

Reply to criticism on a model of birth spacing

Jan Van Bavel

(22 April 2004)

First of all, it is very important to stress that my duration model of birth spacing (Van Bavel 2003b; 2004) is not intended to model an ordinary fertility rate. The fertility rate can be decomposed into a probability of a birth on the one hand, and elapsed time to birth, given that it took place, on the other. The birth spacing model tries to assess the effect of a number of covariates on the elapsed time only, which can be called the speed or pace of reproduction, given that reproduction took place. The decomposition of fertility into a likelihood and the conditional speed does not negate the fact that both are related.

Multicollinearity

“I suspect that the inclusion of both a youth dependency ratio and a count of children is problematic because these two variables ought to be highly correlated. The simultaneous inclusion of both net and crude parity may also be problematic.”

“The inclusion of both the proportion of dependent children (less than 9) and parity is odd. In a nuclear family, the two are likely to have been highly correlated. Did the authors check for this correlation? It would be wiser to have included one or the other, not both.”

What is “odd”? And what do you mean by “check for correlation”? Check what?

A major reason for scientists to do multivariate regression analyses is that, very often, covariates and determinants of the dependent variable are highly correlated. If covariates in a regression analysis would not be correlated, there would be no need to do multivariate regression analysis. If two covariates X1 and X2 are correlated, the only way to assess the net effect of each of them is to include both of them in the same regression. If you drop one of them, say X1, then there is a chance that X2 will capture the effect of X1. As a result, you run the risk of falsely concluding that there is an effect of X2, while in fact the real causal determinant is X1. This is commonly called a spurious effect of X2. Conclusion: it would not at all “be wiser to have included one or the other, not both”. If your sample is big enough (see *infra*), the correct method is to include both X1 and X2, especially when they are highly correlated.

“Regarding the question of multicollinearity, my experience is that covariates usually do not change a great deal when adding more covariates, but if the correlation is "too high" it may be difficult to interpret the estimated effects, because you never know which covariate captures the effects. But here again, I think the best way is to test different covariate set-ups to see what happens if you add or remove covariates. My feeling is that a robust model is one where the results do not change a great deal when adding or removing single covariates.”

I don't agree with this test of robustness, although I know that this is a very common approach. Suppose that two predictor variables X1 and X2 are really causally and positively related with Y, and that X1 and X2 are very highly and positively correlated. If you drop X1 from the regression, the estimated effect of X2 on Y will increase. With this "incomplete regression", the effect of X1 may wrongfully be attributed to X2. The best way is to include both X1 and the "highly correlated" X2 in the analysis. But as a result, the estimated effect of X2 will weaken significantly. Your test of robustness would imply in this case that the correct full model is not robust. Or am I missing a point here?

If I am not mistaken, correlation between covariates is *in principle* never "too high" if one is not an exact linear combination of other variables. In practice, multicollinearity is really a sample size problem. High correlation between covariates is the main reason why we prefer multivariate over bivariate analysis. The problem of multicollinearity arises when the number of cases in some combinations of values of the covariates becomes too low for robust statistical inference. (That's why the econometrician Arthur Goldberger has renamed the so-called multicollinearity problem a problem of "micronumerosity", see Wooldridge 2003, pp.96-100). If there are too few cases, the estimated standard errors will be inflated and you will conclude correctly that the estimated effect parameters (or elasticities) are not reliable, as a consequence of multicollinearity/ micronumerosity.

“Finally, you have too many variables that measure the same things. Age, marital duration, parity, and proportion of dependent children (% <9) are all strongly related to the same time dimension. "Crude legitimate parity," "Net parity," and % <9 depend upon the number of children ever born. The issue is not multi-collinearity. Multi-collinearity is not a violation of the assumptions of the statistical model. It just makes larger standard errors, which isn't a problem if you have enough data. I am concerned about interpreting the model. When several variables measure the same thing it is very difficult to tell what each one is doing”

Again, I think that the whole point of multivariate regression analysis is that many of the variables we observe are highly correlated. If covariates with a potential causal effect would not be correlated, we would not need to do multivariate analysis and could just do bi-variate analyses to assess their effect.

The fact that the inclusion of all these correlated variables works fine in many models is a clear indication that there are independent effects of these variables. A clear indication

from a statistical point of view is that the estimated standard errors of the estimated regression parameters are not too much inflated, even with the small N's I use in Van Bavel (2003a&b; 2004).

Only closed birth intervals = bias!

"Wouldn't it be better to use all information, and not limit the analysis to the closed intervals, which could give you a sense of the practical impact of restricting the analysis to closed intervals."

The point is, as I have argued in the PopStudies article, that I want to condition on purpose on a birth taking place for two reasons: (1) exclude sterile couples (while most analyses in historical demography do as if the onset of sterility is at age 45 or 50 for all women); (2) decompose the fertility (i.e. a hazard) rate in a probability on the one hand, and the speed of events occurring (given that they eventually occur) on the other. So in the birth spacing analysis, I do not have the ambition to model ordinary fertility rates (events divided by exposure), but rather conditional rates, i.e. the pace of reproduction, given that reproduction occurs.

"I'm still not completely convinced that the exclusion of open birth intervals is the best way to study fertility behaviour. You may lose information that could be relevant especially when assured truncations for younger women are given. If we use Cox models, is it not because of censored data? Otherwise you could directly model the durations."

I agree that excluding open birth intervals is no good way to study "fertility behaviour" generally, but I do think that it's a good way to analyse the covariates of the length of closed birth intervals. If you proceed with closed intervals only, you should be aware that you are not analysing the unconditional fertility rate, but the conditional speed (and speed only) of parity progression, given that parity progression indeed did take place.

I agree that hazard rate models like the Cox model are particularly well suited to handle the problem of right truncation or censoring, but Cox did not design his model at all for that purpose. He designed it in order to be able to estimate multivariate regression parameters for the kind of rates you find in a life table (Cox 1972, p.187). Anyway: what do you mean by modelling the durations "directly"? Directly? You mean by OLS?

Some historical demographers seem to believe that Cox models are only applicable if dealing with censored data. Moreover, they commonly neglect the fact that the inclusion of open birth intervals generally biases analyses of fertility rates. That is: it is common practice to do as if the timing of onset of sterility is equal for all women, namely at age 45 or 50 (which is then taken to be censoring data), which is not true. It could be argued that this practice leads itself to serious biases, especially if you want to model birth (s)spacing.

If fertility is natural, the unconditional hazard rate is determined by two groups of couples in the risk set: couples who are (still) fecund and couples who are (already) sterile. If some are controlling their fertility by means of stopping behaviour, this will add to the hazard-depressing effect of sterility. For analysis of spacing, therefore, the unconditional hazard rate is heavily biased by sterility (or extreme subfecundity) and possibly also by stopping behaviour, if present. These two factors introduce a bias if we are interested in the length of birth intervals of fecund couples only, and in the determinants of this length, including intentional spacing behaviour. We want to model the speed of subsequent reproduction here, not the probability of subsequent fertility.

“In your analysis of birth spacing you exclude open birth intervals and use closed ones only. You state “that this is to exclude sterile and extremely sub-fecund couples. But by definition it would also exclude information from open intervals that resulted from a deliberate decision to stop. A couple that deliberately quit having births at age 30, for whatever reason, would have all of the information about its experience after age 30 eliminated. This biases the results in favor of the couples that continue to have children at regular intervals, that is, the non-stoppers. Why not include the open intervals, and treat the attainment of age 45 as a censoring event?”

In an analysis of birth spacing as a fertility control strategy, it is not a bias to eliminate information from people who have stopped reproducing. Birth pacing (or spacing) can only be observed as a fertility controlling strategy in the behaviour of those who are still reproducing. “Why not include the open intervals, and treat the attainment of age 45 as a censoring event?” Because such an analysis would include sterile couples as well as people who have deliberately stopped reproducing.

“I do not believe that you can use only the completed intervals and exclude the incomplete intervals in a Cox regression model. The whole point of event history models is to include censored observations. In effect, we assume that all intervals will be completed eventually, but some are censored before that happens. The first rule of event history analysis is that one does not terminate a life history in a way that provides information about whether the event will occur. Women do not know when they will have a birth or not. They are always waiting for the next birth to occur. In your approach all of the intervals end in a birth. This is informative censoring”

There are several important points here.

Of course, observations need not be censored in order to apply event history techniques - everybody would prefer uncensored duration data in principle. But indeed, the nice thing about modelling durations determined by rates, like in Cox regression, is that it is possible to include censored observations. However, this is only possible if we know how long observations are in the risk set; only if we know the censoring time.

This is usually not the case in the kind of data available to demographers. My dataset also lacks information on the onset of sterility. Yet, it is common practice to do as if the age of 45 or 49 years is "censoring time".

For the current purpose, I don't believe that this common practice is a good way to go. If you look at the literature about sterility (Pittinger 1973, Trussell and Wilson 1985, Larsen & Menken 1989, Wood et al. 1994), you see that even when considering sterility in women only, the proportion sterile increases quite steeply from the age of 35 years.

If we want to find out what are the covariates of the length of birth intervals, the risk set includes all people who are still at risk of giving birth. So not the couples who are sterile. The problem is that we do not know who these are. Doing as if all women are censored at the age of 45 or 49 years heavily biases the results of an analysis of birth pacing – even if that is common practice.

“Using only complete intervals violates your own assumption that some women were using spacing without intending to stop childbearing completely. Women that have longer birth intervals will also have longer incomplete birth intervals. So, you are systematically excluding more time at risk from women who may have been spacing. This is one of the problems that makes spacing so difficult to distinguish from stopping”

I use only those intervals that end with a birth (so before they are censored at some unknown point in time as a consequence of sterility) and you are right to point out that this is informative censoring. Indeed, there is a problem of selectivity in that couples with short birth intervals have a higher probability of closing their birth interval before the onset of sterility. But:

1. my (and I think most people's) best guess - I cannot prove this - is that the bias as a result of this violation of the independence assumption is much less severe than the bias introduced by including open birth intervals and, hence, including a considerable proportion of people who are actually sterile. What the model tries to capture is how the length of interbirth intervals can be described as a function of a set of covariates. The violation of the independence assumption makes my set of closed birth intervals selective in the sense that only the last birth (censored) interval of those with relatively long intervals tend to be selected out of the sample.
2. the bias introduced as a consequence of my informative censoring is in the conservative direction. Indeed, I agree with you that I tend to select people who are trying to delay the next birth out of the sample, at least their last (censored) birth interval (not their previous ones).

Proportion of dependent children

"I have very serious doubts about interpreting "proportion of dependent children <9." This variable is mostly measuring age and marital duration."

But those variables, age and marriage duration, are controlled for in the model. The effect of the proportion of children <9 years is surely not limited to high marriage duration. Of course, the fit of the model would improve if I would take into account the fact that the effect of marriage duration is not linear. I have checked this and it does not alter the other parameters substantially.

"By construction, this variable must equal one for women who have been married less than ten years. So, all of the variance occurs among older women, mostly after age 35."

The proportion of children under age 9 varies from the beginning; in the beginning of marriage, it varies between 0 and 1 as a function of (1) infant and child mortality, and (2) the number of children born already before and legitimised by marriage. After 10 years, the speed of parity progression is an additional source of variation.

"Are women with one child under 10 the same as women with 4 children under age 10?"

No, but I also include the number of births (crude parity) and the number of children alive (net parity). This captures any size effects.

"You are correct to say that couples with short birth intervals will have a higher proportion of young children. Unfortunately, you eliminated much of that effect by dropping the incomplete intervals. This variable is mostly comparing young couples (who all have a value of 1.0) to couples that began childbearing more than 10 years earlier. Since every couple starts with all children <9, this variable is really about the proportion of children 10 and older. You find that women with more children over age 10 have shorter birth intervals. The number of children under age 10 does not really matter."

Well, it does. Again, this argument implies that my estimates are conservative ones, as explained in the paper. A high proportion of children <9 is correlated with low marriage duration, and low marriage duration is negatively correlated with the length of birth intervals. This makes it very hard to detect the positive effect, if present, of proportion of children <9 on the length of birth intervals. This is only possible after controlling for marriage duration.

"I think that it is much better to use separate variables for numbers of children <5, 5-9, >10, etc."

My approach is a departure from that approach, which has been more common in past analyses. The problem is that in that way, different effects (size and proportion) are mixed up. The estimates of that kind of variables are therefore often hard to interpret.

Final birth interval and post factum modelling

“the models include as explanatory variables information about future events, which is a major no-no in estimation of event-history models. Specifically, there is an indicator of whether or not this is a final birth interval, that is, whether or not this is the last birth. Obviously, this includes information about the length of the interval after this birth, that it is infinite, or censored. [a] This is not only problematic because it is information about the future. [b] It should also be correlated with the parity measures discussed earlier. This indicator is more likely to be true at higher parities. It shouldn't be in the same model as the measures of parity and so forth.”

First, let me state clearly why I include this variable in the regression of closed birth intervals. From the literature on fecundity and natural fertility, we learned that there is a period of subfecundity preceding the onset of 100% sterility. As a result, final closed birth intervals tend to be much longer than non-final ones. Failed attempts to stop childbearing have the same effect (Anderton & Bean 1985; Knodel 1987; 1988). The model uses the knowledge that no more birth occurs after the next one as a proxy for these major determinants of the speed of reproduction, i.e. subfecundity preceding sterility and/ or failed attempts to stop childbearing.

Now, the above criticism identifies two problems: [a] using information about the future and [b] multicollinearity.

[a] Of course, the “final birth interval” is a post factum covariate, but recall that the model is from the outset a post factum model, not a Cox regression of an ordinary fertility rate. It tries to describe the speed of parity progression as a function of the enclosed covariates. The final birth interval variable can only feature in models of closed birth intervals. By definition, these are conditional and post factum because they are describing a rate given that a next birth took place within five years, which we can only know after the facts.

The criticism correctly states that the covariate “includes information about the length of the interval after this birth”, but the latter censored intervals are not to be explained in this regression. There is indeed a correlation between the length of birth intervals and whether or not they are final ones. For reasons explained above, that’s exactly why I prefer to include this covariate. Within the context of this conditional model, I don’t see how the inclusion of this variable would bias the parameter estimates of the other covariates. Rather on the contrary.

Although it does not alter the results substantially, excluding all final intervals from the analysis is a bad idea, because then it is implicitly assumed that prolonged final intervals are *always* the result of failed stopping attempts. However, spacing as well leads to longer final intervals. By including the final interval dummy, we guarantee that the effects of the other covariates are not limited to the final interval.

[b] The fact that the final interval dummy is more likely to be true at higher parities is not an argument against but an argument in favour of including both the parity measures and the final interval dummy. See the section on multicollinearity. Or am I missing the point made here?

Non-linear effects

“The operationalization of parity is peculiar. By using a count variable, it assumes a linear effect of parity. If the effect of parity is non-linear, its effects will be underestimated. It might make more sense to categorize parity.”

“You might want to categorize the counts of numbers of past births. Including a count variable for them will only pick up the linear portion of the effect of numbers of children.”

Too some extent, I agree in principle – although it will depend on the shape of the functional relationship between the variables whether or not some portion of the relationship will be “picked up” and/ or whether the picked-up part will overestimate or underestimate the real strength of the relationship. In practice, categorizing is not always a good solution because of sample size limitations. There are other ways to capture non-linear relationships. I agree that the model could be improved in theory (but not necessarily in practice because of the limited sample size) by accounting for non-linear relationships.

“Odd” age effects in the birth spacing model

“The consequences of informative censoring are apparent in your results. The relative risks by age of mother are completely unrealistic. Your tables imply that 40-year old women were almost as likely to give birth as women in their 20s. We know that is not true. You probably get this result because you have systematically excluded women with longer birth intervals at older ages.”

First, the words “almost as likely” are not applicable here: the hazard rate modelled in the birth spacing model is conditional in the sense that it applies only to closed birth intervals, which implies a conditional probability of birth equal to one for all women, regardless of age. Recall that I have excluded sterile couples out of the sample on purpose. The literature on natural fertility tells that sterility is by far the major determinant of the age pattern of fertility (Wilson, Oeppen and Pardoe 1988). I do not want to model the (unconditional) fertility rate here, but the post factum speed of parity progression. Post factum, because conditional, i.e. given that we know that another birth eventually occurred. In the youngest generations, the effect of age on the conditional

speed of parity progression disappears as a consequence of the presence of stopping behaviour in those generations, but *only after controlling for marriage duration*. Indeed, the length of birth intervals *for nonsterile women* is more a function of marriage duration than a function of age. ("The saddest curve", in Charles Westoff's words).

In other words: I think the best way to understand the age pattern of natural fertility is that it is a mixed function of the age pattern of sterility on the one hand, and the age-at-marriage pattern (or, equivalently, marriage duration at a given age). What I have done in this paper is to exclude the influence of the first function.

Heterogeneity

"I would encourage the authors to consider experiments with models that allow for heterogeneity in fecundity between women. We know that some women are more likely to conceive than others, for purely physiological reasons. This may even account for the positive association between parity and subsequent pace of births. In a discrete-time analysis, you could estimate a fixed effect model through use of a conditional logit. I think there are analogies for Cox regression that allow each woman to have their own fecundity."

First of all, I agree that the model may be improved by including a random effect (or "frailty") parameter to allow for unobserved heterogeneity on the level of women/ couples/ families. But I think that the model already does a better job than previous attempts to control for differences in fecundity. I argue, both in Van Bavel (2003a) and (2004), that the heterogeneity introduced by differential fecundity is not completely unobserved (see also Heckman & Walker, 1992), and that we can control for the observed heterogeneity by including the crude parity measure in the model. Indeed, *after controlling for age and marriage duration*, crude parity is a proxy for fecundity – given that all these variables are observed at the beginning of the current interval. See Van Bavel (2004): "Crude legitimate parity is defined as the number of children already born within the current marriage at the beginning of the interval. It represents natural fecundity differences between couples: couples with on average short birth intervals and, hence, more births at a given age and marriage duration, can be expected to have shorter birth intervals in the future as well. Differences between couples reflect differential fecundability and breastfeeding habits (Knodel 1988; Wood 1994)."

References

Anderton, D. L. and Bean, L. L., 1985. 'Birth spacing and fertility limitation: A behavioral analysis of a nineteenth-century frontier population'. *Demography* 22: 169-83.

- Cox, D. R. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society* B34:187-202.
- Heckman, J. J. and J. R. Walker. 1992. Understanding third births in Sweden, in J. Trussell, R. Hankinson, and J. Tilton (eds.), *Demographic Applications of Event History Analysis*. Oxford: Clarendon Press, pp.157-208.
- Knodel, J., 1987. 'Starting, stopping, and spacing during the early stages of fertility transition: the experience of German village populations in the 18th and 19th centuries'. *Demography* 24: 143-62.
- Knodel, J. 1988. *Demographic Behavior in the Past: A Study of Fourteen German Village Populations in the Eighteenth and Nineteenth Centuries*. Cambridge: Cambridge University Press.
- Larsen, U. and Menken, J., 1989. 'Measuring sterility from incomplete birth histories'. *Demography* 26: 185-201.
- Pittinger, D., 1973. 'An exponential model of female sterility'. *Demography* 10: 113-21.
- Trussell, J. and Wilson, C., 1985. 'Sterility in a population with natural fertility'. *Population Studies* 39: 269-86.
- Van Bavel, J., 2003a. 'Does an effect of marriage duration on pre-transition fertility signal parity-dependent control? An empirical test in 19th century Leuven, Belgium'. *Population Studies* 57: 55-62.
- Van Bavel, J., 2003b. 'Birth spacing as a family strategy. Evidence from 19th century Leuven, Belgium'. *The History of the Family: A n International Quarterly* 8: 585-604.
- Van Bavel, J., 2004. 'Deliberate birth spacing before the fertility transition in Europe: evidence from 19th century Belgium'. *Population Studies* 58: 95-107.
- Wilson, C., Oeppen, J., and Pardoe, M., 1988. 'What is natural fertility? The modelling of a concept'. *Population Index* 54: 4-20.

Wood, J. W. 1994. *Dynamics of Human Reproduction. Biology, Biometry, Demography.* New York: Aldine De Gruyter.

Wood, J. W., Holman, D. J., Yashin, A. I., Peterson, R. J., Weinstein, M., and Chang, M.-C., 1994. 'A multistate model of fecundability and sterility'. *Demography* 31: 403-26.

Wooldridge, J. M., 2003. *Introductory Econometrics. A Modern Approach.* Mason (Ohio): Thomson/ South-Western.
