = 1, \ldots, J$ constituting the demographic events of interest. If state j is one of several competing risks, then the interpretation of $1 - S_j$ involves a counterfactual in which all other competing risks are eliminated (Cox and Oakes 1984).

percentiles of T_1 across groups than to compare expectations of T_1 across groups. Then let t_{1p} denote the pth percentile of the distribution of T_1 and suppose again that A and B are two groups of substantive interest, with covariate means given by $\overline{\mathbf{x}}_A$ and $\overline{\mathbf{x}}_B$, respectively. Since $S_1(t)$ varies between 0 and 1, to use (7) to obtain predictions for the timing of T_1 for groups A and B, evaluated at the pth percentile of the T_1 distribution, let $\pi = 1 - (p/100)$; then from (5), we have:

$$t_{1p} = Q_1^{-1}[-\log(\pi)/\exp(\mathbf{ax})]\,, \tag{8}$$

where $Q_1^{-1}(v)$ is the function such that if $v = Q_1(t)$ then $t = Q_1^{-1}(v)$. Given the above, note that compositional differences between groups A and B will generate differences in the pth percentiles of T_1 corresponding to:

$$\begin{aligned}\Delta t_{1p} &= t_{1p}^B - t_{1p}^A \\ &= Q_1^{-1}[-\log(\pi)/\exp(\mathbf{a}\overline{\mathbf{x}}_B)] - Q_1^{-1}[-\log(\pi)/\exp(\mathbf{a}\overline{\mathbf{x}}_A)]\,,\end{aligned} \tag{9}$$

where t_{1p}^A and t_{1p}^B denote the pth percentiles of T_1 in groups A and B, respectively. Note that obtaining t_{1p}^A, t_{1p}^B, and Δt_{1p} requires inverting the function for the integrated hazard $Q(t)$ either analytically or numerically. As noted above, analytic expressions are available for choices of $q(t)$ such as the exponential, Weibull, and Gompertz models and piecewise variants of these models. However, the Cox proportional hazard model (1972) is more difficult to use in this context because it does not specify a parametric form for $q(t)$; in particular, standard proposals for estimating the integrated hazard under a Cox model (see, e.g., Breslow 1974) will not, in general, yield a well-defined inverse function for $Q_1^{-1}(x)$.

The Transition from the T_1 to T_2 Event

As noted above, one way in which T_1 influences the T_2 process is that individuals are not at risk of T_2 until the occurrence of T_1. In addition, when modeling T_2, one can condition on any relevant aspect of an individual's history (Aalen 1978; Tuma and Hannan 1984), including the timing of the event T_1.

To simplify the exposition of ideas, we first focus attention on the case in which $q_{22}(u) = 1$ in (4), with the survivor function $S_2(t|t_1, \mathbf{x})$ then given by:

$$\begin{aligned} S_2(t|t_1, \mathbf{x}) &= \exp\big[-H_2(t|t_1, \mathbf{x})\big] \\ &= \exp\Big[-\int_{t_1}^{t} r_2(s|t_1, \mathbf{x})\, ds\Big] \\ &= \exp\Big[-\exp(\alpha t_1 + \mathbf{bx}) \int_{t_1}^{t} q_{21}(s)\, ds\Big]\,. \end{aligned} \tag{10}$$

Note that the quantity t_1 appears in (10) both as a right-hand-side covariate in the expression $\exp(\alpha t_1)$ and by left-truncating the period of risk via the lower limit of integration.

Now suppose that groups A and B are identical in all respects except that x_1 for groups A and B differs by a constant, i.e., $x_{1B} = x_1 + \Delta$. Consider the cumulative relative risk defined as the ratio of the cumulative hazard for group B to that for group A:

$$\begin{aligned} \frac{H_2(t|t_1, \mathbf{x}_B)}{H_2(t|t_1, \mathbf{x}_A)} &= \frac{\displaystyle\int_{t_1}^{t} q_{21}(s) \exp[\alpha t_1 + b_1(x_{1i} + \Delta) + \cdots]\, ds}{\displaystyle\int_{t_1}^{t} q_{21}(s) \exp[\alpha t_1 + b_1 x_{1i} + \cdots]\, ds} \\ &= \frac{\exp[\alpha t_1 + b_1(x_{1i} + \Delta) + \cdots] \displaystyle\int_{t_1}^{t} q_{21}(s)\, ds}{\exp[\alpha t_1 + b_1 x_{1i} + \cdots] \displaystyle\int_{t_1}^{t} q_{21}(s)\, ds} \\ &= \exp(b_1 \Delta)\,. \end{aligned} \tag{11}$$

Thus, the direct effect on the cumulative relative risk of a shift from x_1 to $x_1 + \Delta$ is given by the usual estimate of relative risk.

How is (11) related to prevalence? Substantively, one might be interested in two quantities related to prevalence, one involving *absolute* prevalence—the arithmetic difference in prevalence between groups A and B—and the other involving *relative* prevalence—the ratio of prevalence for the two groups. In this paper, we focus on arithmetic differences in prevalence. Then recalling that the expression in (6) relates $H(t)$ to $1 - S(t)$ and under the assumptions outlined above, the

arithmetic difference in T_2 prevalence is given by:

$$\begin{aligned}[1 - S_2(t|t_1, \mathbf{x}_B)] - [1 - S_2(t|t_1, \mathbf{x}_A)] &= S_2(t|t_1, \mathbf{x}_A) - S_2(t|t_1, \mathbf{x}_B) \\ &= \exp[-H_2(t|t_1, \mathbf{x}_A)] - \exp[-H_2(t|t_1, \mathbf{x}_B)] \\ &= \exp[-H_2(t|t_1, \mathbf{x}_A)] - \exp[-\exp(b_1\Delta)H_2(t|t_1, \mathbf{x}_A)]\,,\end{aligned} \tag{12}$$

Effect of the timing of T_1 on T_2 prevalence

Variation in T_1 will affect the prevalence of T_2 even in otherwise homogeneous populations because some individuals will have longer durations of exposure to the risk of T_2 by virtue of quicker T_1 transitions. Standard hazard regressions *adjust* for such variations in exposure in the hazard rate, but do not quantify the magnitude of the effect of exposure on prevalence. However, such exposure effects of T_1 on T_2 can be derived via the same ideas as used above. Consider two groups of individuals, A and B, who are identical in all respects save for the timing T_1, and suppose that the random variable T_{1A} is realized as t_1 for group A and that T_{1B} is realized as $t_1 + \Delta$ for group B. Then the cumulative relative risk is given by:

$$\begin{aligned}\frac{H_2(t|t_{1B}, \mathbf{x})}{H_2(t|t_{1A}, \mathbf{x})} &= \frac{\displaystyle\int_{t_1+\Delta}^{t} q_{21}(s)\exp[\alpha(t_1 + \Delta) + \mathbf{bx}]\,ds}{\displaystyle\int_{t_1}^{t} q_{21}(s)\exp[\alpha t_1 + \mathbf{bx}]\,ds} \\ &= \frac{\exp[\alpha(t_1 + \Delta) + \mathbf{bx}]\displaystyle\int_{t_1+\Delta}^{t} q_{21}(s)ds}{\exp(\alpha t_1 + \mathbf{bx})\displaystyle\int_{t_1}^{t} q_{21}(s)ds} \\ &= \exp(\alpha\Delta)\left[\int_{t_1+\Delta}^{t} q_{21}(s)ds \Big/ \int_{t_1}^{t} q_{21}(s)ds\right].\end{aligned} \tag{13}$$

To motivate the notion of an exposure effect, we posited a population at risk of T_2 that was homogeneous in all ways save for the timing of T_1. Such a homogeneous population would imply $\alpha = 0$, leaving only the bracketed ratio of integrals in (13). Thus, the "pure" effect of exposure generated by a shift from t_1 to $t_1 + \Delta$ is given by the bracketed ratio of integrals in (13).

By contrast, $\alpha \neq 0$ suggests that the observed realization t_1 of the random variable T_1 has an effect on T_2 as a usual right-hand-side covariate in the T_2 equation. Such a situation could arise if T_1 has a causal effect on T_2 conditional on the other covariates in the model or if the realization t_1 of the random variable T_1 was correlated with unobserved covariates that influence T_2.

The expression in (13) decomposes the cumulative relative risk into two multiplicative components corresponding to the exposure effect given by the bracketed ratio of integrals and a more "standard" proportional effect of T_1 on T_2 given by $\exp(\alpha\Delta)$. Note that while the "standard" effect does not vary with t by assumption, the effect of exposure will in general vary in nonlinear ways with t. As a result, it can be useful to evaluate the effect of exposure over a range of t.

The arithmetic difference in T_2 prevalence corresponding to (13) is given by:

$$\begin{aligned}[1 - S_2(t|t_1, \mathbf{x}_B)] - [1 - S_2(t|t_1, \mathbf{x}_A)] &= (1 - \exp[-H_2(t|t_{1B}, \mathbf{x})]) - (1 - \exp[-H_2(t|t_{1A}, \mathbf{x})]) \\ &= \exp\Big[-\exp(\alpha t_1 + \mathbf{bx}) \int_{t_1}^{t} q_{21}(s)ds\Big] - \\ &\quad \exp\Big[-\exp(\alpha[t_1 + \Delta] + \mathbf{bx}) \int_{t_1+\Delta}^{t} q_{21}(s)ds\Big] .\end{aligned} \tag{14}$$

Indirect effect of $\mathbf{x}$ *on* T_2 *prevalence*

Assessing the indirect effect of $\mathbf{x}$ on T_2 proceeds in the same way, via an indirect effect of exposure and a more "standard" indirect effect. A first step is to trace the effect of x on the timing of T_1. Consider the pool of individuals who have not yet experienced the event T_1 and suppose that two groups, A and B, are identical in all respects save for their values of x_1. As before, set $x_{1B} = x_{1A} + \Delta$; then from (8), the effect of composition on the timing of T_1 is given by:

$$\begin{aligned}\Delta t_{1p} &= t_{1Bp} - t_{1Ap} \\ &= Q_1^{-1}[-\log(\pi)/\exp(\mathbf{bx}_B)] - Q_1^{-1}[-\log(\pi)/\exp(\mathbf{bx}_A)] \\ &= Q_1^{-1}[-\log(\pi)/\exp(b_1(x_1 + \Delta) + \cdots)] - Q_1^{-1}[-\log(\pi)/\exp(b_1 x_1 + \cdots)] ,\end{aligned} \tag{15}$$

where $\pi = 1 - (p/100)$ and p corresponds to the pth percentile for the distribution of T_1. Note that

the function Q^{-1} is highly nonlinear; hence, the predicted effect of a covariate x on the percentile distribution T_1 will vary with the percentile p at which the effect is evaluated.

Recall from (13) that a shift from t_1 to $t_1 + \Delta$ influences the cumulative relative risk in two ways, through an indirect effect of exposure and a "standard" indirect effect. But a shift from x_1 to $x_1 + \Delta$ will induce a shift in t_1, thus generating both direct and indirect effects for the T_2 equation. This is given by combining (13) and (15), from which one can derive the indirect effect of shifting x_1 to $x_1 + \Delta$ on the cumulative relative risk:

$$\begin{aligned}\frac{H_2(t|t_{1B},\mathbf{x}_B)}{H_2(t|t_{1A},\mathbf{x}_A)} &= \exp(b_1\Delta + \alpha\Delta t_{1p})\Big[\int_{t_{1A\pi}+\Delta t_{1p}}^{t} q_{21}(s)ds \Big/ \int_{t_{1A\pi}}^{t} q_{21}(s)ds\Big] \\ &= \exp(b_1\Delta)\exp(\alpha\Delta t_{1p})\Big[\int_{t_{1A\pi}+\Delta t_{1p}}^{t} q_{21}(s)ds \Big/ \int_{t_{1A\pi}}^{t} q_{21}(s)ds\Big]\end{aligned} \tag{16}$$

Thus, the consequence of shifting from x_1 to $x_1+\Delta$ in the T_1 equation appears in three places in the T_2 equation in (16): a direct effect of x_1 on T_2 represented by the quantity $b_1\Delta$, and two indirect effects of x_1 via T_1—an indirect effect of exposure represented by the lower limit of integration in (16), and a more usual indirect effect represented by the quantity $\alpha\Delta t_{1p}$. When T_2 depends on both age and duration, the cumulative relative risk involves double integrals:

$$\frac{H_2(t|t_{1B},\mathbf{x}_B)}{H_2(t|t_{1A},\mathbf{x}_A)} = \exp(b_1\Delta)\exp(\alpha\Delta t_{1p})\frac{\displaystyle\int_{t_{1A\pi}+\Delta t_{1p}}^{t} q_{21}(s)ds \int_{0}^{u-\Delta t_{1p}} q_{22}(v)dv}{\displaystyle\int_{t_{1A\pi}}^{t} q_{21}(s)ds \int_{0}^{u} q_{22}(v)dv} \tag{17}$$

Thus as in (16), the consequence of shifting from x_1 to $x_1 + \Delta$ in the T_1 equation appears in three places in (17): a direct effect of x_1 on T_2 represented by the quantity $b_1\Delta$, and two indirect effects of x_1 via T_1—an indirect effect of exposure represented by the lower and upper limits of integration, and the more usual indirect effect represented by the quantity $\alpha\Delta t_{1p}$.

The arithmetic difference in T_2 prevalence between groups B and A similarly involves a change to double integrals

$$
\begin{aligned}
S_2(t,u|t_1,\mathbf{x}_B)] - S_2(t,u|t_1,\mathbf{x}_A) =& (1-\exp[-H_2(t,u|t_{1B},\mathbf{x})]) - (1-\exp[-H_2(t,u|t_{1A},\mathbf{x})]) \\
&= \exp\Big[-\exp(\alpha[t_1+\Delta]+\mathbf{bx})\int_{t_1+\Delta}^{t} q_{21}(s)\,ds\int_0^{u-\Delta} q_{22}(v)\,dv\Big] - \\
&\quad \exp\Big[-\exp(\alpha t_1+\mathbf{bx})\int_{t_1}^{t} q_{21}(s)\,ds\int_0^{u} q_{22}(v)\,dv\Big],
\end{aligned}
\tag{18}
$$

To summarize, we have shown that indirect effects of covariates can be decomposed into "standard" and "exposure" components. The "standard" component of the indirect effect arises because the value of a covariate x influences the timing of T_1, with the timing of T_1 influencing the T_2 process via t_1 as a right-hand-side covariate in the T_2 equation. The "exposure" component of the indirect effect comes from differences in exposure in which x influences the timing of T_1, and in which the timing of T_1 influences durations of exposure to the risk of T_2, with these influences appearing in the upper and lower limits of integration in the expressions for cumulative relative risk H and prevalence $1-S$.

DATA

We use data from the 1979 National Longitudinal Survey of Youth (NLSY), a household-based national probability sample of persons aged 14-21 in 1979. The original 12,686 cases consist of a main sample of 6,111 respondents, an oversample of 5,295 minorities and poor whites, and a sample of 1,280 Armed Forces personnel. The military sample was suspended in 1985, with 1,079 (out of the original 1,280) cases affected. Retention has been high in the NLSY, with for example, 10,485 (90.3 percent) Of the 11,607 non-military respondents reinterviewed in the 1987 wave, for a retention rate of 98.8 percent.

Of the 6,283 women present at the initial 1979 interview, we excluded women: (1) of all race and ethnicities other than non-Hispanic whites, non-Hispanic blacks, and Hispanics (n = 875); (2) with missing data on first intercourse (n = 156); (3) who reported not knowing their biological mother (n = 7); (4) with missing data on age at menstruation (n = 93); or (5) with missing first

birth or first marriage histories (n = 138). These selection criteria yielded a sample of $n = 5,014$ women.

Data on age at first sexual intercourse were obtained in the 1984–1986 interviews, when all respondents were at least 18 years old. In the 1984 wave, age at first intercourse was obtained to the nearest year. In the 1985 wave, questions on the calendar month and year of menarche and first sexual intercourse were administered to all female respondents; these questions were repeated in 1986 for 1985 female nonrespondents. We computed the young woman's age in months at first premarital sexual intercourse using data from the 1985 and 1986 waves, using a hot-deck procedure to impute missing data on calendar month at first sexual intercourse. Wu, Martin, and Long (2001) find that these self-reports are of reasonable quality, with comparisons of these data in close agreement with data on sexual onset for a comparable birth cohort of women from the 1995 National Survey of Family.

For women who report never having engaged in sexual activity, we censored their first sexual intercourse history at their age at interview in 1985 or 1986, depending on the year in which they were asked the question. We likewise censored women's first sexual intercourse history at their age at first marriage if they reported that they had initiated sexual intercourse on or after the date of first marriage. We similarly censored a woman's premarital birth history at either her age at last interview or at her age at first marriage if she did not report a first birth prior to last survey observation or first marriage.

RESULTS

Figure 1 presents smoothed nonparametric estimates using a procedure described in Wu (1989) for the age-graded risk of entry into sexual activity, the age-graded risk of a premarital first birth, and the duration-graded risk of a premarital first birth conditional on entry into sexual activity. The top panel of Figure 1 plots smoothed nonparametric estimates of the logarithm of the hazard rate of first sexual intercourse by age, the middle panel plots two different estimates of the logarithm of

the hazard rate for a premarital first birth, and the bottom panel plots estimates of the logarithm of the hazard rate for a premarital first birth by duration since sexual onset. In the upper two panels, the curves for the logarithm of the rate rise in a roughly linear fashion to about age 18.5, after which the curves decline, again in a roughly linear fashion.

[Figure 1 about here]

In the middle panel of Figure 1, the two curves differ in the assumptions they make about when women become at risk of a premarital first birth. The solid curve presents estimates that do not place a woman at risk of a premarital first birth until she reports becoming sexually active; hence, for this curve, we use a woman's report of age at first intercourse to left-truncate her premarital birth history. The dotted curve presents estimates that ignore this left truncation; hence, while this curve can be viewed as the average of the logarithm of premarital first birth risks in the population, it ignores variation in onset of sexual activity and implicitly assumes that women are at risk of a premarital first birth even if they have not initiated sexual activity, an implausible assumption.

A comparison of the two curves in the lower panel of Figure 1 shows that left truncation affects estimates substantially, with the curve ignoring left truncation systematically underestimating premarital first birth risks relative to the curve that incorporates left truncation. Differences between these two curves are especially apparent at younger ages, reflecting the tendency for premarital births risks to be especially high for teen women in the period following the initiation of sexual activity.

The nonparametric estimates in the bottom panel of Figure 1 exhibit a non-monotonic pattern of duration dependence in which premarital first birth risks first rise and then decline. Based on these nonparametric results, we model age dependence in both the T_1 and T_2 equations using a splined piecewise Gompertz specification with nodes at ages 15 and 18 (e.g. Wu and Tuma 1990, Lillard 1993). For the T_2 equation, we modeled duration dependence using a piecewise constant specification for durations 0 to 14, 15 to 29, 30 to 59, and 60+ months. Estimates from these

models are presented in Table 1.

[Table 1 about here]

The first two columns in Table 1 adopt a conventional approach to modeling premarital first birth risks by examining women's age-specific risks of a premarital first birth but ignoring the timing of first sexual intercourse. We present estimates from two proportional hazard specifications, the Cox proportional hazard model and a piecewise splined Gompertz model with proportional effects of covariates. Estimates from these models reveal substantially higher relative risks for blacks compared to whites, but no significant difference in relative risks for white and Hispanic women. The next four columns present corresponding estimates for the transition to first sexual intercourse and the transition to a premarital first birth conditional on entry into sexual activity. Compared to white women, black women have significantly higher risks of first sexual intercourse (corresponding to earlier ages at onset) as well as significantly higher premarital first birth risks following onset. However, the Hispanic/white contrasts are opposite in sign for the two transitions, with significantly *lower* risks of first sexual intercourse but significantly *higher* premarital first birth risks following onset for Hispanic women relative to white women,

The next row presents estimated coefficients for a time-varying dummy variable equal to one at all ages after first menses. A conventional modeling approach suggests that this variable is associated with significantly higher premarital first birth risks; however, estimates from our sequential approach shows that this variable is associated with significantly higher risks of sexual onset but is not significantly associated with premarital first birth risks following onset. The next three rows present estimated coefficients for mother's education, number of siblings, and income-to-needs. All three variables have associations in the expected directions with unconditional premarital first birth risks, small and statistically insignificant associations with age at first intercourse, and associations in the expected directions with premarital first birth risks conditional on sexual onset. Thus, our results suggest that conclusions obtained from our

sequential approach can yield qualitatively different insights than those obtained from a more conventional approach.

Results for family structure, religion, and ability are reported in the next three rows of Table 1. These associations show qualitative agreement between approaches, with the signs and significance levels similar for estimated coefficients of the risks for the unconditional transition to a premarital first birth, to first intercourse, and to a premarital first birth conditional on sexual initiation.

The results in Table 1 also close agreement between estimates the Cox and piecewise splined Gompertz specifications. As noted above, our decomposition derivations require explicit estimates of the various baseline hazards, which are not easily obtained from a Cox specification; hence, we henceforth restrict our discussion to estimated coefficients from the piecewise splined Gompertz models.

We now turn to results for selected decompositions. Table 2 presents decomposition results comparing black and white women. Predicted median ages at onset of sexual activity are reported in Panel A of Table 2 and are calculated using the estimated coefficients in column 2 of Table 2 and the expressions in (8) and (A6) in Appendix 1, with other covariates set to their sample means. The values of the predicted medians are 17.55 and 17.73 (210.6 and 212.7 months) for black and white women, respectively. The resulting difference, while in the expected direction, is thus relatively small, corresponding to the black coefficient (.10) in Table 1. Although observed black/white differences in age at first sexual intercourse are larger, these differences do not control for other variables; thus, our results suggest that much of the unconditional difference in age at onset of sexual activity can be attributed to the association of variables other than race on women's age at onset of sexual activity.

[Table 2 about here]

As noted above, decomposition results will vary with duration of exposure; hence, Panel B presents results for 60 and 90 months. Panel B reports the arithmetic difference in the predicted

percentage of premarital first births, obtained using the expression in (18) and using the estimated coefficients in columns 4 and 6 of Table 1. As expected, there are substantial differences in prevalence, even holding constant other variables, with a predicted black/white difference of 11.7 and 13.2% in the percentage of women having a premarital birth at 60 and 90 months, respectively, following sexual onset.[2] As noted above, the derivations of the previous show that the 11.7 and 13.2 coefficients can be decomposed into three components, a direct component (labeled "D"), corresponding to the estimated coefficients in columns 4 and 6 of Table 1, and two indirect components, one corresponding to the estimated right-hand-side coefficient for age at first intercourse in Table 1 ("E") and a second due to black/white differences in exposure to risk ("F"). Thus at 60 months of exposure, the predicted black/white difference of 11.7% can be decomposed into a direct effect of 10.3% and two indirect effects of .3 and 1.0%, respectively. This shows that the direct effect is substantially larger than either of the two indirect effects, which is in qualitatively agreement with the estimated coefficients in Table 1 (black/white coefficient of .10 for onset of sexual activity and .54 for premarital first birth risks conditional on onset).

Table 3 presents parallel decompositions for Hispanic and white women, with all other covariates set to their sample means. Recall that the parallel Hispanic/white coefficients in Table 1 for the piecewise splined Gompertz model were negative for age at first sexual ($-.47$), but positive and significant for premarital first births conditional on age at onset (.38). The differences in predicted median ages at onset of sexual activity correspond to the $-.47$ coefficient in Table 1, with predicted values of 18.65 and 17.73 (223.8 and 212.7 months) for Hispanic and white women, respectively

[Table 3 about here]

[2]Because a premarital first birth and first marriage are competing risks, our predicted probabilities at 60 and 90 months of duration should be interpreted under the counterfactual in which women cannot marry during these durations of exposure. This counterfactual is substantively most appropriate when the variables examined in our decompositions do not have a strong association with the competing risk of first marriage. These cautions affect possible interpretations of our results, an issue especially important for our black/white decompositions.

Because the Hispanic/white contrasts in Table 1 take opposite signs, the sign of the arithmetic difference in Table 3 could in principle be positive or negative, depending on the magnitude of the Table 1 coefficients. Our empirical results show lower prevalence for Hispanic women relative to their white counterparts (-6.9 and -4.9% at 60 and 90 months of exposure, respectively), conditional on onset of sexual activity and after setting all other covariates to their sample means. The magnitude of these differences is roughly half that of the corresponding black/white differences in the decompositions in Table 2.

It is informative to contrast the above results with the black/white (.64) and Hispanic/white (.02) coefficients in Table 1 for the unconditional transition to a premarital first birth. That is, adopting a conventional approach that examines a woman's age at a first premarital birth but ignores the timing of first sexual intercourse implies large black/white differences but small Hispanic/white differences in the percentage with a premarital first birth, holding other covariates. If, however, premarital first birth risks are assumed to be negligible prior to onset of sexual activity, our results suggest, as before, more premarital first births to blacks relative to whites (about 12 or 13% for 60 and 90 months of exposure), but *fewer* premarital first births to Hispanics relative to whites (between 5 and 7% for 60 and 90 months of exposure). These comparisons show that the results from our sequential model generate insights that are qualitatively different from more conventional approaches.

Table 4 present decomposition results for AFQT. Recall that AFQT had significant effects in Table 1 for both onset of sexual activity ($-.14$) and premarital first birth risks given onset ($-.38$). In these decompositions, we compare women with low and high AFQT scores, defined as a score half a standard deviation below or above the mean, respectively, corresponding to standardized scores of ± 0.7. Panel A of Table 4 shows that varying AFQT in this way corresponds to just under a 3 month difference in the predicted median age at first sexual intercourse (215.1 vs. 213.2 months) because of the relatively modest magnitude of this effect given these scores ($-.14 \times [\pm 0.7] \approx \pm 0.1$). Differences in prevalence, even holding constant other variables, are 8.5 and 9.5% for the

percentage of women having a premarital birth at 60 and 90 months, respectively, following sexual onset. These correspond to direct and indirect effects of 6.7 (direct), 0.3, and 1.4% (indirect) at 60 months following sexual onset and 7.9, 0.4, and 1.3% at 90 months following sexual onset. Thus, these decompositions show substantially smaller indirect effects of AFQT on premarital birth risks, and a far larger direct effect, holding constant all other variables.

[Table 4 about here]

Table 5 present decomposition results for women from intact and nonintact families at age 14. In Table 1, results from the piecewise splined Gompertz specification were that residing in a nonintact family at age 14 was associated with a .34 higher risk of onset of sexual activity and a .26 higher risk of a premarital first birth conditional on onset of sexual activity. Panel A of Table 1 shows that these results yield a predicted difference in the median age at first sexual intercourse of 7 months (209.6 vs. 216.6). The corresponding differences in the probability of a premarital first birth are 15.8 and 17.7% for 60 and 90 months of exposure following sexual onset, respectively, holding all other covariates at their sample means. The decompositions for the 15.8% difference at 60 months of exposure show that the largest portion comes from the direct effect (11.1%), with the next largest portion stemming from the indirect effect of differential exposure (3.8%). The results for 90 months of exposure are similar, with the largest portion of the overall 17.7% difference stemming from the direct effect (13.3%) and far smaller portions from the two indirect effects (3.4% from the indirect exposure effect and 1.0% from the indirect right-hand-side covariate effect of age at onset). Thus, these decomposition results show that direct effects dominate indirect effects even though the relative risks for nonintact family structure in Table 1 are larger for first sexual intercourse than for premarital first births.

[Table 5 about here]

Tables 6–8 present parallel decompositions for mother's education, timing of menses, and income-to-needs in the woman's family of origin. [Paragraphs not written]

[Tables 6–8 about here]

DISCUSSION

In this paper, we have proposed a sequential model for the subprocesses underlying a premarital first birth in which we conceptualize the risk of such a birth in terms of an initial period during which women are at risk of initiating sexual activity and a subsequent during which women are at risk of a premarital first birth following initiation of sexual activity. This sequential model differs from a more conventional hazard specification modeling women's age at a premarital first birth, with our sequential model positing that women become at risk of a premarital first birth only after initiation of sexual activity. Although highly stylized, in that sexual activity will vary in intensity and frequency following first intercourse, this sequential approach nevertheless highlights key periods during which premarital first birth risks can be expected to vary substantially. It also follows a long demographic tradition that holds that better approximating durations of exposure to risk is a central task in understanding demographic phenomena such as fertility.

Does a sequential hazard model provide insights different than more conventional hazard models? For premarital first births, we conclude that the answer is yes. Our empirical results suggest numerous examples in which the effects of covariates on the transition into sexual activity and to a premarital first birth are zero for one transition but substantial and statistically significant for the other transition, or in which coefficients are substantial in magnitude and statistically significant but opposite in sign.

Following Wu and Martin (2008), we also decompose the arithmetic difference in the probability of a premarital birth into direct and indirect components of covariates. That is, black/white differences in the probability of a premarital first birth can reflect, for example, a direct effect reflecting higher premarital first birth risks in the period following sexual initiation for blacks relative to whites. But black/white differences in the probability of a premarital first birth can also result from an indirect effect reflecting, for example, earlier black entry into sexual

activity, which in turn will imply longer durations of exposure to risk for blacks relative to whites.

Our decompositions provide additional insights not easily obtained from a simple inspection of the hazard coefficients in our sequential model. A first insight provided by our decompositions is that direct effects typically outweigh indirect effects. This pattern holds for 6 of the 8 decompositions we present, including the decomposition for family structure, where the coefficient for first sex is larger and attains a higher level of statistically significance than the coefficient for premarital birth conditional on sexual onset. The two exceptions were our decompositions comparing white and Hispanic women and for early and late menses. Our white/Hispanic decompositions indicate a substantially later entry of Hispanic women (12 months) into sexual activity, net of the other covariates in our models. By contrast, our results for age at menarche show that early onset of menses is associated with earlier age at onset of sexual activity, but has no association with premarital first birth risks conditional on sexual onset.

For demographers, this result will not be surprising given that proximate determinants tend to take precedence over more distal determinants. Nevertheless, policies targeting teen and nonmarital fertility have often assumed that delaying sexual activity or encouraging abstinence will produce substantial reductions in these outcomes, with far less attention paid to policies targeting how premarital first birth risks might be reduced in the period following initiation of sexual activity.

A second insight is that among indirect effects, the indirect effect of exposure outweighs the indirect effect. Earlier age at first sex clearly results in longer exposure to the chance of a nonmarital birth. By a rough average, moving first sex earlier by a month increases the proportion of nonmarital births by .5 percent, net of other controls. Earlier age at first sex also increases the rate of premarital births following first sex, but this difference while statistically significant is only about a third as important as the exposure effect.

This second insight also carries potential policy implications for debates about abstinence education. To whatever extent abstinence education is effective, the large majority of that effect is up front; reducing non-marital births due to pregnancies in the first few months after initiation

of sexual activity. The longer-term (covariate) effects of early first sex on women's long term behavior and nonmarital fertility are quite small.

Although our empirical results provide no firm causal estimates of the effects of covariates on either sexual initiation or premarital first birth risks, they nevertheless run counter to arguments that delaying sexual onset will lead to substantial reductions in nonmarital fertility. More generally, our models and empirical results represent a first step toward conceptualizing premarital first births in terms of subprocesses such as the initiation of sexual activity, and the subprocesses following sexual initiation, including contraceptive effort by sexually active women, pregnancy risks following sexual onset conditional on contraceptive effort, and, conditional on a pregnancy, how a pregnancy is resolved.

APPENDIX 1

Obtaining the indirect effect of **x** on T_2 prevalence requires inverting the integral of $q_1(t)$. The examples in this paper employ a piecewise splined Gompertz specification for the baseline hazard $q_1(t)$ of T_1. Under proportionality, we have

$$r_1(t) = q_1(t)\exp(\mathbf{ax})\,. \tag{A1}$$

Consider partitioning the time interval (τ_0, ∞) into K prespecified intervals $(\tau_0, \tau_1]$, $(\tau_1, \tau_2]$, . . ., $(\tau_{K-1}, \infty]$; then a piecewise splined Gompertz specification for $q_1(t)$ can be written as:

$$q_1(t) = \begin{cases} \exp(\beta_1 + \gamma_1 t) & t \in (\tau_0, \tau_1]\,; \\ \exp(\beta_2 + \gamma_2 t) & t \in (\tau_1, \tau_2]\,; \\ \ldots & \\ \exp(\beta_K + \gamma_K t) & t \in (\tau_{K-1}, \infty]\,, \end{cases} \tag{A2}$$

for $\gamma_k \neq 0$ and $k = 1, \ldots, K$. Under (A1) and (A2), $H(t)$ is given by:

$$H_1(t) = \int_{\tau_0}^{t} r_1(s)ds = \exp(\mathbf{ax}) \int_{\tau_0}^{t} q_1(s)ds = \exp(\mathbf{ax})Q_1(t)\,. \tag{A3}$$

Integrating $Q_1(t)$ yields:

$$Q_1(t) = \begin{cases} e^{\beta_1}\left(e^{\gamma_1 t} - e^{\gamma_1 \tau_1}\right)/\gamma_1 & t \in (\tau_0, \tau_1]\,; \\ Q_1(\tau_1) + e^{\beta_2}\left(e^{\gamma_2 t} - e^{\gamma_2 \tau_2}\right)/\gamma_2 & t \in (\tau_1, \tau_2]\,; \\ \ldots & \\ Q_1(\tau_{K-1}) + e^{\beta_K}\left(e^{\gamma_K t} - e^{\gamma_K \tau_K}\right)/\gamma_K & t \in (\tau_{K-1}, \infty]\,. \end{cases} \tag{A4}$$

As noted in the text, our goal is to determine the p percentile of the T_1 distribution; this corresponds to inverting the integral of $q_1(t)$. Set $Q_1(t) = x$ and define the inverse function implicitly through $Q_1^{-1}(x) = t$. Suppose the desired percentile lies in the kth interval $(\tau_{k-1}, \tau_k]$; then from (A4)

$$\begin{aligned} x &= Q_1(\tau_{k-1}) + [\exp(\beta_k + \gamma_k t) - \exp(\beta_k + \gamma_k \tau_{k-1})]/\gamma_k \\ \exp(\beta_k + \gamma_k t) &= \exp(\beta_k + \gamma_k \tau_{k-1}) + \gamma_k[x - Q_1(\tau_{k-1})] \\ t &= \left[\log\left(\exp(\beta_k + \gamma_k \tau_{k-1}) + \gamma_k[x - Q_1(\tau_{k-1})]\right) - \beta_k\right]/\gamma_k \end{aligned} \tag{A5}$$

Hence for $t \in (\tau_{k-1}, \tau_k]$,

$$Q_1^{-1}(x) = \left[\log\left(\exp(\beta_k + \gamma_k \tau_{k-1}) + \gamma_k [x - Q_1(\tau_{k-1})]\right) - \beta_k\right] / \gamma_k \,. \qquad (A6)$$

Minor complications arise when the distribution of T_1 is defective—that is, when some individuals will not experience the event T_1 even when $t \to \infty$. For the piecewise Gompertz specification, this is determined by parameters in the last open interval, $(\tau_{K-1}, \infty]$. In this interval, the T_1 distribution will be defective if

$$\gamma_K e^{-\beta_K} [x - Q_1(\tau_{K-1})] < e^{\gamma_K \tau_{K-1}} \,. \qquad (A7)$$

Inspecting (A7) shows that $\gamma_K < 0$ is a necessary but not sufficient condition for the distribution of T_1 to be defective.

REFERENCES

Aalen, Odd O. 1978. "Nonparametric Inference for a Family of Counting Processes." *Annals of Statistics* 6(4): 701–26.

Breslow, N. 1974. "Covariance Analysis of Censored Survival Data." *Biometrics* 30(1): 89–99.

Cox, D. R. 1972. "Regression Models and Life Tables" (with discussion). *Journal of the Royal Statistical Society* B34(2): 187–220.

Cox, D. R., and D. Oakes. 1984. *Analysis of Survival Data*. London: Chapman and Hall.

Davis, Kingsley, and Judith Blake. 1956. "Social Structure and Fertility: An Analytical Framework." *Economic Development and Cultural Change* 4(4): 211–35.

Duncan, Otis Dudley. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.

Kalbfleisch, John D., and Ross L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Stolzenberg, Ross M. 1979. "The Measurement And Decomposition Of Causal Effects In Nonlinear And Nonadditive Models." Pp. 459–88 in Karl Schuessler (Ed.), *Sociological Methodology 1980.* San Francisco: Jossey-Bass.

Winship Christopher, and Robert D. Mare. 1984. "Regression Models with Ordinal Variables." *American Sociological Review* 49(4): 512-25.

Wu, Lawrence L. 1989. "Issues in Smoothing Empirical Hazards." Pp. 127–59 in Clifford C. Clogg (Ed.), *Sociological Methodology 1989*. Washington: American Sociological Association.

Wu, Lawrence L., Steven P. Martin, and Daniel A. Long. 2001. "Comparing Data Quality of Fertility and First Sexual Intercourse Histories." *Journal of Human Resources* 36(3): 520–55.

Wu, Lawrence L., and Steven P. Martin. 2008. "Effects of Exposure on Prevalence and Cumulative Relative Risk: Direct and Indirect Effects in a Recursive Hazard Model." Unpublished manuscript, Department of Sociology, New York University.

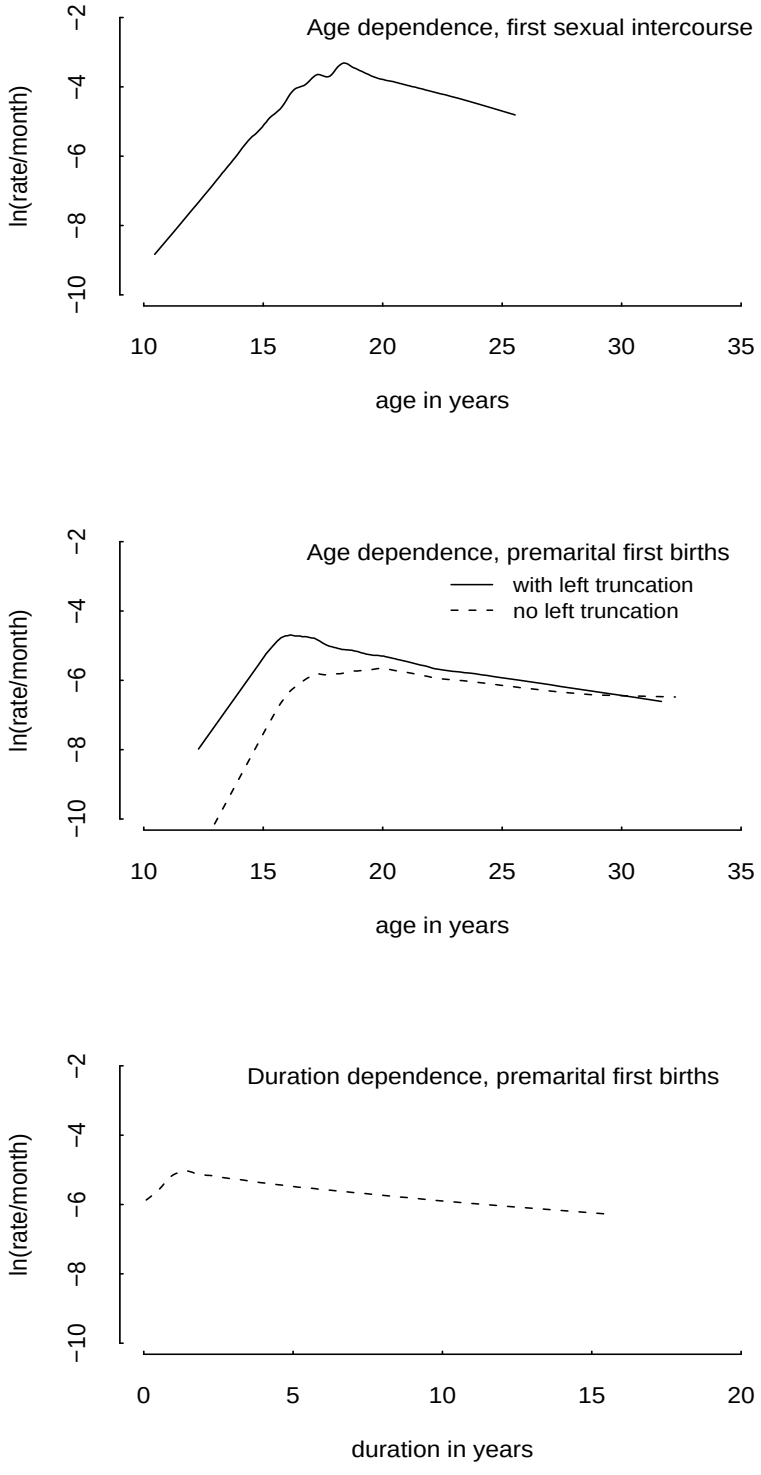

Figure 1: Smoothed nonparametric estimates of: (a) age dependence in the logarithm of the hazard for the transition to first sexual intercourse, (b) age dependence in the transition to a premarital first birth, and (c) duration dependence in the transition from sexual onset to a premarital first birth.

Table 1. Estimated coefficients from Cox and piecewise splined Gompertz proportional hazard models for: (a) the unconditional transition to a premarital first birth; (b) the transition to first sexual intercourse; and (c) the transition from onset of sexual activity to a premarital first birth.

	Unconditional transition to a premarital first birth		Transition to first sexual intercourse		Transition to a premarital first birth given sexual onset	
	Cox	Gmp	Cox	Gmp	Cox	Gmp
Race and ethnicity						
black	.64***	.64***	.10*	.10*	.54***	.54***
	(.08)	(.08)	(.04)	(.04)	(.08)	(.08)
Hispanic	−.02	−.02	−.46***	−.47***	.38**	.38**
	(.12)	(.12)	(.06)	(.06)	(.12)	(.12)
Sexual maturation						
menstruation (time-varying dummy variable)	1.38*	1.32**	.95***	.91***	.49	.47
	(.52)	(.47)	(.12)	(.11)	(.51)	(.47)
Family background						
mother's education	−.03**	−.03**	.00	.00	−.04***	−.04***
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
number of siblings	.04***	.04***	.00	.00	.05***	.05***
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
income-to-needs ratio (in 1000s)	−.055***	−.055***	−.003	−.003	−.052***	−.052***
	(.008)	(.008)	(.003)	(.003)	(.007)	(.007)
mother's age at first birth	−.04***	−.04***	−.03***	−.03***	−.04***	−.04***
	(.01)	(.01)	(.00)	(.00)	(.01)	(.01)
nonintact family at age 14	.50***	.50***	.34***	.34***	.25***	.26***
	(.06)	(.06)	(.04)	(.04)	(.06)	(.06)
catholic	−.12	−.12	−.06	−.07	−.07	−.07
	(.09)	(.09)	(.04)	(.04)	(.09)	(.09)

Table 1. (continued)

	Unconditional transition to a premarital first birth		Transition to first sexual intercourse		Transition to a premarital first birth given sexual onset	
	Cox	Gmp	Cox	Gmp	Cox	Gmp
Ability						
AFQT	$-.42^{***}$	$-.42^{***}$	$-.14^{***}$	$-.14^{***}$	$-.38^{***}$	$-.38^{***}$
	(.04)	(.04)	(.02)	(.02)	(.04)	(.04)
Onset of sexual activity						
age (in months)					$-.38$	$-.67^{**}$
					(.25)	(.21)
missing calendar month, onset of sexual activity					.04	.04
					(.06)	(.06)
Duration dependence						
15 to 29 months					$.26^{**}$	$.21^{*}$
					(.09)	(.09)
30 to 59 months					.12	.01
					(.12)	(.11)
60 months or more					.21	$-.07$
					(.21)	(.18)

All models also include dummy variables for missing values of: mother's education, mother's age at first birth, family structure at age 14, number of siblings, AFQT, and income-to-needs.

$^{*}\ p < .05 \quad ^{**}\ p < .005 \quad ^{***}\ p < .0005$ (two-tailed tests)

Table 2. Decomposition of the difference in the percentage of women predicted to have a premarital first birth, controlling for other covariates. Comparison of non-Hispanic blacks and non-Hispanic whites.

Panel A	age in months
Predicted median, age at first intercourse	
blacks	210.6
whites	212.7

Panel B	months of exposure	
	60	90
Predicted percentage, premarital first birth		
A: blacks	28.3	33.4
B: whites	16.6	20.2
Predicted difference, premarital first birth		
C: A − B	11.7	13.2
Decomposition of C		
D: direct component	10.3	12.0
E: indirect component, differential age at onset	.3	.3
F: indirect component, differential exposure	1.0	0.9

Table 3. Decomposition of the difference in the percentage of women predicted to have a premarital first birth, controlling for other covariates. Comparison of Hispanics and non-Hispanic whites.

Panel A	age in months
Predicted median, age at first intercourse	
Hispanics	223.8
whites	212.7

Panel B	months of exposure	
	60	90
Predicted percentage, premarital first birth		
A: Hispanics	9.7	15.3
B: whites	16.6	20.2
Predicted difference, premarital first birth		
C: A − B	−6.9	−4.9
Decomposition of C		
D: direct component	4.8	6.2
E: indirect component, differential age at onset	−1.1	−1.3
F: indirect component, differential exposure	−10.8	−9.7

Table 4. Decomposition of the difference in the percentage of women predicted to have a premarital first birth, controlling for other covariates. Comparison of women with AFQT scores half a standard deviation above and below the mean.

Panel A	age in months
Predicted median, age at first intercourse	
mean −.5 s.d.	212.2
mean + .5 s.d.	215.1

Panel B	months of exposure	
	60	90
Predicted percentage, premarital first birth		
A: mean −.5 s.d.	24.0	28.7
B: mean + .5 s.d.	15.5	19.2
Predicted difference, premarital first birth		
C: A − B	8.5	9.5
Decomposition of E		
D: direct component	6.7	7.9
E: indirect component, differential age at onset	.3	.4
F: indirect component, differential exposure	1.4	1.3

Table 5. Decomposition of the difference in the percentage of women predicted to have a premarital first birth, controlling for other covariates. Comparison of women from intact and nonintact families at age 14.

Panel A	age in months
Predicted median, age at first intercourse	
nonintact	209.6
intact	216.6

Panel B	months of exposure	
	60	90
Predicted percentage, premarital first birth		
A: nonintact	28.4	33.8
B: intact	12.6	16.1
Predicted difference, premarital first birth		
C: A − B	15.8	17.7
Decomposition of C		
D: direct component	11.1	13.3
E: indirect component, differential age at onset	.8	1.0
F: indirect component, differential exposure	3.8	3.4

Table 6. Decomposition of the difference in the percentage of women predicted to have a premarital first birth, controlling for other covariates. Comparison of women with mother's education half a standard deviation above and below the mean.

Panel A	age in months
Predicted median, age at first intercourse	
mean −.5 s.d.	213.7
mean + .5 s.d.	213.5

Panel B	months of exposure	
	60	90
Predicted percentage, premarital first birth		
A: mean −.5 s.d.	20.8	25.0
B: mean + .5 s.d.	18.8	22.7
Predicted difference, premarital first birth		
C: A − B	2.0	2.3
Decomposition of C		
D: direct component	2.1	2.5
E: indirect component, differential age at onset	−.0	−.0
F: indirect component, differential exposure	−.1	−.1

Table 7. Decomposition of the difference in the percentage of women predicted to have a premarital first birth, controlling for other covariates. Comparison of women with early and late menses.

Panel A	age in months
Predicted median, age at first intercourse	
early menses	211.7
late menses	216.5

Panel B	months of exposure	
	60	90
Predicted percentage, premarital first birth		
A: early menses	20.4	24.6
B: late menses	17.4	21.7
Predicted difference, premarital first birth		
C: A − B	3.0	2.9
Decomposition of C		
D: direct component	−.1	−.1
E: indirect component, differential age at onset	.6	.7
F: indirect component, differential exposure	2.6	2.3

Table 8. Decomposition of the difference in the percentage of women predicted to have a premarital first birth, controlling for other covariates. Comparison of women with low and high income-to-needs in her family of origin.

Panel A	age in months
Predicted median, age at first intercourse	
low income-to-needs	213.4
high income-to-needs	213.8

Panel B	months of exposure	
	60	90
Predicted percentage, premarital first birth		
A: low income-to-needs	22.9	27.5
B: high income-to-needs	16.9	20.5
Predicted difference, premarital first birth		
C: A − B	6.0	7.0
Decomposition of C		
E: direct component	5.8	6.8
E: indirect component, differential age at onset	.0	.1
F: indirect component, differential exposure	.2	.2