J Popul Econ (1997) 10:23–65

Journal of Population Economics

Dynamic microeconomic models of fertility choice: A survey

Cristino R. Arroyo[1], **Junsen Zhang**[2]

[1] Department of International Economics, Johns Hopkins University, Nitze School of Advanced International Studies, 1740 Massachusetts Ave. NW, Washington, DC 20036-1984, USA
(Fax: (202) 663-5683, e-mail: carroyo3@aol.com)
[2] Department of Economics, Chinese University of Hong Kong, Shatin, NT, Hong Kong
(Fax: (852) 2603-5805, e-mail: jszhang@cuhk.edu.hk)

Received August 26, 1994/Accepted October 23, 1996

Abstract. We review existing approaches to the specification and estimation of dynamic microeconomic models of fertility. Dynamic fertility models explain the evolution of fertility variates over the life-cycle as the solution to a dynamic programming model involving economic choices. Dynamic models may be classified into structural and reduced-form models. Structural models generally require solution of the underlying dynamic programming problem. Reduced-form models, while based on a structural specification, do not. Recent innovations in estimation methodologies make both types practical and realistic alternatives to static models of lifetime fertility.

JEL classification: J13, C41, C61

Key words: Fertility, dynamic micro models

---

1. Introduction and motivation

This paper surveys dynamic microeconomic models of fertility choice that have appeared in the recent literature. By dynamic models we mean ones that explicitly model the *time profile* or *intertemporal evolution* of fertility choice

All correspondence to Junsen Zhang. The authors thank Chunrong Ai, Susan Cochrane, John Geweke, V. Joseph Hotz, Robert Miller, Ali Tasiran, and Kenneth Wolpin for providing us with source material. We are also grateful to the anonymous referees, and the responsible editor of this journal for their encouragement and patience, and for contributing several helpful suggestions that greatly increased the value-added of this paper. Any remaining errors are the sole responsibility of the authors. *Responsible editor:* Klaus F. Zimmermann.

and outcomes, as distinct from models that study once-and-for-all lifetime fertility decisions. Within the class of dynamic fertility models, we offer the following general taxonomy. A model may be considered *structural* when its estimable components require the specification and exact solution of an explicit dynamic maximization problem. *Reduced-form* dynamic models are those which may estimate structural parameters or relationships arising from a dynamic programming problem, but do not rely on the exact solution of the dynamic equations. For instance, the form of a reduced-form model may be specified in conformity with some tractable econometric framework not tied to the exact form of the model's solution. Our paper concentrates on these two general approaches to analyzing fertility dynamics.[1]

To further sharpen our focus we survey only *economic* models of fertility, i.e., models in which fertility variates are substantially outcomes of an economic choice process. This deliberately omits discussion of an important research program that studies fertility dynamics but which emphasizes biological rather than economic determinants (i.e., the seminal work of Sheps and Menken 1973 and derivative papers). We neglect these models, but do so without prejudice to their valuable contributions. Finally, we also omit aggregative or macroeconomic models of fertility,[2] even if these have microfoundations.

In relation to other surveys of this literature, this paper fills a gap that is not covered by the recent surveys of Montgomery and Trussell (1984), Eckstein and Wolpin (1989), Rust (1994), and Olsen (1994). Olsen (1994) discusses several economic approaches to fertility notably what may be called the Easterlinian, Schultzian, and Beckerian approaches after those respective authors. All of the approaches surveyed in Olsen lead to essentially static models of lifetime fertility choice.[3] While not their sole emphasis, the more technical surveys of Montgomery and Trussell, Eckstein and Wolpin, and Rust (1994) issues relevant to the specification of dynamic fertility models. Eckstein and Wolpin, and Rust (1994) outline a general dynamic framework that encompasses many discrete-choice models used in other research besides fertility. Like these studies, we cover general approaches, but our emphasis is specifically on dynamic fertility models. Montgomery and Trussell provide a more focused discussion of fertility issues, albeit in the context of the handful of dynamic models around at the time. (They also give a good perspective of the relationship between life-cycle models of fertility against the research on lifetime fertility behavior and female labor supply.) Our paper may be regarded as both an updating of Montgomery and Trussel's (1986) survey and an expansion upon the issues of dynamic specification and estimation.

Recent research has turned up many stylized empirical regularities with a time dimension.[4] None of these facts can be adequately explained by a model of lifetime fertility choice, however sophisticated, since lifetime fertility models lack a true time dimension. Treating fertility decisions as once-and-for-all choices makes it difficult, if not impossible, to connect realized fertility to variations in contraception costs, wages, income, education, mortality risks, or women's labor market participation over time, nor can these explain birth-timing or spacing. Recent efforts, consequently, have focused on tractable models that incorporate fertility dynamics. This new wave of dynamic analysis reflects both the conscious attempt to overcome prior analytical limitations and the influence of methodological developments in other areas of economics.

Dynamic approaches based on life-cycle choice represent a significant philosophical departure from the traditional static view of the fertility decision. Unlike the one-shot family size decision of static models, fertility choices over the life-cycle become a plan for *each time period* and *each contingency* (i.e., a *policy*). Dynamic models also result in optimal fertility policies that are *forward-looking* and *time-consistent*.[5]

From an econometric standpoint two further differences may be noted. *Structural* dynamic models generally prespecify stochastic elements *as an integral component* of the model, whereas lifetime choice models will introduce these directly into the estimating relations. This, however, can make a big difference not just for interpretation of the econometric estimates, but for the appropriate choice of the estimator itself. Further, it makes more sense to imbed certain types of randomness in a dynamic setting; for instance, uncertainty about natural fecundity or preferences on family size can evolve as individuals acquire fertility experience. Finally, to the extent that some static models represent reduced-forms or time-aggregated versions of a dynamic model, information is lost in reducing or aggregating the model over time. When possible, estimation of a structural model can be more informative.

These considerations work in favor of dynamic life-cycle fertility models. Counterbalancing these advantages, however, is the fact that, at the present juncture, the use of dynamic models still requires a trade-off between *model realism* and *model tractability*. This will become more apparent in the ensuing text.

Our paper is organized as follows. Section 2 studies a general framework nesting several well-known fertility models of the structural variety, including that of Heckman and Willis (1976), Wolpin (1984), Hotz and Miller (1984), Rosenzweig and Schultz (1985), Newman (1988), and Leung (1991). Section 2 also covers recent estimation/simulation techniques that have been proposed for structural models of the discrete-choice type. Section 3 reviews reduced-form models, focusing primarily on specification and estimation issues of hazard-rate models and linear approximations to optimal decision rules. Section 4 surveys other dynamic models that are neither structural nor reduced-form in nature.

2. Structural models

2.1 General framework

Introducing more general and "realistic" types of uncertainty, patterns of serial correlation, and time-variation with economic or biological variables requires an explicitly dynamic framework. A *structural* approach (i) models observed fertility as part of the solution of an explicit dynamic programming problem, and (ii) derives estimable or computable relationships from this solution.

A general framework nesting several well-known dynamic models[6] may be given as follows. An individual i maximizes the (expected, discounted) value of utility over her life-cycle $t = \tau, ..., T$:

$$\max E_\tau \sum_{t=\tau}^{T} \beta^t U(N_t, M_t, X_t, H_t, u_t, \theta) \tag{1}$$

using all available information known as of τ. (We have suppressed the i index for notational clarity.) The uncertainty underlying the conditional expectation E_τ generally comes from the risk of a new birth and from infant mortality risk, though there can be other sources (e.g., income shocks, wage shocks, preference shocks). Where convenient we represent all sources of uncertainty as a general error term ε_t with well-defined joint density function $f(\varepsilon_t)$.

The term $\beta < 1$ is a discount factor, U is the period utility function at t, and θ is a preference parameter which, for the moment, we regard as a once-and-for-all shock realized at τ. (θ thus reflects heterogeneity among individuals.) The time-varying arguments of the utility function, N_t, M_t, X_t, H_t, are, respectively, current births, family size[7], the quantity of market goods consumed, and non-work (leisure) time. In some models, such as Hotz and Miller (1984), leisure is an input into a production function for home goods Z_t that enter the period utility function. Here H_t gets subsumed into utility directly. The variable u_t measures contraceptive efficiency, generally assumed to lie between a lower bound $\underline{u} \geq 0$ and an upper bound $\bar{u} \leq 1$. Table 1 summarizes the forms of U used in several models that are special cases of the general framework above.

The analyst's choice of which arguments to include in U depends largely on the research question of interest. For example, non-work time, H_t, appears in Hotz and Miller (1984), a study intimately concerned with the

Table 1. Specifications of period utility, various structural models

Heckman and Willis (1976):
$U = W(\psi N_t, X_t) - f(u_t)$
Wolpin (1984):
$U = W(M_t, X_t, \theta)$
$= (\alpha_1 + \theta) M_t - \alpha_2 M_t^2 + \beta_1 X_t - \beta_2 X_t^2 + \gamma M_t X_t$; γ any sign
Hotz and Miller (1984):
$U = W(M_t, Z_t)$; $Z_t = Z(H_t, X_t, \zeta_t)$
where Z_t is household production
ζ_t is a random error
Rosenzweig and Schultz (1985):
$U = W(N_t, M_t, X_t, H_t, \theta)$
$= \phi_1 N_t - \phi_2 N_t^2 + \alpha_1 M_t - \alpha_2 M_t^2 + \beta_1(\theta) X_t - \beta_2 X_t^2 + \delta_1 H_t - \delta_2 H_t^2 + \gamma H_t M_t$; γ any sign
Newman (1988):
$U = W(M_t, X_t, u_t)$
$= \alpha_1 M_t - \alpha_2 M_t^2 + \beta_1 X_t - \beta_2 X_t^2 + \gamma M_t X_t + \rho_1 u_t - \rho_2 u_t^2$; γ any sign
Leung (1991):
$U = W(M_t, X_t) - f(\pi(u_t))$
where $\pi(.)$ is the probability of a birth

Note: All parameters above, unless specified, are positive. All variables are described in the text of Sect. 2.1.

interaction between labor force participation of women and fertility decisions. There is, however, one important consequence of omitting contraceptive efficiency u_t in period utility U. As Montgomery and Trussell (1984, p. 261) point out, deletion of u_t as an argument in U leads to the result that the individual always chooses a level of contraceptive efficiency equal to one of the corners $\underline{u}$ or $\bar{u}$. Hotz and Miller (1984) and Rosenzweig and Schultz (1985) omit u_t; Heckman and Willis (1976), Newman (1988), and Leung (1991) do not. Wolpin assumes that individuals can choose the number of new births directly, but this is tantamount to selecting a zero or 100% effective contraceptive regime, if one were to assume further that a birth is sure to occur without contraception.

Maximization is subject to a sequence of budget constraints for each period t:

$$I_t + w_t(\bar{H} - H_t) = X_t + p_t^M M_t + p_t^u u_t, \tag{2}$$

where I_t is current (husband's) income, w_t is the individual's market wage, $\bar{H}$ is the total amount of time available for work, so that $(\bar{H} - H_t)$ is the mother's labor market time. The variables p_t^M, p_t^u are, respectively, the full dollar costs of an unit of M_t, and u_t, relative to the numeraire good X_t. Labor market returns, $w_t(\bar{H} - H_t)$, figure in Hotz and Miller (1984) and Rosenzweig and Schultz (1985) but not in any of the other studies noted above. Contraceptive costs $p_t^u u_t$ appear only in Rosenzweig and Schultz (1985) and Newman (1988). One common feature to all these models, however, is that capital markets for intertemporal transfers of wealth are nonexistent. This lack of an efficient mechanism for consumption-smoothing and the fact that children are durable goods imparts added value to children as "assets," over and above their value as "current goods" in U.[8]

The individual's choice variables are, generally, consumption X_t, nonwork time H_t, and contraceptive efficiency u_t, while family size evolves according to

$$M_{t+1} = M_t + N_t, \tag{3}$$

where N_t is the number (zero or one) of surviving newborn children in period t. Whether a net birth occurs or not typically depends on some stochastic birth and death processes, as well as the level of contraceptive efficiency. In general, one has

$$N_t = N(\pi^b(1 - u_t), \pi^{\mathrm{m}}), \tag{4}$$

where π^b is the probability of a birth assuming no contraceptive control (i.e., an individual's natural fecundity) and π^{m} is infant mortality risk. The probabilities π^b and π^{m} can vary with different biological, socio-economic, or even choice variables of interest to the analyst, and such variables are commonplace in empirical work based on these models.

At the theoretical level, however, the six models mentioned above treat π^b and π^{m} as exogenous (these are also generally assumed to be serially independent). On the other hand, that π^b and π^{m} are *known* to either the individual or the researcher is not a common assumption. Rosenzweig and Schultz (1985), for one, explore the effects of an individual's incomplete knowledge about π^b on the decision to contracept. But in Newman (1988) individuals do not, over time, learn more about their natural fecundity than they know at the onset of the fertile period. We assume that $\pi^b > 0$ over the individual's fertile cycle, although in Hotz and Miller (1984) the individual faces a probability of infertility before the end of the cycle.

Within specification (4) are nested two general types of fertility models: models of *perfect fertility control*, and *hazard models* in which fertility control is imperfect. If we set $\pi^b = 1$ always and we constrain the choice of u to 0 or 1 (either directly, or indirectly via omitting u in period utility U) then control over births is perfect as in Wolpin (1984) or Rosenzweig and Schultz (1985). Hotz and Miller (1984) differ slightly in that u is either $\underline{u} \geq 0$ or $\bar{u} \leq 1$. When $\underline{u} > 0$ and $\bar{u} < 1$, the Hotz-Miller model implies a nonzero birth hazard in each period, though it takes only discrete, dichotomous values. A true continuous hazard model arises in Heckman and Willis (1976), Newman (1988) and Leung (1991) as the optimal u is continuously-varying in the closed interval $[0, 1]$.[9, 10]

2.2 Predictions of the general framework

Contraceptive efficiency, birth spacing, and birth timing. In the general framework above, the rigor of contraceptive efforts employed over the life-cycle is the critical behavioral determinant of the onset, pace, and spacing of births, as well as realized family size over the life-cycle. *Ceteris paribus*, a more lax contraceptive strategy produces more frequent and more narrowly-spaced births. A useful starting point, then, is the analysis of dynamics of the contraceptive decision. To study this, define the value function V as the maximized value of the individual's problem (1) when X_t, H_t, amd u_t are optimally chosen. That is, if N_t^*, X_t^*, H_t^*, and u_t^* are the solution values for the maximization, then

$$\begin{aligned} & V(M_t;\, t,\, \pi^b, \pi^{\mathrm{m}},\, I_t,\, w_t,\, p_t^M,\, p_t^u,\, \beta,\, \theta) \\ & \quad = E_\tau \sum_{t=\tau}^{T} \beta^t U(N_t^*,\, M_t,\, X_t^*,\, H_t^*,\, u_t^*,\, \theta). \end{aligned} \tag{5}$$

To derive some analytical results the following simplifying assumptions will be convenient: (i) let $U_{N_t} = 0$, so that a current birth enters utility via next period's family size M_{t+1} only, and (ii) let the probability of a surviving birth be $\pi_t = \pi^b (1 - u_t) - \pi^{\mathrm{m}}$, so that mortality risk comes into play only in the same period as a birth. Suppressing other arguments in V save M and t, Bellman's optimality principle allows us to rewrite V as

$$\begin{aligned} V(M_t;\, t) = & \max \{U(M_t,\, X_t,\, H_t,\, u_t,\, \theta) + \beta E_t V(M_{t+1}; t+1)\} \\ = & \max \{U(M_t,\, X_t,\, H_t,\, u_t,\, \theta) + \beta [\pi_t V(M_t + 1;\, t+1) \\ & + (1 - \pi_t) V(M_t;\, t+1)]\}. \end{aligned} \tag{6}$$

(Note that $M_{t+1} = M_t$ with probability $1 - \pi_t$ and $M_{t+1} = M_t + 1$ with probability π_t.) Replace X_t with $I_t + w_t(\bar{H} - H_t) - p_t^M M_t - p_t^u u_t$ via (2).

Treating M_t as fixed, the partial derivative of the right-hand-side with respect to u_t gives the optimal contraception rule:

$$-U_{X_t} p_t^u + U_{u_t} - \beta \pi^b [V(M_t + 1;\ t+1) - V(M_t;\ t+1)] = 0$$

or

$$U_{X_t} p_t^u - U_{u_t} = \beta \pi^b [V(M_t;\ t+1) - V(M_t + 1;\ t+1)]. \qquad (7)$$

The left- and right-hand-sides of (7) are, respectively, the marginal costs of contraception (MCC) and the expected marginal benefits of contraception (EMBC). These are plotted in Fig. 1 over the range of u. The graph of $U_{X_t} p_t^u + U_{u_t}$ is drawn as an upward sloping curve taking on positive values, and this reflects the usual assumptions (see Heckman and Willis 1976, or Leung 1991) that $U_X > 0$, $U_{XX} < 0$, and that $U_u < 0$, $U_{uu} < 0$ (i.e., the increments of contraceptive disutility fall as one increases contraceptive levels). On the other hand, the graph of the expected marginal benefits of contraception is a flat curve in u, whose exact sign and location is determined by the sign and value of

$$\Delta V(M_{t+1};\ t+1) = V(M_t;\ t+1) - V(M_t + 1;\ t+1), \qquad (8)$$

which is the capitalized value (at t+1) of preventing a birth at t, given parity M_t. In all the models considered here, variations in the contraceptive

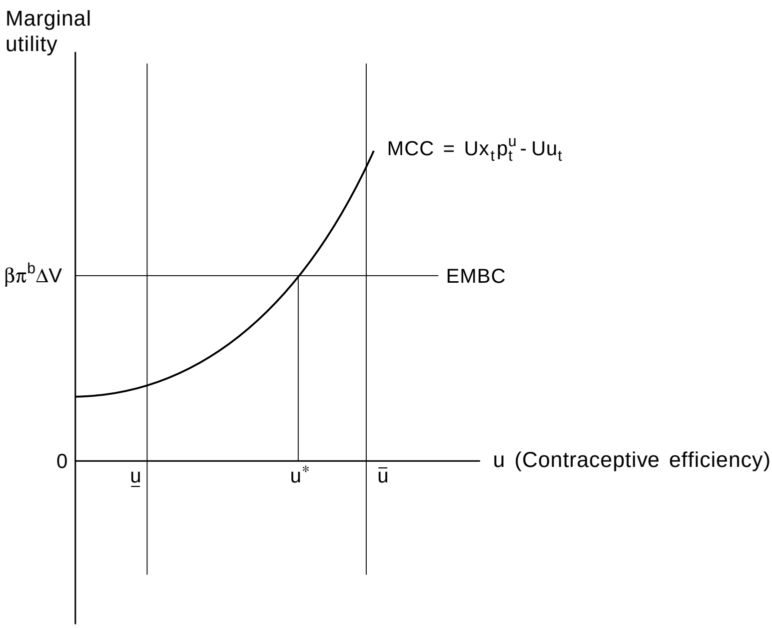

Fig. 1. Marginal costs of contraception, expected marginal benefits of contraception, and optimal contraceptive efficiency

decision (or the birth decision, in the case where fertility control is perfect) are due to shifts in their respective MCC and EMBC curves. These shifts govern the evolution of fertility control over time.[11] Via their effect on these curves, one may analyze the impact of time-variations in exogenous variables or parameters on the likelihood of a birth. The resulting changes in birth timing and spacing patterns, and completed fertility levels follow directly.

Equation (8) highlights the principal practical difficulty attendant to current structural models. With the exception of Newman's (1988) model, virtually no structural fertility model generates closed-form solutions for the EMBC curve as a function of basic variables and parameters. This creates substantial difficulties not just for arriving at an econometrically estimable version of the theory, it also makes it difficult to conduct comparative dynamics cleanly, with respect to the effects of changes in variables of interest on birth control and the birth hazard. For this reason, we rely heavily on the results of Newman (1988) to guide our discussion of comparative dynamics that follows. In addition, as the models nested inside the above general structure make different structural assumptions, the mapping from basic variables and parameters to MCC or EMBC will vary somewhat from model to model. Where these lead to substantive differences in predictions, we provide some extra discussion.

Effects of variations in family size. The six prototypical models mentioned above produce different predictions about the response of the birth likelihood and contraceptive efficiency to changes in family size M_t (i.e., parity). We now attempt to sort out these apparently conflicting findings.

A key theoretical result that frequently appears in the literature is concavity of the value function V in family size M_t. Under reasonable conditions, it has been shown that when the underlying period utility function U is concave in M_t, the value function V will also be concave in M_t.[12] This implies that ΔV in (8) will increase with parity, as the negative of ΔV will decrease with M_t. So as the number of births increase, the EMBC curve shifts up, which tends to raise contraceptive efficiency. The effect of larger M_t on the MCC curve, however, is not clear-cut, and its sign depends to some degree on how strongly children substitute for market goods. Studies with contraception in the utility function typically assume that the cross-partial $U_{Xu} = 0$. With this, the derivative of the left-hand-side of (7) with respect to M_t is

$$dMCC/dM_t = (U_{XM} - U_{XX}\, p_t^M)\, p_t^u.$$

Thus, if $U_{XM}/p_t^M - U_{XX}$ exceeds (is below) zero the MCC curve shifts up (down) as parity M_t rises. As $-U_{XX}$ is nonnegative, the MCC curve is guaranteed to shift up with rising M_t if children are gross complements to market goods, that is, $U_{XM} > 0$. In this case, the tension between the separate upward shifts of EMBC and MCC confound the final effect of rising M_t on the level of contraception u_t. More generally, when $U_{XM}/p_t^M - U_{XX} > 0$, the final effect on contraceptive efficiency is *ambiguous*, depending on the relative size of the shifts of EMBC and MCC.[13]

The shift of EMBC is large when (i) the discount factor β is large (i.e., people do not discount future utility so heavily), (ii) the probability of a birth π^b is high, and (iii) V is very concave in M_t, so that the response of ΔV to higher M_t is large (this usually means that the marginal utility of additional children falls quickly). All of these are more plausible for developed economies rather than developing economies.[14] The shift of MCC due to increased family size will be smaller and less positive (iv) the lower are the explicit costs of contraception, and (v) the larger is the utility loss from being able to consume fewer market goods. These are also more likely to hold in developed societies. The above five considerations are consistent with Newman (1988). Generally then, in developed economies, mothers with larger families would tend to have higher contraceptive efficiency levels. In these situations, births will tend to be fewer and more widely-spaced as parity rises. In developing economies, it could go either way.

There are two cases in which contraceptive efficiency increases *unambiguously* with parity. The first is when $U_{XM}/p_t^M - U_{XX} < 0$, which is really a strong form of (v) above. This guarantees that the MCC curve will shift downward, which together with an upward shift in EMBC, will ensure that the optimal u^* will be higher. The other case is when MCC does not shift at all, in which case the concavity of V in M_t, which is behind the upward shift of EMBC, is what drives higher contraceptive efficiency. Quite importantly, a positive or negative shift of MCC requires that p_t^u not be zero, which is the case in Rosenzweig and Schultz (1985) and Newman (1988), but not in any of the other studies.

For instance, Leung (1991) finds that contraceptive efficiency is always increasing in parity (despite the assumption that children and market goods are gross complements) and this result hinges entirely on V being concave in M_t. The omission of explicit contraception costs in the budget constraint, which Leung argues is not critical to his analysis, is partly responsible for his unambiguous result. Thus, unless one includes the effects on marginal contraception costs that channel through explicit costs of contraception p_t^u one does not get a decreasing frequency/probability of births with parity, nor any threshold effects, which have been detected in some developing economies.[15]

Effects of variation in elapsed time without a birth. As shown by Newman (1988) the optimal level of contraceptive efficiency decreases with the amount of time that has passed without a birth. In his model, this happens because with a fixed family size M_t, ΔV ends up being a negative function of time. The secular decline in ΔV, in turn, generates a downward drift in the EMBC curve while the MCC curve is not affected by elapsed time without a birth, resulting in a declining time path of optimal contraception levels, *ceteris paribus*. The intuition is as follows: suppose the EMBC curve was above the zero level in the preceding period, so that the individual found it beneficial to prevent a birth then. If the individual was lucky and no birth occurred in the previous period she can relax the control a little in the current period, as the remaining amount of time she faces a birth risk is shorter.[16] In effect, the individual finds it optimal to maintain an (almost) fixed "lifetime" hazard of an additional birth, which implies decreasing contraceptive vigilance as the end of the fertile cycle draws near.

Newman's result is a sharpening of an earlier conjecture made by Heckman and Willis (1976) which, in addition to the behavior suggested above, also considered the possibility that ΔV is constant over time, so that the birth hazard remains constant in each period that a birth does not occur, but the individual's "lifetime" hazard actually falls over time. In this case, the contraceptive response is invariant to elapsed time without a birth until the periodic birth hazard finally results in an "accidental" birth. When this happens the individual will then raise contraceptive efficiency because family size has increased, but contraceptive efficiency is kept steady until the next birth occurs, etc. Newman's solution, however, rules out this type of behavior.

Hotz and Miller (1984) supply an additional reason for why a downward drift in ΔV is likely. If, for a fixed family size, the requisite childrearing time falls as children age, then as time elapses without a new birth, the capitalized net benefits of contracepting, captured by ΔV, will fall naturally. This generally leads to a lowering of contraceptive efficiency levels and a higher *unconditional* birth hazard in each period. If, however, the individual faces a sufficiently large positive risk of infertility even bevore menopause is reached, the *conditional* hazard (i.e., the hazard, as of the present time t) of a birth $t+j$ periods away may not vary positively with elapsed time j. Let ρ be the probability of becoming permanently sterile in a period. The probability, then, of remaining fertile for j periods after the last birth is $(1-\rho)^j$. This is decreasing in elapsed time j. Consequently, the chances of a birth occurring at $t+j$ will be smaller the farther away $t+j$ is from t.

Effects of variations in natural fertility. Newman (1988) shows that an increase in the natural probability of a birth π^b (i.e., "fecundity") tends to raise the net birth probability π, even though the optimal response to a perceived increase in π^b is to raise contraceptive levels u_t. This can be seen, in a loose sense, in terms of the shifts in the MCC and EMBC curves, as π^b is a scale factor for EMBC. An increase in π^b would tend to shift the EMBC curve up without affecting the MCC curve. Now there is, however, an additional implicit effect of π^b on EMBC, which happens via the term ΔV. ΔV is implicitly a function of natural fertility π^b through future utility and the future optimal contraceptive levels of u_t^* (which depends on π^b). Intuitively, however, the effect of higher π^b should be to raise future contraceptive levels if it raises current contraception levels. But raising future contraceptive levels is consistent with a rise in ΔV, rather than a fall in ΔV. So the indirect effect on EMBC via ΔV should only reinforce the direct effect of a rise in π^b.

Since the birth hazard π^b $(1–u_t)$ rises, however, the implication is that more fecund women will have more births, and, because of the birth-spacing effect of increases in parity, will tend to cluster their births in the earlier years of their fertile cycle. This creates issues of how one handles heterogeneity in natural fertility levels when one goes about estimating hazard functions.

Rosenzweig and Schultz (1985) also find a theoretically positive contraceptive response to an increase in π^b, although in their framework a distinction is drawn between "permanent" or persistent shocks to natural fertility, and "transitory" ones. They find that the contraceptive response is larger if individuals know that the rise in π^b will persist for many periods, and is smaller if individuals are unsure about whether the rise in π^b comes from the permanent component of natural fertility or the transitory component. This lack of information further confounds one's ability to estimate the size of the effects of heterogeneous fertility levels on the birth hazard. The final effect on the birth hazard of a rise in π^b, which would incorporate the optimal contraceptive response to this change, is much more complicated than in Newman (1988).

Effects of variations in infant mortality risk. As the probability of an infant death, π^{m}), rises, this usually results in a lower *net* birth probability π, given that morality risks are typically modelled as affecting current birth outcomes only. The effect on net fertility should be the same as the case of a fall in the probability of a birth, $\pi^b(1-u)$. That is, the EMBC curve falls and contraceptive vigilance is relaxed. This type of response of increasing the birth hazard in response to higher mortality risk is known as *hoarding.*

If mortality risks also apply to older children, then lower fertility control resulting from *child replacement* motives may also appear. This type of birth hazard effect is distinct from hoarding in that it is a response to a realized infant death rather than a response to an increase in the likelihood of a death. Wolpin (1984) calculates that this effect is small in a sample of Malaysian households – there a child death typically causes families to increase the odds of a birth, but the increase in the expected number of children thereafter is only 0.015 children. In a dynamic model of fertility control the reason for this (cf. Newman 1988) is quite intuitive: the response of the optimal control level $u*$ is usually small when infant mortality rates are high since the optimal level of $u*$ already takes into account the large mortality risks. This effect is more difficult to explain, however, in a non-dynamic context of lifetime family size decisions.

Effects of variations in income. Positive variations in income I_t can come either in the form of (i) an one-time increase in beginning period income, (ii) a permanent increase in the level of I_t for all t, or (iii) keeping permanent income levels the same, a shift in lifetime income profiles with higher incomes occurring in later years. The models considered do not generally distinguish between cases (i) and (ii), as the associated behavioral responses are qualitatively the same and differ mostly in size, rather than in sign.

In the case of (i) or (ii) the one-time increase in income generally causes the MCC curve to shift down and to cause fertility control levels to rise because of negative cross effects with the marginal utility of market goods U_X.[17] The effect on EMBC, however, is not clear-cut and depends on parameters. Together with the shift of the MCC curve this results in an ambiguous outcome on the level of contraceptive efficiency $u*$ and the potential for threshold effects. In cases when the discount factor is large, the net birth probability π^b is high, or childrearing costs are high, etc., most of the models predict a negative effect on EMBC also (Newman 1988; Leung 1991; and Proposition 3 in Hotz and Miller 1984). In developed economies where this effect is expected to be large relative to the shift in MCC, the overall effect is a lowering of fertility control levels, implying larger expected family sizes in developed economies, *ceteris paribus.*

In the case of more steeply rising income profiles (iii), it is the birth-spacing/timing decision that is primarily affected. Heckman and Willis (1976) argue that households will tend to delay and space births until later in the life-cycle, when incomes have risen to levels most favorable for increasing family size. In the context of the MCC and EMBC curves, what one observes are joint shifts of the two curves in response to rising income, leading to progressively lower levels of contraception levels $u*$.[18]

Effects of variations in women's wages. An increase in w_t will have separate effects on the marginal costs and benefits of fertility regulation. The effect on MCC is such that fertility control goes up, due to an overall reduction in contraception costs. Lower marginal costs of contraception at all levels follow because MCC depends (in part) on the foregone marginal utility of consumption U_X. This falls with increases in the level of consumption X. Since consumption levels typically rise with wage rates, contraception costs associated with foregone consumption utility are smaller at all levels. Formally,

$$dMCC/dw_t = U_{XX}\, p_t^u\, (\bar{H} - H_t) < 0 \quad (provided\ \bar{H} - H_t > 0)$$

so that the MCC curve shifts down.

The effect on the EMBC curve, however, depends on the relative size of income and substitution effects that impact the ΔV component. How does a rise a w_t affect the value of preventing an additional child ΔV? The direct income effect of a relaxation of the budget constraint should lead individuals to lower fertility control. The substitution effect, however, can be of either sign. Without explicitly modelling the time costs of child-rearing, the sign of the substitution effect depends on how strongly complementary children are with both consumption goods and leisure time. If wages rise, individuals shift away from leisure into consumption goods. If children are strongly complementary with consumption, but not with leisure, individuals will want more children, and will lower fertility control. In this case the income and substitution effects on EMBC are reinforcing and cause a downward shift in EMBC. The final effect on contraceptive levels $u*$ is ambiguous, depending on the size of the shift in EMBC relative to MCC.

If, on the other hand, children are strongly complementary with leisure time but not with consumption, then the substitution effect on EMBC might outweigh the income effect, and the EMBC curve could conceivably shift upward, leading to an unambiguous increase in contraceptive efficiency $u*$. One model where this occurs is Hotz and Miller (1984). There, an explicit model of the time costs of raising children is grafted onto the basic framework (1)–(4). The result is that children and leisure time become strongly complementary, and EMBC rises in response to higher w_t. Alongside the downward shift of MCC, this leads to a clear-cut increase in contraceptive rigor following a rise in wages.

Effects of variations in childbearing and child maintenance costs. Finally, the effects of changes in the costs of child-rearing and child-maintenance have effects on fertility regulation not unlike those of the effects of changes in women's wage rates, since these are also relative price effects.

After the first child, a rise in p_t^M should generally lead to an upward shift in MCC as

$$dMCC/dp_t^M = U_{XX}\,p_t^u\,M_t > 0 \quad (provided\ M_t > 0).$$

Once again, however, the final effect on contraceptive efficiency u^* is generally ambiguous, as generally ΔV rises in response to higher p_t^M. (Usually, higher maintenance costs of a larger family size not only have negative income effects on household income, they usually also have negative substitution effects on the number of children born.) The rise in ΔV counters the effect of the shift in MCC.

2.3 Vijverberg's (1984) model

Before turning to estimation issues, we pause to examine Vijverberg's (1984) model, which is a legitimate dynamic structural model, but one which does not quite fit the above general framework. Vijverberg sets up a continuous-time counterpart to our general model, but in his view the decision on whether to have a birth in a period consists of two separate decisions: (i) how many children to have in the life-cycle, and (ii) given a chosen family size, how should births be spread out over the life-cycle, taking into account child maintenance costs, labor-market opportunities, and the individual's demand for leisure.

His model has some distinctive and interesting features, particularly in the way the individual's maximization is set up. Vijverberg divides up the life-cycle of an individual into time periods or intervals. The first set of intervals are simply the birth intervals, that is interval $t = 0$ is the time interval from marriage to the first birth, interval $t = 1$ is the time interval between the first and second birth, etc. Once desired lifetime family size I (assumed exogenous in Vijverberg's analysis) is attained, the relevant time intervals are no longer the birth intervals (there are no more). Rather, these become the amount of time it takes to "wean away" children. For instance, the $(I + 1)$st period is the interval between the last child's arrival and the weaning away of the first child, the $(I + 2)$nd interval is the period between the weaning of the first child and the weaning of the second, and so on. There is, finally, a last period, which starts at the point at which the Ith child is weaned away, and ends with the end of the conjugal relationship.

Under Vijverberg's theoretical specification, the individual is now imagined to maximize in stages. First, taking as *given* and *known* the "switchpoints" (i.e., the demarcation points between time intervals as defined above) the individual maximizes expected discounted utility inside each interval and adds these up to get a life-cycle utility number LU^* that is maximal, conditional on the known switchpoints. Next, the individual varies the switchpoints (which, given I, number 2 I+1) and selects that combination which maximizes LU^*. Vijverberg then suggests that a further maximization "stage" is possible in which the individual maximizes the value of LU^* by varying completed family size I.

Like Hotz and Miller (1984) Vijverberg examines the relationship between fertility variates and the time-allocation and labor supply decisions of the household. The key element of his empirical model is a "switchpoint equation", which guides the decision to accelerate or delay a birth. While certainly interesting, the derivation of this equation and the signs of its parameters are fairly involved. We omit this discussion so that we can focus on his empirical findings.

Though his sample did not have a large proportion of women with multiple births, Vijverberg nonetheless found that higher wages tend to cause individuals to delay births, as leisure time and children appear to be complementary. Several other predictions of his model were also confirmed: women appear to choose between a career and raising children, higher child maintenance costs discourage births, higher (permanent) husband's income seems to encourage earlier births, and (contrary to Heckman and Willis' findings) a rising income profile for women (reflected in higher *predicted* future wages) tends to also cause earlier births!

2.4 Direct estimation of structural parameters

We now look at attempts and suggested methods for estimating the structural parameters of a dynamic programming model like (1)–(4). The bases for estimation are either or both of the value function (6) or the optimal policy function (7). Because tractable closed-form solutions for (6) or (7) are unavailable, most approaches to estimation require augmenting maximum-likelihood procedures with numerical solution of the dynamic program.

Continuous-valued choice variables. Though not exactly tractable, Newman's (1988) fertility model possesses a closed-form solution for the optimal policy and value functions. An earlier paper, Newman and McCullough (1984), did in fact estimate a similar-looking model using hazard analysis methods as described in Sect. 3 below. However, as pointed out in Newman (1988), the Newman-McCullough estimation was not fully structural, since it did not use the form of the hazard function with the optimal contraceptive policy inserted. A more thorough application of hazard-rate estimation to Newman's unrestricted model, is not straightforward, and has yet to be worked out.

Newman's (1988) model is one in which the choice variable, contraceptive efficiency, is allowed to vary continuously over an interval. (In addition, his model is set up in continuous rather than discrete time.) In the context of time-series models, some recent methods have been developed for the structural estimation of dynamic programming problems in which the choice variables are continuous (see Taylor and Uhlig 1990; Smith 1992). These new methods, however, are suitable for: (i) time-series rather than panel or cross-sectional data, and (ii) discrete-time, rather than continuous-time equations. Bridging the gap between these new methods and the Newman model may be a fruitful area for future exploration. Moreover, with respect to (ii), one advantage is that the new estimation methods do not require closed-form solutions, so that abandoning the continuous-time formulation of Newman (1988) may preclude a closed-form solution for (6) or (7) but will not compromise estimation.

Discrete choice models. In the realm of models of discrete fertility choice, Rosenzweig and Schultz (1985) is an early model featuring a dynamic structure *a la* (1)–(4). However in estimating the model they apply a time-averaging procedure that renders the estimating model nondynamic. Moreover, they also do not solve the dynamic program, numerically or analytically, or obtain structural estimates of the parameters of the value function (6) or the optimal policy function (7). Hotz and Miller (1988) assume linear approximations to the true optimal policy functions derived in Hotz and Miller (1984). Since it is difficult to map back directly from the fitted linear coefficients to the original structural parameters, their approach is essentially of the reduced-form variety, and is discussed in Sect. 3 below.

Thus far Vijverberg (1984), Wolpin (1984), and, more recently, Hotz and Miller (1993) are the only studies known to us which estimate structural parameters of a dynamic model. Wolpin (1984) appears to be the precursor in the fertility literature of the new breed of estimable dynamic structural models. These, by necessity, combine numerical solution of the dynamic program with maximum likelihood estimation of the structural parameters. It is particularly instructive to examine his approach. Wolpin attempts estimation of a parameter vector consisting of the utility function parameters, the (assumed) fixed prices of a new birth and consumption, the time discount factor, and other parameters of the optimal policy (7). Call this parameter vector α.

Estimation is based on the following decision rule: let $V(M_t;\, t)$ be the value function for the individual's utility from period t onward, given the current value of the state M (in this case, the stock of children). Let θ_t be a random shock to preferences now allowed to vary with t, and let P_t be the probability at t that an infant will survive into the next period. Under the functional form of U assumed by Wolpin (see Table 1), the decision to add a child or not[19] can be shown to depend upon the function:

$$\begin{aligned} J_t &= E_t\left[V(M_t + 1;\, t+1) - V(M_t;\, t+1)\right] + P_t\,\theta_t \\ &= -E_t\left[\Delta V(M_{t+1};\, t+1)\right] + P_t\,\theta_t, \end{aligned} \tag{9}$$

where $E_t\,\Delta V(M_{t+1};\, t+1)$ denotes the capitalized value (at t+1) of preventing a birth at t.[20] The second term on the right-hand side is the expected change in utility due to any preference shocks for more children. The optimal decision rule for births is therefore:

$$N_t^* \begin{cases} = & 1 \text{ iff } J_t > 0 \\ = & 0 \text{ iff } J_t \le 0. \end{cases} \tag{10}$$

Optimal births N_t^* thus depend on the value of the function J_t, which in turn depends on the parameters of the unknown value functions $V(\cdot;\, t+1)$, the parameters of the conditional distribution function that individuals use to calculate the conditional expectation E_t, the survival probability P_t, and the preference shock θ_t. Inspection of (9) and (10) reveals that the model follows a classic probit structure, except that one lacks a closed-form expression for $E_t\left[\Delta V(M_{t+1},\, t+1)\right]$ in terms of α. With some additional assumptions, however, Wolpin demonstrates that this model is still estimable.

Consider (9). Wolpin argues that there is always a unique value of θ_t, denoting this critical value θ_t^*, for which the indicator function J_t will be zero. This value is just

$$\theta_t^* = E_t\,[\Delta V\,(M_{t+1};\ t+1)]/P_t. \tag{11}$$

Assume that individuals know the random shock θ_t at the time of the current period fertility decision, but the analyst does not ever observe θ_t. The analyst cannot then observe J_t directly. Current births' N_t^*, however, are observable. To proceed Wolpin makes the classical probit assumption that θ_t^* is a normally distributed random variable. Then the probability that the individual chooses a birth in period t (conditional on M_t) is

$$Pr\,[N_t = 1|M_t] = 1 - \Phi\,(\theta_t^*/\sigma), \tag{12}$$

where σ is the standard error of θ_t^* and $\Phi\,(\mathrm{x})$ is the value at x of the standard normal cumulative density function. Analogously, the conditional probability of no birth at t is

$$Pr\,[N_t = 0|M_t] = \Phi\,(\theta_t^*/\sigma). \tag{13}$$

Let Ω be the set of time periods where there is a birth, and Ω^c its complement. *For individual i*, the likelihood L^i of any particular birth pattern is

$$L^i = \prod_{t\in\Omega} Pr\,[N_t = 1|M_t] \prod_{t\in\Omega^c} Pr\,[N_t = 0|M_t], \tag{14}$$

where it is understood that Ω, Ω^c, N_t, M_t, a*nd* θ_t^* all depend on i, and that L^i, in general, is a function of parameters α via θ_t^*. Given a sample of I individuals, the sample likelihood is

$$L = \prod_{i=1}^{I} L^i, \tag{15}$$

which is maximized with respect to α.

This the first part of the estimation methodology; the critical second component of the methodology is the numerical solution of the dynamic programming problem. This is where the main computational issues associated with structural models generally arise.[21] Maximizing L involves evaluating θ_t^*, which requires, by (11), evaluation of the conditional expectation $E_t\,V\,(M_{t+1};\ t+1)$ at $M_{t+1} = M_t + 1$ and $M_{t+1} = M_t$. Let ε_{t+1} be next period's errors[22] and $f\,(\varepsilon_{t+1}|M_t)$ their conditional density. The conditional expectation at t of $V\,(M_{t+1};\ t+1)$ is

$$E_t\,V\,(M_{t+1};\ t+1) = \int_{\varepsilon} V\,(M_{t+1};\ t+1) f\,(\varepsilon_{t+1}|M_t)\,d\varepsilon_{t+1}. \tag{16}$$

On the face of it, given the current state M_t and a functional form for f, only one integration appears in (16), and the problem seems easy. But with no closed-form solution for the optimal fertility policy, the form of $V(M_{t+1};\, t+1)$ *is not known* for any of the realizations of ε_{t+1}. Indeed, one must solve the entire dynamic program, which at the time required backward solution from period *T*, following Bellman's recursion:

$$\begin{aligned} E_t\, V(M_{t+1};\, t+1) &= E_t\{\max\,[U(\cdot;\, M_t) + \beta\, E_{t+1}\, V(M_{t+2};\, t+2)]\} \\ &= E_t\{\max\,[U(\cdot;\, M_t) + \beta\, E_{t+1}\, \{\max\, U(\cdot;\, M_{t+1}) \\ &\quad + \beta\, E_{t+2}\,[V(M_{t+3};\, t+3)]\}]\} \end{aligned} \tag{17}$$

and so on until *t+j=T*.

To illustrate the computational complexity involved, consider the following specialization of (17). Suppose that in (17) $T = t+3$. Evaluating $E_{t+2}\,[V(M_{t+3};\, t+3)]$ is trivial and involves no integration because in the last period $V(M_{t+3};\, t+3) = U(\cdot;\, M_{t+3})$. To evaluate $E_{t+1}\,[V(M_{t+2};\, t+2)]$, however, requires a nontrivial maximization which involves an integration for each value of the state M_{t+2}. If the state variable can take on $G(M_{t+2}) = G_{t+2}$ different values, one has to calculate G_{t+2} integrals. This done, one can take a step backward and solve for $E_t\, V(M_{t+1};\, t+1)$. But this requires the calculation of G_{t+1} integrals, one for each possible value of the state M_{t+1}. In total, solving this 3 period problem required $G = G_{t+1} + G_{t+2}$ integrations.

Now imagine an extremely simple life-cycle problem where T=21 periods, and that the state vector M_t consists solely of a family size variable.[23] Conditional on today's state M_t, next period's state M_{t+1} can assume two different values, M_t or $M_t + 1$. This implies a total number of (21–1)×2=40 integrations that must be calculated to solve for the value of $E_t\, V(M_t + 1;\, t+1)$ in the first line of (9). In addition, one also needs values for $E_t\, V(M_t;\, t+1)$, i.e., the expected value function without an extra child. This implies another 40 integrations. In total, 80 integrations are needed to solve the dynamic program *of a single individual* for a single value of θ_t^* in (11).

Suppose one has a sample of 100 individuals. This implies that 8000 integrations must be performed in order to evaluate the value of the likelihood *L* at the initial setting of the parameters α. But maximizing *L* numerically requires iterating on the value of α. At each step, i.e., at each new trial value for α, one must perform 8000 integrations to get a new likelihood value *L*, so that convergence in 10 iterations implies 80,000 integrations. Clearly, even for a very unrealistic model, the computational burden can be quite formidable.[24] These computational demands go up exponentially as one increases the dimensionality of the state vector M_t, the number of time periods, the sample size, the dimensionality of the parameter space, or the number of times one must obtain estimates of α, say for policy experiments. This is essentially Bellman's "curse of dimensionality."

Recent advances in computational methods for discrete-choice models. From the preceding sections it is evident that the principal trade-off involved in structural estimation is that of computational simplicity vs. realism of the model. This trade-off exists because tractable closed-form value functions or optimal policy functions are unavailable for many interesting dynamic discrete- choice models. This requires numerical solutions of the dynamic program instead, which leaves one open to the curse of dimensionality. Some progress around this problem has been made, and we examine several recent proposals that we view as critical for this effort.

The first of the methodologies below relies on simplifying the calculation of the conditional expectation of the conditional expectation $E_t V(M_{t+1}; t+1)$, which appears in the value function recursion (6) or (17). Keane and Wolpin (1994) refer to this as the "EMAX" function, since with $k=1,\ldots, K$ alternative choices one can always write

$$E_t V(M_{t+1}; t+1) = E_t \max_k V_k(M_{t+1}; t+1). \tag{18}$$

Here $k = 1, \ldots, K$ indexes the discrete-valued and mutually exclusive alternatives the individual chooses from. (In Wolpin's model above, k takes two values: $k=0$ if the individual chooses not to have a child and $k=1$ otherwise.) $V_k(M_{t+1}; t+1)$ is the value of remaining lifetime utility as of t+1, assuming the individual chooses alternative k at time t, and behaves optimally thereafter. (In our version of Wolpin's model, $V_k(M_{t+1}; t+1) = V(M_t+k; t+1)$.) Keane and Wolpin (1994) call these k functions the *alternative-specific value functions.* These functions satisfy

$$V_k(M_{t+1}; t+1) = U_k(\cdot; M_{t+1}) + \beta E_{t+1} V(M_{t+2}; t+2); k = 1, \ldots, K. \tag{19}$$

The first (and perhaps most well-known) of the solution methodologies for (19) is Rust (1987). Rust achieved computational simplifications of the EMAX function at the cost of some realism by assuming that: (i) individuals' utility functions display additive separability of deterministic and stochastic components, (ii) given the current level of observable state variables, the stochastic components are serially independent, (iii) the stochastic components have a multivariate extreme value distribution. These given, Rust's two key results were: (i) the EMAX function has a closed-form solution, and (ii) the choice probabilities $\Pr[k|M_t]$ are multinomial logit. As noted in Keane and Wolpin (1994), these results together allow the analyst to avoid costly numerical integrations in solving the dynamic program, and in performing the likelihood estimation.

These assumptions were quite restrictive and ruled out more general models in which the errors display serial correlation or do not follow the extreme value distribution. Since then, several methods have been proposed to relax these restrictions. Rust (1994), in particular, reviews available solution methodologies for the EMAX function in the (realistic) case when the underlying random process is Markovian. For Markovian processes, these include the method of successive approximations, policy iteration methods, or minimum weighted residual methods. He then lays out a general estimation theory for estimates based on these approaches.

Rust also surveys recent results that have been obtained for more general underlying random processes. For these, Monte Carlo integration is the natural solution method, but it can be very computationally intensive. To lower computational costs Keane and Wolpin (1994) and Geweke, Slonim, and Zarkin (1992) propose combining Monte Carlo integration with interpolations/approximations of the value function (6) or the optimal policy function (7).

Keane and Wolpin (1994) rely on the value function recursion (6) or (19) as a basis for estimation, and try to find a form for the EMAX function of the recursion that can be computed more easily, *à la* Rust (1987). However, instead of deriving an exact form for EMAX, Keane and Wolpin propose approximating functions which are arbitrarily close to the true EMAX function $E_t V(M_{t+1};\, t+1)$. The idea is as follows: (i) first solve the dynamic program backwards by simulation methods for a reasonably-sized subset of state points $\{\tilde{M}_t\}_{t=1}^{T}$, (ii) next, using the simulated values of the value function V, fit by regression methods an approximating function to the EMAX function, (iii) use the fitted approximation function to interpolate for EMAX values on the remaining state points, (iv) having calculated all the relevant EMAX values, calculate the value of likelihood function at a prespecified value of α. Then iterate (i)–(iv) on α until the value of the likelihood function is at a maximum.

The general form of the approximating function proposed by Keane and Wolpin is

$$EMAX\,(M_t,\, t) = MAXE\,(M_t,\, t) + g\,(MAXE\,(M_t,\, t) - \bar{V}_k)\,, \qquad (20)$$

where $\text{MAXE}(M_t, t) = \max_k \bar{V}_k$, $\bar{V}_k = E_t V_k(M_{t+1}; t+1)$. The difference $\text{MAXE}(M_t, t) - \bar{V}_k$ inside $g(\cdot)$ is understood here to be a k-vector. The function g is a mapping from $\mathbb{R}^k$ to $\mathbb{R}_{++}$, that is, g is real-valued and always positive. The intuition for this form is that the difference beween EMAX and MAXE, which is always positive, will depend on how far apart are the expected V_k's from each other, and this distance is captured by the difference $\text{MAXE}(M_t, t) - \bar{V}_k$. By increasing the number of state space points used (relative to the size of the state space) in estimating the interpolating function, one can get arbitrarily good approximations to EMAX.

In Monte Carlo experiments, Keane and Wolpin (1994) find that the following form of g worked well:

$$\begin{aligned} EMAX\,(M_t,\, t) - MAXE\,(M_t,\, t) &= \delta_0 + \sum_{j=1}^{K} \delta_{1j}\,(MAXE - \bar{V}_j) \\ &\quad + \sum_{j=1}^{K} \delta_{2j}\,(MAXE - \bar{V}_j)^{1/2} \qquad (21) \end{aligned}$$

By simulating the dynamic program at the subset of state points $\{\tilde{M}_t\}_{t=1}^{T}$ one gets values for both MAXE and $\bar{V}_j$, as well as $\text{EMAX}(M_t, t)$. These values become the data for estimating δ_0, δ_{1j}, δ_{2j} in (21) by linear regression methods. The fitted version of (21) serves as the interpolating function for calculating $\text{EMAX}(M_t, t)$ at the remaining state points. The computational gain is that it is much easier to evaluate MAXE or any of the $\bar{V}_k$'s for the remaining state space points than it is to calculate EMAX.

Geweke et al. also propose an estimation-interpolation procedure that is very similar in spirit to that of Keane and Wolpin. The main difference is that, in this case, the authors work with approximations to the optimal policy function rather than the value function V. In the Wolpin (1984) model, the optimal policy function is the step function (10) which depends on the unknown function $J_t = E_t[V(M_t+1;\, t+1) - V(M_t; t+1)] + P_t\theta_t$. In the context of that example, Geweke et al. propose approximating J_t by the translated logistic of a polynomial-in-M_t:

$$\tilde{J}_t = exp\,(r^p\,(M_t)'\,\delta)/[1 + exp\,(r^p\,(M_t)'\,\delta)] - \frac{1}{2}\ , \qquad (22)$$

where $r^p(M_t)$ is a vector of variables of the form $\prod_{i=1}^{p}(M_{it})^p$ so that $(r^p(M_t)'\delta)$ is a pth-order polynomial function in the state M_t, with coefficients δ. They show that this approximation can be made arbitrarily good by increasing the polynomial order p. To estimate δ, the authors suggest maximum likelihood estimation of the logit model

$$Pr\,[N_t = 1|M - t] = exp\,(r^p\,(M_t)'\,\delta)/[1 + exp\,(r^p\,(M_t)'\,\delta)] \qquad (23)$$

based on R simulations of $\{M_t\}_{t=1}^{T}$ and S possible values of the choice variable. Estimates $\hat{\gamma}$ can now be used to construct the approximate decision rule J

$$\hat{J}_t = exp\,(r^p\,(M_t)'\hat{\delta})/[1 + exp\,(r^p\,(M_t)'\hat{\delta})] - \frac{1}{2}\ . \qquad (24)$$

Using this decision rule in place of the function J_t (evaluation of which requires numerical integration) greatly facilitates further simulations of M_t, as well as subsequent calculations of the values of the choice variables that solve the dynamic program.

Rust (1995) proposes a more promising alternative to Keane-Wolpin which is based on a random Bellman operator. His approach is appealing in that it avoids the need for interpolation and repeated simulation, and in fact breaks the exponential relationship between computation time and dimensionality of the state variable.

Perhaps even more promising is the simulation estimator of Hotz et al. (1994), a technique which has the added advantage of being applicable to more general random processes than Markovian ones. The key discovery of Hotz et al. was that under fairly general conditions, one can "invert" nonparametric estimates of conditional choice probabilities (i.e., the densities $f(\varepsilon_{t+1}|M_t)$ in (16)) to get consistent estimates of the value function or EMAX function in a *normalized* form. This "nonparametrically estimated" normalized value function can then be used to calculate consistent estimates of the optimal decision rule or policy. The estimated decision rule, along with the data on the model variables and the conditional probabilities f, can be combined together in a simulation to trace out a "simulated" normalized value function. Estimation of the parameter vector α is finally done by choosing α to minimize the distance between points on the simulated value function and the "nonparametrically estimated" value function. See Rust (1994) for a more detailed review of this new technique.

Empirical results. With regard to the empirical findings on fertility behavior arising from structural models, we review the results in Wolpin (1984), as this is more or less a representative fertility model. Other models have very unique features and are interesting in their own right, but for space considerations are only referenced here. For instance, Vijverberg (1984) has a labor-supply and time-allocation decision in addition to the fertility choice. Hotz and Miller (1993) study a model with imperfect fertility control by contraception, but with perfect fertility control by irreversible sterilization. Necessarily, their results differ.

Wolpin applied his estimation strategy to a 188-household subsample of households in the 1976 Malaysian Family Life Survey, which contains a retrospective life history for each household, with birth, infant mortality, and household income. Wolpin then matched these up with Malaysian state-specific survival rates. For his application, Wolpin set the number of time periods to 20, each period being an 18-month interval. Time period 1 was set to the onset of individual's fertile cycle (age 15 or marriage, whichever came first), and at the last decision period ($t = 20$) the fertile cycle was assumed to end, with individuals assumed to live for 10 periods (15 years) thereafter.

After testing for and finding no strong evidence of unobservable individual heterogeneity at work, Wolpin estimated the parameters of his quadratic utility function (see Table 1) and the parameters of its budget constraint (consisting of cost parameters for child maintenance costs, and parameters for the cost of new births). Point estimates of the utility function parameters were reasonable – at point estimates marginal utility was positive and diminishing in family size and in the consumption good, and time preference was about 0.09 per annum. A model check using a likelihood ratio test revealed a significant enough difference between his model's results and those from a model of random (Bernoulli) births. So the Wolpin structure explained more than by pure chance.

Interestingly, Wolpin found that children and consumption goods tended to be gross substitutes in utility. The higher the mother's education levels, however, the lower was the utility of additional children. With respect to the cost function parameters, Wolpin found large costs of new births, which followed a decreasing then increasing pattern over the life-cycle. Maintenance costs, however, appeared to be small and reasonable. With respect to income, a rise in husband's current income tended to have a large positive effect on the number of births, with larger income elasticities of a birth appearing at the higher income extremes. The expected number of children ever born did not vary much with the realization of an infant death, so estimated replacement behavior was extremely small (an infant death inducing a rise in children ever born by an average of only 0.015). However, the response to higher probabilities of infant mortality risks was not small. Indeed, Wolpin found that a fall in the mortaility probability by only 0.05 percentage points reduced the number of children ever born by about one-quarter.

Regarding the implied comparative dynamics, Wolpin found that individuals who over time expected rising incomes and falling mortality risks tended to delay births. This finding is consistent with predictions of Heckman and Willis (1976). Further, a lowering of mortality risks generated a tendency to cluster births in early periods of life, while a rising survival probability profile tended to delay childbearing.

In short, Wolpin found: (i) small income effects; (ii) large and negative education effects; (iii) very small positive replacement effects, and a non-monotonic relationship between current birth probabilities and family size; (iv) no unobservable heterogeneity among individuals, and (v) a negative effect on fertility of higher mortality risk.

Of interest is a comparison of these five results against results that could be obtained from a typical static reduced-form model. As a final exercise, Wolpin did a probit estimation of the birth likelihood taking family size, current and expected income, current and expected infant survival probabilities, and mother's schooling as regressors. The comparative-static results were as follows. Like structural estimates, the fitted probit regressions implied: (i) negligible income effects on birth probabilities, and (ii) negative education effects. Unlike the structural estimates, however, the nonstructural probit model produced (iii) an insignificant or positive effect of family size on current birth decisions, the positive effect suggesting (iv) potentially unobserved heterogeneity; and (v) a positive effect on fertility of higher mortality risk. Evidently, the two approaches see the relationship between births and fertility variates differently. It remains an open question as to which approach gives the more believable results.

The empirical properties of Wolpin's (1984) model indicate that a relatively simple version of the dynamic programming model (1)–(6) is capable of replicating some fairly complex properties of birth history data. While the computational complexity of the estimation strategy is an issue, it is fair to ask how complex a comparable non-structural alternative would have to be to replicate all the possible birth patterns in the data. Wolpin figured that, to do this, a comparable nonstructural econometric model would require the estimation of around 400 parameters, instead of the 13 of his structural model. Computationally, such a model would also be fairly demanding. Given the recent advancements in the computation of structural estimates for even richer models, there appears to be an expanding empirical role for structural dynamic models. Nonetheless, dynamic reduced-form models continue to be useful and are usually more tractable alternatives. We now turn to these models.

3. Reduced-form models

In this section we consider the class of dynamic reduced-form fertility models, with focus on research by economists. As mentioned in the introduction, reduced-form models may have a basis in some dynamic programming problems but do not rely heavily on that structure for the specification of estimating equations. Specification of the estimating relationships, including stochastic elements, is often made in conformity with some tractable econometric framework; these models typically use the underlying theory to posit variables and possible error structures.

In most of these studies the main variable of interest is the birth likelihood. When this variable is regarded as continuously-varying, practically all reduced-form dynamic fertility models adopt a hazard-rate approach to estimation. When the birth probability (and the contraceptive efficiency) takes on discrete (dichotomous) values, the empirical methodology of Hotz and Miller (1988) is a viable alternative.

3.1 Hazard-rate analysis (Heckman and Walker, etc.)

Formulation and estimation. Hazard-rate analysis, also often called duration or event history analsys, is proving useful in a variety of problems in economics.[25] Indeed, most economic research on timing and spacing of births has used the hazard approach. Specifically, this framework is followed in Newman (1983); Heckman et al. (1985); Newman and McCullough (1984); Heckman and Walker (1987, 1989, 1990 a, b, 1991); and David and Mroz (1989). In this approach, women are assumed to be continuously subject to the risk of a birth. The risk is given by the hazard rate, defined as the conditional probability of a birth at time t given no birth immediately before t. The hazard rate may vary randomly across the population (referred to as *heterogeneity*) and may vary over time spent in a birth interval (referred to as *duration dependence*).

In a series of papers, Heckman and Walker present semiparametric multispell fertility models. Their models are first discussed here because they are more general than others. It will become clear that models used in Newman (1983), Heckman et al. (1985), and Newman and McCullough (1984) can be considered as special cases. Following the notation of Heckman and Walker, a woman's birth history is assumed to evolve in the following way. The woman becomes at risk for the first birth at calendar time $\tau=0$. Define a finite-state continuous-time birth process $\{Y(\tau),\ \tau>0\}$, $Y(\tau)\in\Gamma$, where the set of possible attained birth states (parities) is finite ($\Gamma=\{0, 1, 2, \ldots, c\}$, $c<\infty$). An element of Γ defines the number of children born. $Y(\tau)$ is parity attained at τ. Transitions occur on or after $\tau=0$. Child mortality is not considered.

The basic component for multistage duration models is the conditional hazard. Define $H(\tau)$ as the relevant conditioning set at time τ, which may include anticipitations about the future formed at time τ and relevant past information up to time τ (previous birth intervals, etc.).

Define the potential durations by $T_1, \ldots, T_c$. If a woman becomes at risk for the jth birth at time $\tau(j-1)$, the conditional hazard at duration t_j is defined to be

$$h_j\left(t_j|H\left(\tau\left(j-1\right)+t_j\right)\right). \tag{25}$$

Assuming that T_j is absolutely continuous given H, we may integrate (25) to find the survivor function

$$S\left(t_j|H\left(\tau\left(j-1\right)+t_j\right)\right)=exp\left[-\int_0^{t_j} h_j\left(u|H\left(\tau\left(j-1\right)+u\right)\right)du\right]. \tag{26}$$

The assumption of absolute continuity rules out conditioning on variables that perfectly predict the fertility outcome.

A woman at risk for a first birth at $\tau=0$ continues childless a random length of time governed by the survivor function

$$Pr\left(T_1 > t_1 | H\left(\tau\left(0\right) + t_1\right)\right) = exp\left[-\int_0^{t_1} h_j(u|H(\tau(k-1)+u))\,du\right]. \qquad (27)$$

At time $\tau(1)$, the woman conceives and moves to the state $Y(\tau)=1$. In the general case, $Y(\tau)=k–1$ for $\tau(k–1) \le \tau < \tau(k)$ and $T_k = \tau(k)–\tau(k–1)$ is governed by the conditional survivor function

$$Pr\left(T_k > t_k | H(\tau(k-1)+t_k)\right) = exp\left[-\int_0^{t_k} h_k(u|H(\tau(k-1)+u))\,du\right]. \qquad (28)$$

The conditional density function of duration $T_k = t_k$ is given by the product of the hazard and survivor functions

$$g(t_k|H(\tau(k–1)+t_k)) = h_k(t_k|H(\tau(k–1)+t_k)) \cdot S(t_k|H(\tau(k–1)+t_k)). \quad (29)$$

Modelling unobserved heterogeneity (e.g., fecundity) is necessary because it is virtually impossible to measure all important covariates appropriately.[26] Recent work shows the importance of controlling for heterogeneity in hazard models (Heckman and Walker 1985; Heckman et al. 1985; Trussell and Richards 1985; Manton et al. 1986; Struthers and Kalbfleisch 1986; Sturm and Zhang 1993; Zhang and Sturm 1994). Heckman and Walker distinguish two types of unobservables: (i) those that are known to the woman being studied but unknown to the researcher, and (ii) those that are not known to both the woman and the researcher. They point out that the latter type of heterogeneity can produce dynamics of its own if the agents being studied learn about their unobservables over the life cycle, and provides the rationale for including lagged birth intervals as explanatory variables as in Rodriguez et al. (1984).

Heckman and Walker point out that the study of heterogeneity in multistage duration models is in its infancy. The few studies that include unobserved heterogeneity consider the first type and universally assume the heterogeneity can be represented by a scalar random variable θ which is time-invariant with distribution $M(\theta)$. Densities are now defined conditional on $H(\tau)$ and θ:

$$\begin{aligned} g\left(t_k|H(\tau(k-1)+t_k);\,\theta\right) = {} & h_k(t_k|H(\tau(k-1)+t_k);\,\theta) \\ & \cdot S(t_k|H(\tau(k-1)+t_k);\,\theta)\,. \end{aligned} \qquad (30)$$

The conditional density of $T_1, \ldots, T_c$ given $H(\tau(0) + \sum_{i=1}^{c} t_i)$ is then

$$g\left[t_1, \ldots, t_c | H\left(\tau(0) + \sum_{i=1}^{c} t_i\right)\right] = \int_{\underline{\theta}} \prod_{k=1}^{c} g\left(t_k|H(\tau(k-1)+t_k);\theta\right) dM\left(\theta\right) \qquad (31)$$

where $\underline{\theta}$ is the support of θ, i.e., its domain of definition.

In their empirical specification, Heckman and Walker approximate the jth conditional hazard using the following functional form

$$h_j(t_j|H(\tau(j-1)+t_j);\,\theta) = exp\left[\gamma_{0j} + \sum_{k=1}^{K_j}\gamma_{kj}\left(\frac{t_j^{\lambda_{kj}}-1}{\lambda_{kj}}\right) + \boldsymbol{Z}\beta_j + f_j\theta\right], \tag{32}$$

where $\boldsymbol{Z}$ includes all observed regressors possibly including durations from previous spells and spline functions of current durations. Parity dependence is incorporated by allowing coefficients to bear parity-specific subscripts.

An important feature of hazard specification (32) is that it encompasses a variety of widely-used models. Setting $\beta_j=0$, $K_j=1$, and $f_j=0$, (32) specializes to a Weibull model if $\lambda_{1j}=0$, to a Gompertz hazard if $\lambda_{1j}=1$, and to a quadratic model if $K_j=2$ and $\lambda_{1j}=1$ and $\lambda_{2j}=2$. It becomes an exponential model if $\gamma_{kj}=0$ for all k. This general framework allows us to use likelihood ratio tests to choose among many conventional competing specifications. Specification (32) also extends previous duration models by allowing for general time-varying covariates and by introducing unobserved heterogeneity that is correlated across spells.[27] Permitting the f_j to vary by parity j allows the scalar unobservable to play a different role in different spells.

The conventional way to include unobserved heterogeneity is to assume a parametric function for $M(\theta)$, as in Newman and McCullough (1984). Such procedures have been called into question because of the sensitivity of the estimated parameters to choices of functional forms for the hazard and heterogeneity. Heckman and Singer (1984) show that $M(\theta)$ can be estimated nonparametrically. However, Trussell and Richards (1985) demonstrate nonrobustness of estimates to choices of functional forms for hazard even when $M(\theta)$ is estimated nonparametrically. These sensitivity results are widely cited as "discouraging" evidence on the value of incorporating unobserved heterogeneity into hazard models. Montgomery and Trussell (1985, p. 30) capture this negative mood in the demographic literature:

> This sensitivity to the choice of functional forms of the distribution of unobservables or of the hazard leaves us profoundly depressed about where next to proceed. We fear that the advances in statistical technique have far outpaced our ability to collect data and our understanding of the behavioral and biological processes of interest.

Heckman and Walker (1987) argue that missing from the pessimistic assessments of models with unobservables is any discussion of the fit of alternative models to the data. They demonstrate that, in contrast to the presumption in the demographic literature that preferred alternative models fit the data equally well, few models can fit the data at all. Because many models are non-nested, we confront the problem of non-nested model selection when seeking a "best" model. Heckman and Walker (1990 a, 1991) propose and describe four widely used criteria: (i) the Leamer-Schwarz criterion that penalizes likelihood for parametric estimation; (ii) the criterion of picking a model on the basis of coefficient stability across cohorts; (iii) the effectiveness of fitted micro-models in predicting aggregate time series; and (iv) a criterion widely used by demographers – predicting fertility attained (parity) at different ages.

In Heckman and Walker (1987, 1990 a, b, 1991) they estimate $M(\theta)$ by the nonparametric maximum likelihood (NPMLE) procedure of Heckman and Singer (1984). This procedure approximates any distribution functions of unobservables with a finite mixing distribution, $\{p_i, \theta_i\}_{i=1}$, where p_i is the weight placed on θ_i. The NPMLE procedure estimates support points $\theta_i, i=1,\dots,I$ and the weight placed on the support points $\left(p_i, i=1,\dots,I \ where \sum_{i=1}^{I} p_i = 1\right)$ along with other parameters of the model.

Finally, Heckman and Walker's empirical framework allows for period-specific stopping behavior. The survivor function for the jth birth is

$$S_j(t_j|H(\tau(j-1)+t_j);\theta) = P^{(j-1)} + (1-P^{(j-1)})\, exp\left[-\int_0^{t_j} h_j(u|H(\tau(j-1)+u);\theta)\,du\right], \qquad (33)$$

where $P^{(j-1)}$ is the probability that a woman with j–1 children is never at risk to have the jth birth and thus captures permanent biological or behavioral sterility (i.e. a parity-specific mover-stayer mixture distribution).[28] The contribution to sample likelihood of a woman with fertility history $T_1=t_1, T_2=t_2,\dots, T_k=t_k$ and an incomplete k+1 spell of length $\bar{t}_{k+1}$ is

$$\sum_{j=1}^{I}\prod_{j=1}^{k}\left[-\frac{\partial\, ln\, S_j(t_j|H(\tau(j-1)+t_j);\theta_i)}{\partial t_j}\right]\cdot S_j(t_j|H(\tau(j-1)+t_j);\theta_i) \cdot S_{k+1}(\bar{t}_{k+1}|H(\tau(j-1)+t_j);\theta_i)\,p_i\,. \qquad (34)$$

A general multistate computer program, CTM, applicable to multistate competing risks models much more general than a birth process, is used to estimate the model. The details of the computer program are in Yi, Walker, and Honore (1987) and Heckman and Walker (1987).

Other formulations can be seen as special cases of Heckman and Walker's formulation. Newman and McCulloch (1984) model duration dependence as a three point linear spline and assume a parametric Gamma distribution for unobserved heterogeneity for $M(\theta)$. Heckman et al. (1985) is almost identical to the general formulation of Heckman and Walker, except that the former study models duration dependence as a quadratic polynomial and the latter uses a general Box-Cox transformation. All these studies allow for time-varying covariates. Allowing for period-specific stopping behavior is an unique feature of the general Heckman-Walker formulation.

David and Mroz's (1989) empirical model is similar to the general Heckman-Walker framework in several aspects. David and Mroz model unobserved heterogeneity and period-specific stopping behavior (called as secondary sterility) as in Heckman and Walker. However, there are several differences in specifications. Child mortality is ignored in Heckman and Walker, but is modeled as an independent random censoring of the birth process. In another word, they define the duration as the minimum of the waiting time to a live birth and the waiting time to the death of the youngest child. Unlike Heckman and Walker, David and Mroz assume a log-logistic function for the conditional hazard. (See Kalbfleisch and Prentice (1980) for a discussion of the relation between log-logistic functions and other functions such as Weibull.) As a minor difference, David and Mroz allow the weight placed on the support points (p_i, $i = 1,..., I$) to be determined by a multinominal logistic function of some observable characteristics.

Empirical results. Table 2 summarizes the empirical results in Newman and McCullouch (1984), Heckman et al. (1985), David and Mroz (1989), Heckman and Walker (1990 a, b, 1991), and Tasiran (1995). Newman and McCullouch estimated their hazard model with Gamma heterogeneity for birth history up to age 30 for four cohorts of women from ages 30 to 49 at five-year intervals. It is found that higher female education delays the first birth of all birth cohorts and delays subsequent births of all cohorts except the birth cohort 40–44.[29] While the education of the male is not (statistically) significant in determining the risk of the first birth, it is important in delaying subsequent births. Women in younger cohorts living in areas with higher child mortality tend to have the first and subsequent births earlier. The family planning variable is never significant at a conventional level.

Heckman et al. find that some “stylized facts” of the demographic literature are not robust to unobserved heterogeneity. Rodriguez et al. (1984) suggest that age at marriage and/or the entry into parenthood are the crucial determinant of life cycle fertility and that variation in completed fertility across the population comes primarily from these initial conditions, and that subsequent childbearing is largely determined by initial events and by previous birth intervals. Using the Swedish data, Heckman et al. (1985) find that, in models that do not control for heterogeneity, the longer a preceding birth interval the longer the subsequent one, exhibiting the “engine of fertility” phenomenon noted by demographers.

Controlling for heterogeneity, the “well-noted empirical regularity” either vanishes or reverses in sign. For a sample of married women, it vanishes entirely. Furthermore, for all women and controlling for marital status as a covariate, inclusion of heterogeneity produces a “reverse engine of fertility” phenomenon: the longer the preceding birth interval the shorter the subsequent one. For the married women sample, the importance of age at marriage on the spacing of births is considerably reduced in size and statistical significance. For the sample of all women, controlling for heterogeneity eliminates the effect of age at marriage on all but the final birth transition.

Emphasizing the roles of female wages and male income, Heckman and Walker (1990 a, b, 1991) find that female wages delay times to all conceptions and reduce total conception. This result is shown to be robust to a variety of empirical specifications. Higher male income reduces times to conceptions (strongest effect for the first birth) and increases total conceptions when marital status is not controlled for. The estimated male income effect is weaker when marital status is included as a separate regressor. Somewhat surprisingly, unobserved heterogeneity correlated across spells is not an important

Table 2. Empirical results from reduced-form models

Authors (Year)	Emphasized (or primary) covariates	Other or additional covariates	Dataset
Heckman, Hotz and Walker (1985)	Age at marriage, previous birth intervals	Current spell duration, labor participation, schooling, urban, white-collar, attended university	1981 Swedish Fertility Survey
Heckman and Walker (1990 a, b, 1991)	Female wage and male income	Current spell duration, urban, white-collar, age, time trend, attended university, ever-married, unemployment rate, policy measures	1981 Swedish Fertility Survey
Tasiran (1995)	Female wage, male income, female schooling, and working experience	Current spell duration, urban, white, age at union start, birth cohorts, ever-married	1981 Swedish Fertility Survey, Swedish 1984 and 1988 Household Market and Non-Market Activities, 1985–1988 PSID
Newman and McCulloch (1984)	Male and female years of schooling	Female's birth year, child mortality, family planning measure	1976 Costa Rica National Fertility Survey
David and Mroz (1989)	Age of husband and wife, previous child's sex, number of boys and girls alive, number of boys and girls alive aged 10 or older	Number of prior deaths of boys and girls before the 3rd month, number of prior deaths of boys and girls aged 3–11 months, village mortality	Rural France data relating to the marriage cohorts of 1749–1789

feature of modern Swedish fertility data.[30] In all models in which nonzero duration dependence is permitted, they find positive duration dependence. Their major finding is that a neoclassical model of fertility outperforms demographic models which exclude wage and income variables in terms of several model selection criteria.

As Heckman and Walker point out, the most controversial aspect of their study is its use of cohort average wages as proxies for missing micro wages. While the use of average wages almost eliminates the possibility of simultaneous equation bias in estimating the relation between wages and fertility, it introduces an errors-in-variables problem and a spurious relationship between strongly time-trended variables (i.e., wages and fertility). Heckman and Walker demonstrate that the results on female wages and male income are robust to those considerations.

Heckman and Walker (1989) use estimates of individual fertility dynamics obtained in Heckman and Walker (1990 a, b, 1991) to forecast aggregate annual birth rates for each cohort. They examine the ability of microdynamic models to account for time series variation in cohort fertility. They find that, for most Swedish women, their estimated models pass some important time series specification tests. They also find that their aggregated neoclassical microdynamic models explain the time series better, in the sense of mean squared error of forecast, than do time series regressions.

Building on Heckman and Walker's work, Tasiran (1995) provides new evidence for the role of female wages and male income in the fertility of Swedish women. Tasiran modified the macro wage and income series used by Heckman and Walker, and made use of micro wage and income data in individual income-tax records from Statistics Sweden. He constructed two combined macro-micro wage series. For the first combined series, he used Heckman and Walker's series for the years 1948–1967 and 1981, and for the period 1968–1980 the micro wage data derived from income-tax records. For the second combined series he used wages of salespersons and shop assistants for the years 1948–1967 and 1981, and for the period 1968–1980 the micro wage data. He also included a variable for women's years of working experience.

Tasiran reports that the Heckman-Walker results of a negative wage effect and a positive income effect do not hold generally. The sign and statistical significance of female wages and male income are sensitive to parity level, measurement of wages and income, controlling for unobserved heterogeneity, inclusion of working experience, and data sets. Tasiran's main conclusion is that the effects of female and income are not robust and are much weaker than Heckman-Walker indicated. He also finds that years of female schooling and of working experience significantly delays the first birth but have weaker effects on higher births.

In a reply to Tasiran (1995), Walker (1996) criticizes Tasiran's casual use of micro wages. Specifically, Walker points out that a more careful analysis would have addressed four issues on the use of individual wages. First, he argues that a simple merging of aggregate and micro data is inappropriate. Second, the issue of wages of nonworking women was not addressed by Tasiran. Third (and most importantly, according to Walker), inadequate attention was paid to measurement error in individual wages. Finally, exogeneity of wages and selection bias should be addressed.

The main finding of Walker's thoughtful analysis is that there is substantial measurement error in individual wages in birth years. Controlling for measurement error by dropping birth year observations, the selection effect appears to be minimal. Estimates from a wage regression (excluding birth years observations) were then used to backcast female wages before 1968 and to fill in wages for nonworking women. As Walker finds,

> For the first transition the estimated wage effects are roughly two-thirds the magnitude of the effects estimated by aggregate wages. Even in the higher parities, the estimated wage effects are quantitatively large and statistically significant. Although the cohort drift which was present in the estimates obtained from aggregate data is not present, the statistical significance and importance remain. (Walker 1996)

Replying to Walker's comment, Tasiran (1996) writes

> The results in rows 5 of Walker's Table 3 is an excellent illustration of the sensitivity of the estimates of the wage rate effects on fertility and that the first Heckman and Walker estimates were on the high side. Comparing rows 5 with rows 3, 11 out of 12 estimates in rows 5 are less in absolute value than the corresponding estimates in rows 3 and several are substantially lower.

Walker's analysis of the individual wages certainly highlights the importance of careful and thoughtful work on micro data. His findings also lend support for the results of Heckman and Walker obtained by aggregate data in wages. The debate between Walker and Tasiran, however, is valuable, as it illuminates the complicated fertility dynamics in Sweden. The central problem is a lack of high quality micro wage data. Given this problem, it is not surprising to find different results when individual wages are constructed and handled differently. The issue of robustness of results should be dealt with carefully. When magnitude differs across different wage series, Tasiran appears to interpret the results to be sensitive. Walker, on the other hand, believes that as long as statistical significance is preserved, the results are robust. Interpreted with care, Walker's new results show that the original Heckman-Walker findings are, to some extent, robust.

We believe that there are still some issues that require further investigation. First, Walker's analysis shows that there is substantial measurement error in wages for birth year observations. The question is: why? Tasiran (1996) thinks that the measurement error is mainly in working hours which he obtained from Heckman and Walker. Rather than simply deleting birth years observations (when estimating the wage regression) and thus losing useful information, a reexamination of the data on earnings and working hours might be necessary to determine the sources of, and possible remedies for, measurement errors in micro wages.

Second, Walker (1996) also used the female wage series for shop assistants and found the wage effects were small for younger cohorts. Walker argues that the inability to recover statistically significant wage effects for the youngest cohorts stems from the flatter age-profile of wages during a period of low wage growth. Specifically, he mentions that the observed decline in the wage dispersion during the 1970s reflects the compression arising from the adoption of wage solidarity. If that was the case, how can the use of aggregate macro wages show statistically significant wage effects in Heckman and Walker? That would be possible if there was more variation in macro wages than in female shop assistant wages. But why should the slow-wage-growth *cum* flatter-age-profile argument apply to female shop assistants, but not to macro wages? Therefore it appears necessary to investigate whether the age-profile of wages is flatter for shop assistants than for other workers.

David and Mroz (1989) estimate their model using data from rural France relating to the marriage cohorts of 1749–1789, a population that is supposed by many demographers to be identified with "natural fertility", namely, the absence of parity dependence in age-specific marital fertility rates. Their central concern is to re-examine this widespread "natural fertility" supposition with suitable econometric methods. They obtain strong evidence against the "natural fertility" characterization of the demographic regime that prevailed in rural France immediately before the Revolution of 1789.

There are clear indications that marital fertility rates were being regulated in congruence with the differential valuations placed upon children, according to their gender and age. In particular, controlling for the total numbers of young children and old children, the presence of a larger number of young girls leads to a lower hazard rate, but that of a larger number of young boys has no such effect. A larger number of older children, regardless of gender, reduces the hazard rate. They attribute the differential response by age of children to risk aversion and the greater risk of losing a younger boy to mortality. Village mortality is found to have a positive effect on the hazard, suggesting the presence of hoarding behavior in fertility. They also find strong replacement effects to the deaths of boys (especially boys aged two months or above). To the contrary, there is weak evidence that the death of young girls reduces the hazard rate. They argue that these results suggest the endogeneity of child mortality.

3.2 Quasi-maximum likelihood estimates of linear decision rules

When the contraceptive efficiency variable takes on a few discrete values (say $\underline{u}$ or $\overline{u}$), an estimation strategy can be based on binary linear decision rules that approximate the structural model's exact optimality conditions. As the exact dynamic decision rules are generally nonlinear and may lack closed-forms, this approach offers a reasonable operationalization of an intractable theoretical specification. In what follows we outline a much simplified version of an approach due to Hotz and Miller (1988). Our simplifications highlight their basic ideas as applied to the contraceptive decision rule only. For a more complete model that features joint estimation with labor supply and child care decision rules, the interested reader may consult the original paper.

Formulation and estimation. Hotz and Miller (1988) assume that the contraceptive decisions and the birth hazard are guided by the values of an index function q_{it} that is linear in several underlying causal variables. Let $a_{it}=p_{it}^M M_{it}$ be the current expenditure of household i at time t for a family size of M_{it}. Further, let c_{it} be the amount of time committed to child-rearing in period t. Noting that t indexes the mother's age and that the mother's entire birth history is given by the sequence of binary numbers $\{N_{it-k}\}_{k=1}^{t}$, Hotz and Miller now specify q_{it} as

$$q_{it} = v_0 + v_1 I_{it} + v_2 c_{it} + v_3 a_{it} + \sum_{k=1}^{t} v_{4k} N_{it-k} + v_5 t_i + v_6 \theta_i + \varepsilon_{it}, \qquad (35)$$

where v_j, $j=1,...,6$ are linear coefficients, some of which can be separately identified and estimated. The level of the index q_{it} is thus assumed to be function of income I_{it}, maternal time inputs c_{it}, expenditure on market goods a_{it}, the mother's birth history, the mother's age t, plus a set of fixed effects parameters θ_i, and an error ε_{it} assumed $N(0,\sigma^2)$. This index determines the level of the birth hazard, which is a dichotomous variable in Hotz and Miller's (1988) framework:

$$\pi_{it}\begin{cases} = & \bar{\pi} \text{ } if \text{ } q_{it} \geq 0 \\ = & \underline{\pi} \text{ } if \text{ } q_{it} < 0, \end{cases} \tag{36}$$

where $\bar{\pi}$ a*nd* $\underline{\pi}$ are, respectively, the upper and lower values of the birth hazard.[31]

Hotz and Miller posit that total expenditure a_{it}, and total maternal care time c_{it} are linear sums of the mother's birth history $\{N_{it-k}\}_{k=1}^{t}$:

$$\mathrm{a}_{it} = \sum_{k=1}^{t} \mathrm{a}_k N_{it-k} \tag{37}$$

$$c_{it} = \sum_{k=1}^{t} c_k N_{it-k}. \tag{38}$$

Using (37)–(38), one can rewrite the index function as

$$q_{it} = \tilde{v}_0 + \tilde{v}_1 I_{it} + \sum_{k=1}^{t} (\tilde{v}_2 c_k + \tilde{v}_3 \mathrm{a}_k + \tilde{v}_4 k) N_{it-k} + \tilde{v}_5 t_i + \tilde{v}_6 \theta_i + \tilde{\varepsilon}_{it}, \tag{39}$$

where $\tilde{\varepsilon}_{it}$ is now a $N(0,1)$ error and the parameter vector $\tilde{v}$ is just the normalization $\tilde{v} = \sqrt{\sigma}v$, where v is the vector of identifiable parameters (v_0, v_1, $\{v_2 c_k + v_3 a_k + v_4 k\}_{k=1}^{t}$, v_5, v_6) in the original index function (35).

This last equation is now estimable. One, however, does not observe the actual levels of the index q_{it}; what is observable is the sample of individual birth histories, and births could have happened whether behavior put an individual at the low birth risk $\underline{\pi}$ or the high birth risk $\bar{\pi}$. Thus, a probit-style approach that takes into account both possibilities is called for. To this end, define the probability of observing the ith individual's birth sequence over the sample period, conditional on period incomes, previous births, and the fixed effects θ_i. This is

$$\begin{aligned} &\prod_t Pr\left(N_{it} | I_{it}, \{N_{it-k}\}_{k=1}^{t}, \theta_i\right) \\ &= \prod_t \{N_{it}\left[\underline{\pi} + (\bar{\pi} - \underline{\pi}) Pr\left(q_{it} \geq 0 | I_{it}, \{N_{it-k}\}_{k=1}^{t}, \theta_i\right)\right] \\ &\quad + (1 - N_{it})\left[1 - \underline{\pi} - (\bar{\pi} - \underline{\pi}) Pr\left(q_{it} \geq 0 | I_{it}, \{N_{it-k}\}_{k=1}^{t}, \theta_i\right)\right]\} \\ &= \prod_t \{(1 - N_{it}) + (2 N_{it} - 1)\left[\underline{\pi} + (\bar{\pi} - \underline{\pi}) \Phi(\tilde{v}_{it})\right]\}, \end{aligned} \tag{40}$$

where Φ is the standard normal cumulative distribution function and the product $\prod_t$ ranges over the entire lifetime of the ith individual, that is t goes from $\underline{t}_i$ to t_i. Some explanation is necessary for the above derivations. In the second line of (40), the first term applies when $N_{it} = 1$, that is, a birth occurs in period t. The (conditional) likelihood of a birth having taken place at t is the sum of the baseline risk of a birth $\underline{\pi}$ (which the individual is exposed to regardless of the level of contraceptive efficiency) and the additional birth risk when the minimal level of contraception is employed. This additional risk is just the term $(\bar{\pi} - \underline{\pi})\, Pr\,(q_{it} \geq 0 | I_{it}, \{N_{it-k}\}_{k=1}^{t}\ \theta_i)$. Finally, the second component of the second line of (40) applies when $N_{it}=0$, that is no birth occurs. The probability of this event is just one minus the probability of a birth, which is the first expression in braces in (40). In the absence of any other equations to be jointly estimated, (the log of) equation (40) is the (log) likelihood function for the birth sample, and this can be estimated by a standard maximum-likelihood procedure.

In the case when there is more than one equation to be estimated, things become more complicated. In the original Hotz and Miller (1988) paper, for instance, the index function (39) is only one equation that needs to be estimated alongside similar linear decision rules for childcare time and the mother's labor market participation. As Hotz and Miller note, this creates two problems. First, since the system of equations generally possesses a non-recursive covariance structure, full-information maximum likelihood (FIML) estimation will be computationally burdensome, as it involves repeated calculation of bivariate (or trivariate) distribution functions involving almost all of the model's parameters. Secondly, since the amount of childcare time enters into both the labor supply decision rule and the index function, there are now cross-equation restrictions on the parameters of the latter two, which one would have to either impose in the estimations or test for.

In the face of these two problems of FIML estimation, Hotz and Miller (1988) propose the following alternative estimator. Let $L_{i1}\,(\underline{\pi}, \bar{\pi}, \tilde{v}, \theta_i)$ denote the log of (40) given the model parameters $(\underline{\pi}, \bar{\pi}, \tilde{v}, \theta_i)$, and suppose one could define similar *quasi*-likelihood functions L_{ip} for the other p decision rules for household i. Now form the following *quasi*-likelihood function

$$\begin{aligned} Q\,(\underline{\pi}, \bar{\pi}, \tilde{v}, \tilde{v}^{+}, \theta_i) &= \sum_i Q_i\,(\underline{\pi}, \bar{\pi}, \tilde{v}, \tilde{v}^{+}, \theta_i) \\ &= \sum_i \left[L_{i1}(\underline{\pi}, \bar{\pi}, \tilde{v}, \theta_i) + \sum_p L_{ip}(\underline{\pi}, \bar{\pi}, \tilde{v}, \tilde{v}^{+}, \theta_i) \right], \quad (41) \end{aligned}$$

where $\tilde{v}^{+}$ are any extra parameters appearing in the p equations L_p. *Given estimates for the fixed effects* $\{\theta_i\}$, denote the remaining parameters in L_{i1} as $\mathbf{z}^1 = (\underline{\pi}, \bar{\pi}, \tilde{v})$, and the remaining parameters in L_{ip} as $\mathbf{z} = (\mathbf{z}^1, \tilde{v}^{+})$. *Given* $\{\theta_i\}$, the parameter estimates that maximize the quasi-likelihood function Q are the solutions of

$$\sum_i \mathrm{m}_i(\mathbf{z}, \theta_i) = \sum_i \partial L_{i1}(\mathbf{z}^1, \theta_i)/\partial \mathbf{z}^1 + \sum_i \sum_p L_{ip}(\mathbf{z}, \theta_i)/\partial \mathbf{z} = 0\,. \qquad (42)$$

Equation (42) suggests that, given $\{\theta_i\}$, a consistent estimate $\hat{\mathbf{z}}$ would have to make $\sum_i \mathrm{m}_i\,(\mathbf{z}, \theta_i)$ as close to zero as possible. Such an estimate can be found by minimizing the following average of the *score functions* $m_i(\mathbf{z}, \theta_i)$:

$$J_I = \left[(1/I)\sum_i \mathrm{m}_i(\mathbf{z},\,\theta_i)\right]' W_I \left[(1/I)\sum_i \mathrm{m}_i(\mathbf{z},\,\theta_i)\right]. \tag{43}$$

The consistent estimate $\hat{\mathbf{z}}$ of z based on (43), is $\hat{\mathbf{z}} = \text{argmin } J_I$. Essentially, $\hat{\mathbf{z}}$ is the generalized method of moments (GMM) estimator for the set of I orthogonality conditions $m_i(\mathbf{z},\,\theta_i) = 0$ (Hansen 1982). Here I is the number of individuals in the sample, and W_I is a symmetric, positive definite weighting matrix that converges, with large I, to a limit matrix W. W_I is used to construct an estimated asymptotic covariance matrix for $\mathbf{z}$.

As Hotz and Miller (1988) show, this quasi-maximum likelihood procedure avoids the computational difficulties of FIML on the joint system of decision rules, since it requires only calculation of an univariate normal distribution function. Further, one can impose the cross-equation parameter restrictions directly in the definition of the score functions m_i prior to GMM estimation. One can use (43) to get $\hat{\mathbf{z}}$ given an initial estimate $\{\hat{\theta}_i\}$. Using (43) again, one can then obtain a new estimate of $\{\hat{\theta}_i\}$ given the first-round estimate $\hat{\mathbf{z}}$, and continue to iterate on the estimate of $\{\theta_i\}$ and $\mathbf{z}$. The only issue that remains is obtaining initial estimates of the fixed effects $\{\theta_i\}$. Hotz and Miller (1988) obtain initial fixed effects estimates by maximizing the sum of the other quasi-log-likelihoods, $\sum_i \sum_p L_{ip}(\mathbf{z},\,\theta_i)$ with no parametric restrictions. The $\{\hat{\theta}_i\}$ were then used in (43) to estimate $\mathbf{z}$ imposing all cross-equation parameter restrictions.

4. Other models and approaches

To close our survey, we briefly discuss other dynamic fertility models which we could not neatly categorize as “structural” or “reduced-form”. Generally, this was because they were either (i) not strongly motivated by a theoretical dynamic model, or (ii) not concerned with the estimation of the model’s dynamic equations *per se*. At some level, however, this was a matter of degree rather than a hard-and-fast rule of categorization.

Some early research treats current fertility as a stock-adjustment process whereby individuals attempt to achieve some desired level of completed fertility (Lee 1981; Schultz 1980). Current fertility is modelled as a function of the gap between desired lifetime family size and current family size, among other things. This induces a lag structure for fertility in that lagged fertility levels become important explanatory variables for current fertility.[32] This approach represents a direct attempt to “dynamicize” earlier static lifetime fertility models, without having to solve the underlying dynamic maximization problem. Applied to fertility data at higher levels of aggregation than individual households, fitted stock-adjustment equations were shown to have some descriptive value (Schultz 1980; Lee 1981).

The stock-adjustment approach has the advantage that econometric procedures for estimating such equations are straightforward and well-known. However, application of this approach to household data may face difficulties in the presence of individual heterogeneity, which can vary over time and confound the presumably fixed parameters of individual stock-adjustment equations. Moreover, at the aggregate level, the issue of parameter stability in *ad hoc* equations can be even more serious, as the well-known critique of Lucas (1976) shows. Besides this issue, there is the more general question of how to interpret reasonable empirical findings, when it is unclear what kind of individual household behavior underlies the stock-adjustment process that one is fitting.

Other approaches set aside altogether the issue of fitting the actual dynamic structural equations (or approximations thereof) in favor of testing model implications directly. The estimating model can be based closely or loosely on an underlying theoretical dynamic model, but it serves mainly to test implications or predictions. A case in point is Cigno and Ermisch (1989), which outlines an explicit theoretical framework (based on Cigno 1988) of the timing and spacing of births, emphasizing the effects of women's earnings profiles over the life-cycle. Their model differs substantially from the models mentioned in Sect. 2 in that it assumes perfect capital markets and ignores the stochastic nature of the birth process. The authors explicitly recognize, however, that they cannot estimate the structural parameters of the theoretical model. Empirical analysis is based, therefore, on a posited ordered probit model of current family size as a function of individual characteristics. They use the probit framework to test theoretical predictions regarding the effects of fertility variates (age at marriage, work experience, last occupation, education, lifetime earnings, time) on the time profile of birth rates.

Rosenzweig and Schultz (1985, 1987, 1989) analyze a *time-aggregated* version of what they call an individual's "reproduction function." This function, essentially a lifetime fertility equation, is a key component[33] of their 1985 theoretical model nested within the dynamic framework of Sect. 2. A similar type of lifetime fertility equation is also employed by Moffitt (1982). In both cases, the time-aggregated reproduction function regards *lifetime* conception rates as a function of the individual's age (or average age during the fertile cycle), along with other variables of interest. (Moffitt looks at a "relative" lifetime wealth variable, while Rosenzweig and Schultz study persistent individual heterogeneity and average contraceptive usage over the individual's lifetime.) Despite the incorporation of time effects via the age variable, both models, as implemented, reduce to essentially static models of lifetime fertility. Dynamic issues like the pace and spacing of births can be studied (since the age variable acts like a time trend) but only to a very limited extent. Further, to do this requires the analyst to adequately control for *all* individual-varying components of the model, so that the variation of ages in the sample picks up the low-frequency, long-run movements in total births.

As implemented, however, the Rosenzweig-Schultz and Moffitt models are not as well-suited for analyzing fertility dynamics as many of the models discussed previously. Put simply, fertility variates like birth intervals or the time of the first birth are as much affected by temporary shocks as they are affected by persistent, long-run components of behavior, and would be inadequately analyzed by relying on a simple time trend disguised as an age variable. Questions of adjustment over time are more naturally handled by building in lag structures and serial correlation, but time-aggregation minimizes, if not eliminates, the role of these processes. These models did fill a gap in the empirical fertility literature at the time by introducing *ad hoc* dynamics (via the age variable). But their usefulness is mainly for studying the implications of intertemporal optimization for *lifetime* fertility-outcomes, not for analysis of intertemporal behavior itself.

Lastly, to explain recent Swedish fertility behavior, Walker (1995) presents a neoclassical economic framework to assess the effect of public policies on the price of fertility, and on the timing and spacing of births. Specifically, he looks at two measures: (i) the shadow price of fertility for a 25-year old woman, and (ii) relative price of fertility at ages 35 and 24. He finds that the estimated time series of prices of fertility can well explain the recent Swedish fertility pattern. Both economic conditions and public programs contriubed to the observed fertility pattern. Walker's use of the shadow price of fertility offers a simple way of incorporating public programs and economic conditions into a single measure to assess their effects on fertility dynamics. However, as noted by the author, the strong and ad hoc nature of assumptions needed to generate the estimates of the shadow price caution against direct interpretation of the causations involved.

5. Conclusions

This paper surveys dynamic microeconomic models of fertility, which we have taken to mean those that explain the evolution of fertility behavior and outcomes of economizing individuals over many time periods. We have focused attention on dynamic fertility models of the structural and reduced-form variety.

Well-specified structural models can accommodate complicated interactions between fertility variates and other economic variables, leading to a rich set of analytical predictions about observed behavior. They can also accommodate many general and realistic patterns of uncertainty and serial correlation, and offer clear and explicit interpretations of estimated parameters and relationships. Since structural models impose relatively strong restrictions on the data, one would expect them to be less useful as descriptive models of fertility data. Wolpin (1984), however, shows that a well-designed parametric structural model can adequately characterize the complex time-profile of fertility variates in a parsimonious way. Analytically, the principal problem with a structural approach is the lack of closed-form solutions for any except very stylized versions of these models. In applied practice this necessitates the use of fairly complex, if accurate, procedures for the numerical calculation of structural estimates, procedures which are also quite computational expensive from the standpoint of the average user.

Reduced-form approaches are viable alternatives that trade-off strict adherence to exact dynamic solutions in favor of approximate solutions that yield tractable closed-forms. Reduced-form models, such as hazard or duration models, can also accommodate complex interactions between fertility variates, while permitting fairly general error structures. This flexibility confers great advantages in terms of their ability to describe actual data. At the same time reduced-form models are set up in a way that makes estimation relatively straightforward. These advantages are likely behind the continuing popularity of hazard models in applied fertility research.

Counterbalancing these advantages, however, are the interpretation issues associated with the use of reduced forms. Specifically, issues like that of unobserved heterogeneity, which may be better handled by an explicit structural model, bedevil the interpretation of results from reduced-form models. Finally, despite the large body of empirical work that relies on reduced-form models, more work with better micro data is still called for, and perhaps closer adherence to exact theoretical restrictions may also be needed to sharpen empirical results.

Heckman and Walker (1990 a) note that there is little consensus on theoretical models at the moment. This has two implications for the development of structural models. On the one hand, disagreements about the appropriate theoretical model will caused researchers to rely more heavily on less stylized methodologies, such as hazard models, that are conformable with several possible underlying theories. On the other hand, disagreements of this sort are healthy for the development of structural models to the extent that experimentation with functional forms, error structures, etc., is encouraged.

A critical problem in the application of structural models is the lack of sufficiently realistic models that are easily estimated. This is related to the general lack of closed-form dynamic solutions to structural models. To some extent, these difficulties are being alleviated by recent advances in computational techniques such as those covered in Sect. 2.4 above. These represent a definite step forward, but researchers are still far from having tractable and realistic structural fertility models that can be estimated in transparent and straightforward ways.

Our belief is that much of the demand for applied analysis is currently being satisfied by reduced-form models of the Heckman-Walker or Hotz-Miller type. There exists, however, a real possibility for significant discoveries from application of models of the structural variety, especially if empirical techniques like that of Hotz and Miller (1993) or Rust (1995) lead to dramatic reductions in estimation complexity. But even empirical studies of the predictions of a dynamic model (e.g., Cigno and Ermisch 1989) can offer useful insights that can reshape our prior understanding of fertility relationships. Inasmuch as these priors affect fundamental things, such as the choice of regressors of a reduced-form hazard model, the practical impact of structural models can be nontrivial.

It is clear, however, that fertility models have entered a new phase in their development. In this new era, dynamic issues such as the onset, pace, and spacing of births, and the evolution of contraceptive behavior over time are the central analytical and empirical issues of interest. The precise nature and structure of uncertainty in a model, the transmission mechanisms, and the adjustment patterns for these random shocks are now among the critical specification issues facing researchers. The complex relationships between fertility variables, economic variables, infant mortality risk, exogenous shocks, and unobserved heterogeneity suggest, indeed demand, sharper lenses with which to examine existing and forthcoming data sets. The methodologies discussed herein represent, in our view, the existing state of a still-evolving art.

Endnotes

[1] The remaining "other" models include partly dynamic models that test model implications but do not fit structural dynamic equations, and models which introduce dynamics in *ad hoc* fashion, e.g., lagged adjustment or distributed-lag models. An example of a partly dynamic framework is the estimating model of Rosenzweig and Schultz (1985). Cigno and Ermisch (1989) test implications of their structural model without estimating structural parameters. Examples of *ad hoc* estimating models are Montgomery and Casterline (1993), and Lee (1981), and Schultz (1980).

[2] Examples are Butz and Ward (1979), Barro and Becker (1988, 1989), Benhabib and Nishimura (1989), Heckman and Walker (1989), and Macunovich (1995).

[3] One exception is the macro-time series framework of Butz and Ward (1979).

[4] To cite a few: (i) fertility rates in developing countries appear to be converging over time; (ii) the spacing of births tends to become narrower among higher fertility women, and becomes less narrow as parity increases; (iii) current and future levels of the woman's wage and husband's income are generally important explanatory variables for the number of births *and the timing of the first birth*; (iv) fertility declines as the risks of infant mortality decline; (v) at the aggregate level, fertility levels in the developed economies like the United States are countercyclical to the business cycle. (As labor force participation rates are procyclical, this may be vievwed as evidence of a fertility-work tradeoff.)

[5] The property of decisions being forward-looking means the following: if one were to look at the segment of the optimal fertility policy beinning from some arbitrary time t_0 in the life-cycle, that segment is optimal for balance of the individual's lifetime, given what the individual knows at t_0 about the future. Time consistency, on the other hand, implies that at any point of time t_0, the optimal policy for the balance of the individual's life-cycle will be unchanged as long as the *state* (i.e., the set of "givens") at t_0 is unchanged. Time-consistency means that the past history of choices or events has no independent influence on optimal choices for the future. The current state is all that matters.

[6] Among the models that fit under this framework are Heckman and Willis (1976), Wolpin (1984), Hotz and Miller (1984), Rosenzweig and Schultz (1985), Newman (1988), and Leung (1991).

[7] As noted by Leung (1991), when a continuous measure of family size M_t is desired (say in order to define the derivative of U with respect to M_t), M_t can be replaced with a flow of "child services" in efficiency units, i.e., $M_t^* = \psi_t M_t$, where ψ_t is a constant "service" coefficient taking continuous values over time. M_t^* is continuous and porportional to family size M_t.

[8] Rosenzweig and Schultz (1987) state that this assumption is not critical for most analytical results, but it is not known whether the quantitative effects matter. In a recent paper by Rosenzweig and Wolpin (1993) Indian farmers' investment in bullocks (cattle) was found to be useful for intertemporal wealth transfers in the face of risk. While children are generally not liquidated for purposes of consumption smoothing, children may be put to work at some stage even before full adulthood to smooth incomes. Moreover, as mentioned in the text, a new child, being durable, can serve as a vehicle (albeit a very inefficient one) for transferring income intertemporally.

[9] Leung (1991) really has two models: a basic fertility model and a model of parental sex preferences. What we have been referring to in this paper is his basic fertility model, which is virtually the same as the Heckman and Willis (1976) model. The choice of u_t, however, is not the focal point of his paper.

[10] Newman's (1988) model is set up in continuous time, so that the probabilities π^b and π^m are continuous-time (Poisson) processes. Newman's model is unique among dynamic fertility models in that it possesses a closed-form solution for the optimal u_t lies in the interior of the interval $[\underline{u}, \bar{u}]$, and the value function (1). His solution assumes that the optimal u_t but this is not restrictive in his view. The presence of any (psychic) costs of using contraception discourages individuals from pushing u_t to its upper bound, while one never goes to the lower bound because there are also implicit costs of attaining maximum fecundity, such as too high a level of coital frequency.

[11] When the MCC curve does not exist, as in models where U does not take u as an argument (Wolpin 1984; Hotz and Miller 1984; Rosenzweig and Schultz 1985) contraceptive choice

reduces to choosing either full contraception $u=\bar{u}$ whenever $\Delta V<0$, or minimal contraception $u=\underline{u}$, whenever $\Delta V>0$.

12 This result appears as early as Heckman and Willis (1976), where the value function V is seen as the discounted sum of several indirect utility functions, each of which is concave in M_t. Hotz and Miller (1984), however, argue that concavity is not straightforward for maximum-value functions such as V. Newman (1988) proves concavity of V in M_t under quadratic utility by the "guess-and-verify" method. To do this Newman exploits the similarity between the functional form of the post-childbearing value function and the (concave) utility function during childbearing years. Leung (1991) also proves concavity (cf. his Proposition 1), this time using an induction argument based on concavity of the post-childbearing value function. Leung mentions the Newman result but claims that his concavity result is a new result. More accurately, his proof is new and more general.

13 This sign ambiguity raises the possibility of threshold effects associated with the values of exogenous variables or parameters.

14 To the extent that mothers in developed economies have higher health stocks than their counterparts in developing economies, the natural odds of a birth π^b will be higher, everything else the same, and thus the more rigorous the contraceptive regime will be as parity rises. In Newman (1988) it is lower infant mortality risk, π^m, rather than higher natural fecundity, π^b, behind the size of the shift in EMBC. The difference between the result in the text and his is due to the simplifying assumption we made that the net birth probability is given by $\pi=\pi^b(1-u_t)-\pi^m$ which assumes no interaction of π^m with u_t.

15 See Wolpin (1984), Table 3, for an example based on Malaysian data.

16 Of course if the EMBC curve were located below zero, so that the individual wants to have a birth, rather than prevent one, she should already be at the lowest contraceptive level $\underline{u}$. In Newman (1988) the solution for u is always an interior one, and corresponds to the situation of EMBC>0 discussed in the text.

17 The expression for the MCC curve is $U_{X_t}\, p_t^u - U_{u_t}$. The derivative of this with respect to income I_t is $U_{XX} p_t^u<0$ (assuming that $U_{uX}=0$). This implies a downward shift of MCC.

18 Heckman and Willis point out that the optimal lifetime family size should be independent of the shape of the income time-profile when capital markets are perfect. The common assumption made in these models, however, is that capital markets do not exist that allow households to borrow on the basis of their future incomes. Under these conditions one may expect lifetime family size to vary with the growth rate of income, for a given level of permanent income. For example, when income is steeply rising, households find it optimal to delay or space births until later periods. Suppose that the household faces a nonzero probability of becoming sterile, as in Hotz and Miller (1984). Then interaction between birth delay/spacing behavior and this probability should lead to lower expected family sizes in households with steeply rising incomes.

19 Recall that in Wolpin (1984) fertility control is perfect.

20 As in Sect. 2.1, all expectations are conditional on the state at t. In Wolpin (1984) the state at t consists of the value of family size M at the beginning of period t, and all prior realizations of the model's random components, namely child mortality risk, preference shocks, and income shocks. There is a slight interpretational difference in the time subscript attached to the beginning-of-period family size M in the text above, and in the original Wolpin paper. In Wolpin (1984) the value of M at the beginning of period t was denote M_{t-1}, whereas we denote this as M_t. This difference is purely semantic – our time notation on M conforms to the timing convention made earlier in (3).

21 From this point on we follow the discussion in Keane and Wolpin (1994), adapting their findings and arguments to the original Wolpin (1984) estimation problem where possible.

22 As indicated at the beginning of Sect. 2.1, the error term ε_{t+1} arises from the birth and mortality risks, preference shocks and possibly other sources of uncertainty like income shocks or wage/price shocks. (Note that in the text, ε_{t+1} contains the preference shock θ_{t+1}.) In a more general sense, ε_{t+1} can be thought of as the forecast error at t.

23 More generally, the state M_t would not just be family size at t but all variables that summarize the current position of the dynamic system, e.g., the realization of any preference shocks at the beginning of t, income shocks that arrive at the beginning of period t, the age of the individual at t, etc. For purposes of our example we have assumed that the state at t is completely summarized by family size M_t.

[24] For the more specific problem of evaluating equation (16), or (9), the computational considerations involved are essentially the same as those involved in the evaluation of (10) of Keane and Wolpin (1994), or (11) in Wolpin (1984). To simplify calculations, Wolpin (1984) assumes normality of the preference shock θ_t alone. (Joint normality with other components of the error term ε_t was not imposed.) But even after simplifications, as per his equation (15) one must still evaluate several conditional expectations by numerical integration, and iterate backward for all time periods.

[25] See Kiefer (1988) and Lancaster (1990) for economic applications. Standard statistical books on hazard models include Kalbfleisch and Prentice (1980) and Cox and Oakes (1984). For a less technical discussion of hazard models and applications in social sciences, see Allison (1984).

[26] Sheps and Menken (1973), among others, develop models of fertility in which persistent differences among women in unobserved fecundity give rise to unobserved heterogeneity.

[27] The unobservable for spell j is $f_j\theta$ and the covariance between $f_j\theta$ and $f_k\theta$ is $f_j f_k \operatorname{Var}(\theta)$ (assuming $E(\theta)=0$).

[28] It is possible to parametrize $P^{(j-1)}$ to depend on regressors (Heckman and Walker 1987) but this was not pursued in Heckman and Walker (1990a, b, 1991).

[29] There appears to be a typo in their Table 2 since the estimated coefficients on female education are negative for all cohorts.

[30] This finding is in sharp contrast to their earlier result using the Hutterite data (Heckman and Walker 1987). Heckman and Walker (1990 a, b, 1991) offer the explanation that behavior swamps biology in modern Sweden.

[31] Hotz and Miller (1984) call equation (35) the "contraceptive index" equation. By (36), the interpretation one should attach to q_{it} is that *higher* levels of q_{it} are associated with higher birth probabilities, and therefore *lower* contraceptive levels. To avoid any confusion with our previous notation, one should remember that the index q_{it} is inversely proportional to our contraceptive efficiency u_t.

[32] In this sense, a recent paper by Montgomery and Casterline (1993) follows in this tradition, as it posits that current fertility in Taiwanese townships is influenced by the township's past fertility levels and past levels of fertility in other neighboring townships.

[33] In terms of the general framework of Sect. 2, the reproduction function is analogous to our net births equation (4), where the explicit birth and mortality probabilities are proxied for by heterogeneity in individual fecundity and the age of the woman.

References

Allison PD (1984) *Event History Analysis: Regression for Longitudinal Event Data.* Sage Publications, Beverly Hills, CA

Barro RJ, Becker GS (1988) A Reformulation of the Economic Theory of Fertility. *Quarterly Journal of Economics* 103:1–25

Barro RJ, Becker GS (1989) Fertility Choice in a Model of Economic Growth. *Econometrica* 57:481–501

Benhabib J, Nishimura K (1989) Endogenous Fluctuations in the Barro-Becker Theory of Fertility. In: Wenig A, Zimmermann KF (eds) *Demographic Change and Economic Development.* Springer-Verlag, Berlin

Butz W, Ward M (1979) The Emergence of Countercyclical U.S. Fertility. *American Economic Review* 69:318–328

Cigno A (1988) The Timing of Births: A Theory of Fertility, Family Expenditures, and Labour Market Participation Over Time. In: Wenig A, Zimmerman KF (eds) *Demographic Change and Economic Development.* Springer-Verlag, Berlin 133–151

Cigno A, Ermisch J (1989) A Microeconomic Analysis of the Timing of Births. *European Economic Review* 33:737–760

Cox DR, Oakes D (1984) *Analysis of Survival Data.* Chapman and Hall, London

David PA, Mroz TA (1989) Evidence of Fertility Regulation Among Rural French Villagers, 1749–1789. *European Journal of Population* 5:1–26, 173–206

Eckstein Z, Wolpin K (1989) The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey. *Journal of Human Resources* 24:562–598

Geweke J, Slonim R, Zarkin G (1992) Econometric Solution Methods for Dynamic Discrete Choice Processes. Unpublished manuscript, University of Minnesota

Hansen LP (1982) Large Sample Properties of Method of Moments Estimators. *Econometrica* 50:1029–1054

Heckman JJ, Singer B (1984) A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* 52:271–320

Heckman JJ, Singer B (1985) Social Science Duration Analysis. In: Heckman JJ, Singer B (eds) *Longitudinal Analysis of Labor Market Data*. Cambridge University Press, New York, 39–110

Heckman JJ, Hotz VJ, Walker JR (1985) New Evidence on the Timing and Spacing of Births. *American Economic Review, Papers and Proceedings* 75:179–184

Heckman JJ, Walker JR (1987) Using Goodness of Fit and Other Criteria to Choose Among Competing Duration Models: A Case Study of Hutterite Data. In: Clogg C (ed) *Sociological Methodology,* vol. 18. Blackwell, Oxford, 247–307

Heckman JJ, Walker JR (1989) Forecasting Aggregate Period-Specific Rates: The Time Series Properties of a Microdynamic Neoclassical Model of Fertility. *Journal of the American Statistical Association* 84:958–965

Heckman JJ, Walker JR (1990 a) The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data. *Econometrica* 58:1411–1441

Heckman JJ, Walker JR (1990 b) The Third Birth in Sweden. *Journal of Population Economics* 3:235–275

Heckman JJ, Walker JR (1991) Economic Models of Fertility Dynamics: A Study of Swedish Fertility. In: Schultz TP (ed) *Research in Population Economics,* vol. 7. JAI Press, Greenwich, CT 3–91

Heckman JJ, Willis RJ (1976) Estimation of a Stochastic Model of Reproduction: An Econometric Approach. In: Terleckyj N (ed) *Household Production and Consumption.* Columbia University Press, New York

Hotz VJ, Miller RA (1984) The Economics of Family Planning. Unpublished manuscript, NORC, University of Chicago

Hotz VJ, Miller RA (1988) An Empirical Analysis of Life Cycle Fertility and Female Labor Supply. *Econometrica* 56:91–118

Hotz VJ, Miller RA (1993) Conditional Choice Probabilities and the Estimation of Dynamic Models. *Review of Economic Studies* 60:497–529

Hotz VJ, Miller RA, Sanders S, Smith J (1994) A Simulation Estimator for Dynamic Models of Discrete Choice. *Review of Economic Studies* 61:265–289

Kalbfleisch JD, Prentice R (1980) *The Statistical Analysis of Failure Time Data.* New York: John Wiley and Sons

Keane M, Wolpin K (1994) The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. *Review of Economics and Statistics* 76:648–672

Kiefer NM (1988) Economic Duration Data and Hazard Functions. *Journal of Economic Literature* 26:646–679

Lancaster T (1990) *The Econometric Analysis of Transition Data.* Cambridge University Press, Cambridge

Lee RD (1981) A Stock Adjustment Model of U.S. Marital Fertility. In: Simon JL, Lindert PH (eds) *Research in Population Economics,* vol. 3. JAI Press, Greenwich, CT

Leung S (1991) A Stochastic Dynamic Analysis of Parental Sex Preferences and Fertility. *Quarterly Journal of Economics* 106:1063–1088

Lucas RE Jr. (1976) Econometric Policy Evaluation: A Critique. In: Brunner K, Meltzer AH (eds) *Carnegie-Rochester Conference Series on Public Policy,* vol. 1. North-Holland, Amsterdam

Macunovich DJ (1995) The Butz-Ward Fertility Model in the Light of More Recent Data. *Journal of Human Resources* 30:229–255

Manton KG, Stallard E, Vaupel JW (1986) Alternative Models for the Heterogeneity of Mortality Risks Among the Aged. *Journal of the American Statistical Association* 81:634–644

Moffitt RA (1982) Postwar Fertility Cycles and the Easterlin Hypothesis: A Life-Cycle Approach. In: Schultz TP, Wolpin K (eds) *Research in Population Economics,* vol. 4. JAI Press, Greenwich, CT, 237–252

Montgomery MR, Casterline JB (1993) The Diffusion of Fertility Control in Taiwan: Evidence from Pooled Cross-Section Time-Series Models. *Population Studies* 47:457–479

Montgomery MR, Trussell J (1986) Models of Marital Status and Childbearing. In: Ashenfelter O, Layard R (eds) *Handbook of Labor Economics,* vol. 1. Elsevier, Amsterdam, 205–269

Newman JL (1983) Economic Analyses of the Spacing of Births. *American Economic Review, Papers and Proceedings* 73:33–37

Newman JL (1988) A Stochastic Dynamic Model of Fertility. In: Schultz TP (ed) *Research in Population Economics,* vol. 6. JAI Press, Greenwich, CT

Newman JL, McCulloch CE (1984) A Hazard Rate Approach to the Timing of Births. *Econometrica* 52:939–962

Olsen RJ (1994) Fertility and the Size of the U.S. Labor Force. *Journal of Economic Literature* 32:60–100

Rodriguez G *et al. (1984) A Comparative Analysis of the Determinants of Birth Intervals. Comparative Studies,* vol. 30. World Fertility Survey, London

Rosenzweig M, Schultz TP (1985) The Supply of and Demand for Births: Fertility and its Life-Cycle Consequences. *American Economic Review* 75:992–1015

Rosenzweig M, Schultz TP (1987) Fertility and Investments in Human Capital: Estimates of the Consequence of Imperfect Fertility Control in Malaysia. *Journal of Econometrics* 36:163–184

Rosenzweig M, Schultz TP (1989) Schooling, Information and Nonmarket Productivity: Contraceptive Use and its Effectiveness. *International Economic Review* 30:457–477

Rosenzweig M, Wolpin K (1993) Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India. *Journal of Political Economy* 101:223–244

Rust J (1987) Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica* 54:755–784

Rust J (1994) Structural Estimation of Markov Decision Processes. In: Engle RF, McFadden DL (eds) *Handbook of Econometrics,* vol. 4. North-Holland, Amsterdam, 3081–3143

Rust J (1995) Using Randomization to Break the Curse of Dimensionality, manuscript. University of Wisconsin

Sah RK (1991) The Effects of Child Mortality Changes on Fertility Choice and Parental Welfare. *Journal of Political Economy* 99:582–606

Schultz TP (1980) An Economic Interpretation of the Decline in Fertility in a Rapidly Developing Country: Consequences of Development and Family Planning. In: Easterlin RA (ed) *Population and Economic Change in Developing Countries.* University of Chicago Press, Chicago

Sheps M, Menken J (1973) *Mathematical Models of Conception and Birth.* University of Chicago Press, Chicago

Smith AA (1992) Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions: Two Approaches. Working paper, Queen's University

Struthers CA, Kalbfleisch JD (1986) Misspecified Proportional Hazard Models. *Biometrika* 73:363–369

Sturm R, Zhang J (1993) Fecundity and Social Development in China: Changes in the Distribution of the First Conception Interval. *Biometrical Journal* 35:985–995

Tasiran AC (1995) Wage and Income Effects on the Timing and Spacing of Births in Sweden and the United States. *Contributions to Economic Analysis,* no. 229. North-Holland, Amsterdam

Tasiran AC (1996) A Reply to Walker's Note: 'A Comment on Tasiran'. Journal of Population Economics (forthcoming)

Taylor JB, Uhlig H (1990) Solving Nonlinear Stochastic Growth Models: A Comparison of Alternative Solution Methods. *Journal of Business and Economic Statistics* 8:1–17

Trussell J, Richards T (1985) Correcting for Unobserved Heterogeneity in Hazard Models: An Application of the Heckman-Singer Model to Demographic Data. In: Tuma NB (ed) *Sociological Methodology,* vol. 16. Blackwell, Oxford, 242–276

Vijverberg WPM (1984) Discrete Choices in a Continuous Time Model: Life-Cycle Time Allocation and Fertility Decisions. In: Schultz TP (ed) *Research in Population Economics*, vol. 5. JAI Press, Greenwich, CT, 51–85
Walker J (1995) The Effect of Public Policies on Recent Swedish Fertility Behavior. *Journal of Population Economics* 8:223–251
Walker J (1996) A Comment on Tasiran's 'Wage and Income Effects on the Timing and Spacing of Births in Sweden and in the United States'. *Journal of Population Economics* (forthcoming)
Yi K, Walker JR, Honore B (1987) CTM: A User's Guide. Unpublished manuscript, NORC, University of Chicago
Wolpin K (1984) An Estimable Dynamic Stochastic Model of Fertility and Child Mortality. *Journal of Political Economy* 92:852–871
Zhang J, Sturm R (1994) When Do Couples Sign the One-Child Certificate in Urban China? *Population Research and Policy Review* 13:69–81