A Multiprocess Model for Correlated Event Histories with Multiple States, Competing Risks, and Structural Effects of One Hazard on Another

Fiona Steele
Centre for Multilevel Modelling, Institute of Education, University of London

May 2005

This research note should be read in conjunction with:

Steele, F., Kallis, C., Goldstein, H. and Joshi, H. (2005) "The relationship between childbearing and transitions from marriage and cohabitation in Britain", *Demography*, 42 (to appear in November issue).

1. Introduction

Steele et al. (2005) describe a general framework for the analysis of correlated event histories, with an application to a study of partnership transitions and fertility among a cohort of British women. A multilevel multistate competing risks model is used to examine the relationship between prior fertility outcomes (the presence and characteristics of children and current pregnancy) and the dissolution of marital and cohabiting unions, and movements from cohabitation to marriage. Using a simultaneous equations model these partnership transitions are modelled jointly with fertility, allowing for correlation between the unobserved woman-level characteristics that affect each process.

In this research note, we show how the model proposed by Steele et al. (2005) may be extended to allow for structural effects of partnership stability, as measured by the latent hazards of partnership transitions, on childbearing within partnerships. The method we describe generalises the approach of Lillard and Waite (1993) to multiple states and competing risks.

2. A recursive multiprocess model for partnership transitions and childbearing

The multiprocess model proposed by Steele et al. (2005) is a system of five simultaneous equations. Each equation defines a multilevel discrete-time event history model.

Partnership transitions

Three types of partnership transition are considered: marital separation, separation from cohabitation, and cohabitation to marriage with the same partner.

<u>Marital separation</u>

We denote by $h_{ij}^{PM}(t)$ the hazard of a marital separation during time interval t of episode i for individual j. The model for marital separation may be written (omitting subscripts) as:

$$\text{logit}\, h^{PM}(t) = \alpha_0^M D^{PM}(t) + \alpha_1^M X(t) + \alpha_2^M F(t) + \alpha_3^M X^{PM}(t) + u^{PM} \ . \qquad (1)$$

$\alpha_0^M D^{PM}(t)$ is the baseline logit-hazard which is a function of marriage duration at time *t* or, for marriages preceded by a period of cohabitation, partnership duration. The potentially endogenous time-varying outcomes of the fertility process which may affect future partnership transitions and fertility are denoted by $F(t)$, with coefficient vector α_1^M. Other covariates which affect marital dissolution, but which are treated as exogenous, are represented by $X(t)$ and $X^{PM}(t)$. Variables contained in $X(t)$ may also appear in any other equation in the system, while $X^{PM}(t)$ is unique to the marital separation equation. Unobserved time-invariant individual-specific factors are represented by normally distributed random effects u^{PM}.

Outcomes of cohabitation

Denote by $h_{ij}^{PC(r)}(t)$ the hazard of a transition of type *r* from cohabitation, in time interval *t* of episode *i* for individual *j*, where *r*=0 (no transition), 1 (separation), or 2 (marriage). Transitions from cohabitation may be modelled using a multilevel discrete-time competing risks model:

$$\log\left[\frac{h^{PC(r)}(t)}{h^{PC(0)}(t)}\right] = \alpha_0^{C(r)} D^{PC(r)}(t) + \alpha_1^{C(r)} X(t) + \alpha_2^{C(r)} F(t) + \alpha_3^{C(r)} X^{PC(r)}(t) + u^{PC(r)} \qquad (2)$$

where $\alpha_0^{C(r)} D^{PC(r)}(t)$ is a function of cohabitation duration at time *t*, $X^{PC(r)}(t)$ are covariates that affect only the hazard of a transition of type *r* from cohabitation, and $u^{PC(r)}$ are individual and transition-specific random effects.

Childbearing within partnerships

Denote by $h_{ij}^{FM}(t)$ the hazard of a conception leading to a live birth within marriage during time interval *t* in partnership episode *i* for individual *j*. We denote by $h_{ij}^{FC}(t)$ the hazard of a conception within a cohabiting union. The model for fertility consists of separate equations for marriage and cohabitation, which are estimated simultaneously.

Marriage

A multilevel event history model for the waiting time to conception within marriage may be written (omitting subscripts):

$$\text{logit}\, h^{FM}(t) = \beta_0^M D^{FM}(t) + \beta_1^M X(t) + \beta_2^M F(t) + u^{FM} \qquad (3)$$

where $\beta_0^M D^{FM}(t)$ is a function of the partnership duration and, for second or higher order births, the duration since the previous birth, and u^{FM} is an individual-level random effect.

Cohabitation

The model for conceptions within cohabitation is written:

$$\text{logit}\, h^{FC}(t) = \beta_0^C D^{FC}(t) + \beta_1^C X(t) + \beta_2^C F(t) + u^{FC} \qquad (4)$$

where u^{FC} is an individual-level random effect.

Estimation

Equations (1)-(4) define a multiprocess model. These equations must be estimated simultaneously as there may be non-zero correlations between the woman-specific random effects across equations. We assume that $u = (u^{PM}, u^{PC(1)}, u^{PC(2)}, u^{FM}, u^{FC}) \sim N_5(\mathbf{0}, \Omega_u)$. Correlated random effects would arise if the unobserved characteristics that influence the timing of partnership transitions are correlated with those that affect childbearing within partnerships. Non-zero correlations between elements of $u^P = (u^{PM}, u^{PC(1)}, u^{PC(2)})$ and of $u^F = (u^{FM}, u^{FC})$ would suggest that any or all elements of $F(t)$ are endogenous with respect to partnership transitions. The multiprocess model can be framed as a multilevel bivariate discrete response model where for each time interval t of a partnership there are two responses: 1) a binary or multinomial response for the partnership status, and 2) a binary response indicating the occurrence of a conception. After converting the data into discrete-time format, the model may be estimated using existing procedures for multilevel binary and multinomial data. See Steele et al. (2004, 2005) for further details.

Identification

Identification of simultaneous equations models typically requires exclusion restrictions to be placed on the covariates. However, the observation of repeated events for a subset of women, with some overlap in events across processes, means that identification is possible without covariate exclusions. The model is identified under the assumption that all residual dependency between processes can be accounted for by allowing cross-process correlation between individual-level residuals that are constant across replications for the same individual. After accounting for this residual correlation, the remaining variation in the fertility outcomes $F(t)$ between partnership episodes represents the effects of prior fertility on the outcomes of marriage and cohabitation, controlling for selection bias. Identification of similar multiprocess models is discussed further in Lillard et al. (1995) and Upchurch et al. (2002).

3. Allowing for structural effects of partnership stability on childbearing

The multiprocess model defined by (1)-(4) allows estimation of the effects of prior fertility on the hazards of partnership transitions, accounting for the joint determination of partnership outcomes and fertility. However, it does not permit estimation of the reverse causal path; that is, the impact of partnership stability on the hazard of a conception. Following Lillard and Waite (1993), we may extend the model for fertility to allow for structural effects of the hazards of partnership transitions on the hazard of a conception to explore whether partnership instability might inhibit or precipitate a birth. Thus equations (3) and (4) become

$$\text{logit}\, h^{FM}(t) = \beta_0^M D^{FM}(t) + \beta_1^M X(t) + \beta_2^M F(t) + \lambda_1 \,\text{logit}\, h^{PM}(t) + u^{FM} \qquad (5)$$

and

$$\operatorname{logit} h^{FC}(t) = \beta_0^C D^{FC}(t) + \beta_1^C X(t) + \beta_2^C F(t) + \lambda_2 \log\left(\frac{h^{PC(1)}(t)}{h^{PC(0)}(t)}\right) + \lambda_3 \log\left(\frac{h^{PC(2)}(t)}{h^{PC(0)}(t)}\right) + u^{FC} \quad (6)$$

Equations (1), (2), (5) and (6) define a structural model. To estimate the structural model we first substitute expressions for the partnership transition hazards, (1) and (2), into (5) and (6) to obtain the reduced form model. Substitution of (1) into (5) gives

$$\operatorname{logit} h^{FM}(t) = \beta_0^M D^{FM}(t) + \gamma_0^M D^{PM}(t) + \gamma_1^M X(t) + \gamma_2^M F(t) + \gamma_3^M X^{PM}(t) + v^{FM}, \quad (7)$$

where

$$\begin{aligned}
\gamma_0^M &= \lambda_1 \alpha_0^M \\
\gamma_1^M &= \beta_1^M + \lambda_1 \alpha_1^M \\
\gamma_2^M &= \beta_2^M + \lambda_1 \alpha_2^M \\
\gamma_3^M &= \lambda_1 \alpha_3^M \\
v^{FM} &= u^{FM} + \lambda_1 u^{PM}
\end{aligned} \quad (8)$$

and substitution of (2) into (6) gives

$$\operatorname{logit} h^{FC}(t) = \beta_0^C D^{FC}(t) + \gamma_0^{C(1)} D^{PC(1)}(t) + \gamma_0^{C(2)} D^{PC(2)}(t) + \gamma_1^C X(t) + \gamma_2^C F(t) + \gamma_3^{C(1)} X^{PC(1)}(t) + \gamma_3^{C(2)} X^{PC(2)}(t) + v^{FC}, \quad (9)$$

where

$$\begin{aligned}
\gamma_0^{C(1)} &= \lambda_2 \alpha_0^{C(1)} \\
\gamma_0^{C(2)} &= \lambda_3 \alpha_0^{C(2)} \\
\gamma_1^C &= \beta_1^C + \lambda_2 \alpha_1^{C(1)} + \lambda_3 \alpha_1^{C(2)} \\
\gamma_2^C &= \beta_2^C + \lambda_2 \alpha_2^{C(1)} + \lambda_3 \alpha_2^{C(2)} \\
\gamma_3^{C(1)} &= \lambda_2 \alpha_3^{C(1)} \\
\gamma_3^{C(2)} &= \lambda_3 \alpha_3^{C(2)} \\
v^{FC} &= u^{FC} + \lambda_2 u^{PC(1)} + \lambda_3 u^{PC(2)}
\end{aligned} \quad (10)$$

Equations (1), (2), (7) and (9) define the reduced form model, which can be estimated in the same way as the model denoted by (1)-(4).

4. Recovering the structural parameters from the reduced form parameters

After estimation of the reduced form model, (8) and (10) are used to derive the structural parameters from the reduced form parameters as described below. Identification of the structural model requires covariate exclusions, specifically variables that affect partnership transitions but do not have direct effects on fertility decisions. The identifying variables are represented by $X^{PM}(t)$ and $X^{PC(r)}(t)$ in (1) and (2).

Fixed part parameters

The β s and λ s are the structural parameters in the fixed part of the model.

i) λ_1

Suppose we have a single instrument $X^{PM}(t)$, i.e. a variable in the model for marital separation, but not marital fertility.

From (8) we obtain

$$\lambda_1 = \frac{\gamma_3^M}{\alpha_3^M}$$

ii) β_1^M and β_2^M

Also from (8) we obtain:

$$\beta_1^M = \gamma_1^M - \lambda_1 \alpha_1^M$$

$$\beta_2^M = \gamma_2^M - \lambda_1 \alpha_2^M$$

iii) λ_2 and λ_3

Suppose that $X^{PC(1)}(t)$ and $X^{PC(2)}(t)$ each contain one variable which, for identification, must be different.

From (10) we obtain

$$\lambda_2 = \frac{\gamma_3^{C(1)}}{\alpha_3^{C(1)}}$$

$$\lambda_3 = \frac{\gamma_3^{C(2)}}{\alpha_3^{C(2)}}$$

iv) β_1^C and β_2^C

Also from (10) we obtain:

$$\beta_1^C = \gamma_1^C - \lambda_2 \alpha_1^{C(1)} - \lambda_3 \alpha_1^{C(2)}$$

$$\beta_2^C = \gamma_2^C - \lambda_2 \alpha_2^{C(1)} - \lambda_3 \alpha_2^{C(2)}$$

Random part parameters

The random part parameters in the structural model are the variances of u^{FM} and u^{FC} and their covariances with $(u^{PM}, u^{PC(1)}, u^{PC(2)})$.

From (8) we obtain

$$u^{FM} = v^{FM} - \lambda_1 u^{PM}$$

and $\Omega_u^{FM} = \Omega_v^{FM} + \lambda_1^2 \Omega_u^{PM} - 2\lambda_1 Cov(v^{FM}, u^{PM})$.

From (10) we obtain

$$u^{FC} = v^{FC} - \lambda_2 u^{PC(1)} - \lambda_3 u^{PC(2)}$$

and

$$\Omega_u^{FC} = \Omega_v^{FC} + \lambda_2^2 \Omega_u^{PC(1)} + \lambda_3^2 \Omega_u^{PC(2)} - 2\lambda_2 Cov(v^{FC}, u^{PC(1)}) - 2\lambda_3 Cov(v^{FC}, u^{PC(2)}) + 2\lambda_2 \lambda_3 Cov(u^{PC(1)}, u^{PC(2)})$$

The covariances between u^{FM} and u^{FC} and elements of $(u^{PM}, u^{PC(1)}, u^{PC(2)})$ are given by:

$$Cov(u^{FM}, u^{PM}) = Cov(v^{FM} - \lambda_1 u^{PM}, u^{PM}) = Cov(v^{FM}, u^{PM}) - \lambda_1 \Omega_u^{PM}$$

$$Cov(u^{FM}, u^{PC(r)}) = Cov(v^{FM} - \lambda_1 u^{PM}, u^{PC(r)}) = Cov(v^{FM}, u^{PC(r)}) - \lambda_1 \Omega_u^{PC(r)} \qquad \text{(r=1,2)}$$

$$\begin{aligned} Cov(u^{FC}, u^{PM}) = Cov(v^{FC} - \lambda_2 u^{PC(1)} - \lambda_3 u^{PC(2)}, u^{PM}) &= Cov(v^{FM}, u^{PM}) \\ - \lambda_2 Cov(u^{PC(1)}, u^{PM}) - \lambda_3 Cov(u^{PC(2)}, u^{PM}) \end{aligned}$$

$$\begin{aligned} Cov(u^{FC}, u^{PC(r)}) = Cov(v^{FC} - \lambda_2 u^{PC(1)} - \lambda_3 u^{PC(2)}, u^{PC(r)}) &= Cov(v^{FM}, u^{PC(r)}) \\ - \lambda_2 Cov(u^{PC(1)}, u^{PC(r)}) - \lambda_3 Cov(u^{PC(2)}, u^{PC(r)}) \end{aligned} \qquad \text{(r=1,2)}$$

5. Discussion

As noted above, identification of the multiprocess model with structural effects of one hazard on another requires exclusion restrictions to be placed on the covariates. In our application, we would need to find variables which influence partnership transitions, but which do not have direct effects on the chance of conception within marriage and cohabitation. Such variables are difficult to find because partnership and childbearing decisions are so closely interlinked. Lillard and Waite (1993) successfully applied a full structural model in their analysis of marital dissolution and fertility in the US by exploiting state-level variation in divorce laws and the cost of divorce to identify the effects of marital stability on the hazard of a conception. In Britain, there is no such geographical variation that can be used for identification and, in any case, one would expect measures of divorce to have only a weak relationship with the outcomes of cohabitation. Alternative candidates for instruments were found to be only weakly correlated with the hazards of partnership transitions and/or had significant direct effects on fertility. As a result, estimates of the structural parameters based on these variables had extremely large standard errors and are likely to be biased.

Another area which would benefit from further research is how to obtain unique estimates of the structural parameters when there are multiple instruments. For example, a two-dimensional vector $X^{PM}(t)$ would lead to two alternative estimates of λ_1, i.e. the model is over-identified. In exploratory analysis, we considered two instruments for marital stability. One of these was found to be a poor instrument as it was only weakly correlated with the hazard of marital separation, leading to a large standard error for the estimate of λ_1 based on this variable. In a small-scale simulation study, we found that, in large samples, the two estimates of λ_1 will be approximately equal if the two instruments are equally correlated with the outcome. However, it is not clear how to proceed when there are alternative estimates. One possible solution would be to take the estimate based on the 'best' instrument, or to replace the set of candidate instruments by a combination of them. Another solution might be to impose constraints on the parameters of the reduced form model, but this could be difficult to implement as the constraints would be nonlinear.

References

Lillard, L. and Waite, L. (1993) "A joint model of marital childbearing and marital disruption", *Demography*, 30: 653-681.

Lillard, L.A., Brien, M.J. and Waite, L.J. (1995) "Premarital cohabitation and subsequent marital dissolution: a matter of self-selection?" *Demography*, 22: 437-457.

Rasbash, J., Steele, F., Browne, W.J. and Prosser, B. (2004) *A User's Guide to MLwiN, Version 2.0*. London: Institute of Education.

Steele, F., Goldstein, H. and Browne, W. (2004) "A general multistate competing risks model for event history data, with an application to the study of contraceptive use dynamics", *Journal of Statistical Modelling*, 4: 145-159.

Steele, F., Kallis, C., Goldstein, H. and Joshi, H. (2005) "The relationship between childbearing and transitions from marriage and cohabitation in Britain", *Demography*, 42 (to appear).

Upchurch, D.M., Lillard, L.A. and Panis, C.W.A. (2002) "Nonmarital childbearing: influences of education, marriage, and fertility", *Demography*, 39: 311-329.