Extended Abstract: Markov Variation and Sex Ratio Dependence in the United States

Michael S. Pollard, mpollard@soc.duke.edu
Carolina Population Center,
UNC-CH 123 West Franklin St.
Chapel Hill, NC 27516-2524

Biological sex is one of the most important and prevalent variables considered in social research, as well as serving as a key characteristic around which most societies are currently structured. The survey measurement of an individual's sex is often assumed to be among the most precise in social science. *The Methods and Materials of Demography* states that "in the statistically developed countries, misreporting of sex is negligible" (p106, Shyrock, Seigel, and associates 1976). More problematic is the measurement of the sex of an individual's children, although this too is typically accepted to be a well measured characteristic, particularly when relying on the mother's report.

If we can assume that the sex of children born to women of parity *x* is exogenous with respect to a wide variety of factors linked to choice behaviors we may use it as powerful tool for identifying causal relations. In addition to precision of measurement, other features of fertility are often highlighted as strengths in their use in research. In the absence of sex preselection, it is typically assumed that there will be a roughly random, 50/50 chance of having a boy or a girl at each birth (although the probability of having a boy is slightly favored over the probability of having a girl). The sex ratio (males to females) at birth is thought to reliably measure 105 males born to every 100 females (Bongaarts and Potter 1983). While the sex ratio among newborn infants has long been of interest (see James [1987] for a review), recent evidence of declines in the sex ratio at birth in several European countries (Dickenson and Parker, 1996; Møller, 1996) and

Canada (Allan et al., 1997) has revitalized research attention in the natural variation in the sex ratio.

The sex of children born to women of parity *x* is typically assumed to be exogenous with respect to a wide variety of factors that may be linked to choice behaviors, such as subsequent fertility and marital disruption. In other words, at a given parity, the birth of a son or a daughter, due to a "random" assignment of X and Y chromosomes, can essentially be considered as Campbell and Stanley's (1966) classic quasi-experiment. This research design provides a powerful tool for identifying a causal relation, and it is thus in our interest to examine this assumption in order to reasonably verify it. That is, we need to examine the issue of sex ratio (in)dependence.

The sex ratio has been shown to vary slightly by a variety of factors, including the age of the mother and father, birth order, race, coital frequency, and multiple birth[1] (Jacobsen, Møller, and Mouritsen, 1999; James, 1987; Shyrock, Seigel, and associates, 1976). Mixed evidence also exists for variation by social class, rural/urban, and season (James, 1987; Maconochie and Roman, 1997).

Variation by sex composition of previous children is also inconsistently identified. Edwards (1960) describes three theoretical kinds of natural variation in the sex ratio that have since been frequently examined in the literature: Lexis, Poisson, and Markov.

1) Lexis variation suggests that the probability of having a boy is constant within individual couples, but varies across couples.

[1] Lower sex ratios are associated with older parents, African American parents, increased coital frequency, multiple birth, parents in lower social classes, rural residence, and fall/winter births.

2) Poisson variation suggests that the probability of having a boy varies from one pregnancy to the next, but has the same mean for all couples.
3) Markov variation suggests that the probability of having a boy varies within couples according to the sex of previous births. Positive Markov variation would be present if the probability of a boy increased with prior male births, while negative Markov variation would be present if the probability of a boy decreased with prior male births.

While most research fails to find evidence of Markov variation (e.g., Greenberg and White, 1967; Jacobsen, Møller, and Mouritsen, 1999; Maconochie and Roman, 1997), and some researchers have concluded it simply does not exist (James, 2000), a few studies do suggest the presence of positive Markov variation (Ben-Porath and Welch, 1976; Edwards, 1966), which if true, would require a reformulation of the methods of testing for Lexis and Poisson variation (James, 2000).

I examine the U.S. data for evidence of Markov variation. While most studies discount the existence of Markov variation, they tend to do so with relatively small sample sizes and reject the hypothesis when sibling sex is not significantly correlated. However, it is possible that these studies are committing type II error, incorrectly rejecting a true hypothesis. Testing for the presence of Markov variation requires data on large numbers of sibships, including sequence, not simply composition, and thus places high demands on the data source.

The task below is an examination of the assumption of sex ratio independence; that is, can we generally accept the assumption that there are no strong associations between the sex of children born and other factors that would weaken the assertion that sex composition of children can be used as a causal mechanism?[2] Specifically, can we identify socio-demographic factors that are associated with male births, including the sex of previous children? In addition to a consideration of secondary sex ratio dependence, I will also consider the potential relationship between other sociodemographic variables and sex of birth. Finally, I will examine the extent and predictors of missing information on sex composition of children in the Current Population Survey (1980, 1985, 1990, 1995).

The Current Population Survey

The Current Population Survey (CPS) is a monthly survey of roughly 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. And has been conducted for over 50 years. The CPS primarily serves as a source of information on the labor force characteristics of the U.S. civilian noninstitutional population. In addition to labor force information, a variety of demographic characteristics including age, sex, race, marital status, and educational attainment are also measured. Supplemental questionnaires focusing on a variety of topics, such as school enrollment, health, and most relevant for the present purposes, fertility and marital histories are generally rotated into the regular CPS questionnaire. The June 1980, 1985, 1990, and 1995 waves of the survey contain additional detailed fertility and marital histories, as well as comprehensive data on personal characteristics such

[2] i.e. considered a truly exogenous characteristic with respect to most social, economic, and cultural factors.

as age, sex, race, educational background, and Hispanic origin for a representative sample of the (civilian, noninstitutional) population, and are used here.

Markov Variation

As described above, there are three theoretical kinds of natural variation in the sex ratio that have since been frequently examined in the literature: Lexis, Poisson, and Markov. Recall that: Lexis variation suggests that the probability of having a boy is constant within individual couples, but varies across couples; Poisson variation suggests that the probability of having a boy varies from one pregnancy to the next, but has the same mean for all couples; and that Markov variation suggests that the probability of having a boy varies within couples according to the sex of previous births. Positive Markov variation would be present if the probability of a boy increased with prior male births, while negative Markov variation would be present if the probability of a boy decreased with prior male births. Here we are most interested in Markov variation.

Most previous research fails to find evidence of Markov variation (e.g., Greenberg and White, 1967; Jacobsen, Møller, and Mouritsen, 1999; James, 2000; Maconochie and Roman, 1997). It is possible that Markov variation does exist, but researchers have been unable to meet the demanding data size requirements.

The large sample size of the CPS, in conjunction with the fertility histories for the first four births, enables us to check for the presence of Markov variation in sibships.

Edwards (1966) argued that if the sexes of non-consecutive siblings were correlated positively there is evidence of Lexis variation. If the only positive correlations observed are between the sexes of consecutive siblings, then there is no Lexis variation, only Markov. This is examined using the CPS 1980, 1985, 1990, and 1995 data and logistic regression as shown in Table 1. Significance is assessed using the likelihood ratio statistics, which can be compared with a chi-square distribution (with degrees of freedom appropriate to the number of parameters estimated).

[Table 1 About Here]

The top section of Table 1 indicates the relative probability of having a male baby by the sex of the immediately preceding sibling. The bottom section indicates the probability of having a male baby by the entire sex composition of preceding siblings. Lexis variation suggests that couples have a predisposition to having children of a particular sex; that is there should be correlations observed in the lower section of the Table (favoring all boys or all girls). Markov variation suggests that the sex of a given child is affected by the sex of the immediately previously born child, which would be indicated in the upper section of the Table. From the lower section of the Table we are not able to see clear indications of a consistent tendency for boys or girls; the only significant correlation is at parity 3, and indicates that parents with both a boy and a girl and more likely to have a girl as their third birth (which is not repeated at parity 4)[3]. However, in the upper panel we do see the consistent suggestion of Markov variation. At parity 2, results are not significant but the correlations are in the direction anticipated. At parity 3, parents with a girl as their second birth are significantly more likely to have a girl for their third birth than are parents who had a boy as their second birth. At parity 4,

[3] I attribute this unanticipated and unpatterned result to sampling variation.

results are not quite significant at the .05 level, but are again in the anticipated direction: for couples where the third child was a girl there is the suggestion that the fourth child is also more likely to be a girl. Additional work examining this result by race/ethnicity, and additional validity checks will be conducted.

Additional Factors Associated With Male Births

We may also investigate the suggestions from previous research in other national contexts that certain factors naturally alter the odds of having a boy or a girl. Specifically, the age of mother, parity, calendar year, season of birth, multiple birth, race/ethnicity, and educational attainment have all been mentioned by previous research as possible factors (e.g., Jacobsen, Møller, and Mouritsen, 1999; James, 1975, 1987; Maconochie and Roman, 1997). With the exception of "season", the distribution of each of these variables has also changed over time, which could potentially contribute to an understanding of any changes in sex ratio at birth at the national level.

Table 2 presents the probability of having a male child among first to fourth order births in the CPS data (all four fertility supplement cycles), by selected characteristics of the birth. Consistent with prior research, there are significant, although slight, relationships between both parity and race and the probability of having a male baby. Higher parity births are disproportionately female, as are births to black women (relative to white women). Despite the large sample sizes, none of the other factors considered were significantly related to the probability of having a male child.

[Table 2 About Here]

Missing Sex of Child Information

In contrast to the opening statements regarding the high quality of sex of children reports, preliminary work indicated a surprising amount of missing data on sex of children in the Current Population Survey. Thus, I examine what factors predict or are associated with the missing sex of child(ren) cases. Table 3 presents the (unweighted) percentage of female respondents in the 1985, 1990, and 1995 CPS at parities 1-10 with missing information on sex for their first, second, or third births. There are clearly a substantial proportion of cases at each parity in each survey with incomplete information, ranging from roughly 13% at lower parities to over 30% at the very high parities. While the number of cases at the highest parities is relatively small, even women at parity 1 present substantial missing information: for example, 17.22% of the female respondents in the 1985 CPS with only one child did not report the sex of that child. As parity increases, generally so too does the percentage of missing responses.

[Table 3 About Here]

Next we may identify characteristics of parents (more precisely the mother) that are associated with missing sex of child information. Using a straightforward set of logistic regressions it is possible to identify background factors that are significantly associated with missing sex of child information. The responses on the dependent variable can only have values of zero or one (in this case, not missing, or missing, respectively), and logistic regression, a maximum likelihood technique, is appropriate in such situations (Maddala 1983). Logistic models estimate the log-odds that a value of the independent variable is associated with the dependent variable, all else being equal. The odds are simply the ratio of two probabilities: the probability of (missing information) and of (not missing information). Odds ratios can be calculated from the logistic models by taking the anti-log of the parameter estimates, exp[β]. A simple transformation, 100(exp[β] - 1), can be interpreted as the percentage change (reduction or increase) in the odds of missing information for a one unit increase in a given independent variable, holding the other variables constant (Long 1997:81).

A set of dummy variables was constructed for race (white [reference], black, Hispanic, other), current marital status (married [reference], divorced/separated, widowed, single/never married), and region of residence (Northeast [reference], Midwest, South, West). Additionally, continuous variables were constructed measuring the number of times the respondent has been married, a 12-point ordinal scale of educational attainment (1=less than grade 1, 12=doctoral degree), parity (total number of live births), age in years, age of oldest child (in years), and duration of current marriage (in years). The variables are based around simple concepts in order to detect systematic patterns in the missing data; are certain groups underrepresented in the completed responses?

Table 4 presents the results of a series of 8 logistic regressions predicting missing sex of child information for specific parities (1 through 3), missing information more generally (missing any of the first three children), and for the age group that will be considered in later chapters (women under age 41; see subsequent chapters for justification). The results in Table 4 are based on the combined CPS-90 and CPS-95. Results were calculated for both surveys individually, and did not substantively differ. Because the results were not substantially different, the two surveys were combined; although the CPS-80 and CPS-85 are not considered, results are likely applicable to the earlier cycles.

[Table 4 About Here]

Models 1 through 3 in Table 4 predict missing information on sex of the first, second, and third birth, respectively, for all women in the combined CPS data file who were at, or above, the required parity. For missing information on parity 1, almost all of the factors considered are significantly related to missing information on sex of first child. In Model 2, an indicator of whether the information on the sex of the first child was missing is included, and it renders almost all the other factors nonsignificant (except for "black" and "south"). In Model 3, an additional indicator of missing information on second child's sex is also included, and all original covariates become nonsignificant. This suggests that if information on sex of children is missing, it will be missing on each of the first three births, rather than being partially complete. This is supported by Model 4, which predicts missing sex of children information on any of the first three births, and which appears virtually identical to the model for missing information on first births. Models 5 adds "age of oldest child" to Model 4, while Model 6 reduces Model 5 to women aged 40 or younger. Models 7 and 8 add duration of current marriage (and thus is restricted only to currently married respondents).

Generally speaking, the results shown in Table 4 indicate that for women overall, information on children's sex is more likely to be missing for non-white women, women who are not currently married, women in the northeast, and less educated women. Information is also more likely to be missing for births that are farther in the past, or if there is a larger number of total births. Additionally, for women 40 and younger, marital duration is negatively associated with incomplete information on children's sex, while educational attainment and age are not relevant for this group. Further, it is more common for information to be missing for all children, rather than for only some of them.

It should also be noted that a model similar to Model 2 was estimated, except that cases with missing sex of first child information were excluded, and an indicator of sex of first child was added (not shown). The results of this supplemental model were virtually identical to Model 2 in that "black" and "south" were the only significant determinants. Missing information on subsequent children was not significantly related to the sex of the first child (p = .21). This is an important feature to note; completeness of information on sex composition of children is not significantly related to the sex of previous children, thus ruling out the possibility of a bias in information based on sex of children. Further, because the factors associated with missing information appeared to be similar in each survey, they should not lead to different biases across sources (i.e., each survey is consistently affected the same way)

References

Allan, B.B., R. Brant, J.E. Seidel, and J.F. Jarrell. 1997. "Declining Sex Ratios in Canada." *Journal of the Canadian Medical Association* 156: 37-41.

Ben-Porath, Y. and F. Welch. 1976. "Do Sex Preferences Really Matter?" *Quarterly Journal of Economics* 90: 285-307.

Bongaarts, John and Robert G. Potter. 1983. *Fertility, Biology, and Behavior*. New York: Academic.

Campbell, Donald T. and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Dickinson, Heather O. and Louise Parker. 1996. "Why is the Sex Ratio Falling in England and Wales?" *Epidemiology and Community Health* 50: 227-228.

Edwards, A.W.F. 1960. "The Meaning of Binomial Distribution." *Nature* 186: 1074.

--------. 1966. "Sex Ratio Data Analysed Independently of Family Limitation." *Annals of Human Genetics* 29: 337-347.

Greenberg, R.A., and C. White. 1967. "The Sexes of Consecutive Sibs in Human Sibships." *Human Biology* 39: 374-404.

Jacobsen, R., H. Møller, and A. Mouritsen. 1999. "Natural Variation in the Human Sex Ratio." *Human Reproduction* 14: 3120-3125.

James, William H. 1975. "Sex Ratio in Twin Births." *Annals of Human Biology* 4: 365-378.

--------. 1987. "The Human Sex Ratio. Part 1: A Review of the Literature." *Human Biology* 59: 721-752.

--------. 2000. "The Variation of the Probability of a Son Within and Across Couples." *Human Reproduction* 15: 1184-1188.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks: Sage Publications, Inc.

Maconochie, Noreen, and Eve Roman. 1997. "Sex Ratios: Are There Natural Variations Within Human Population?" *British Journal of Obstetrics and Gynaecology* 104: 1050-1053.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Møller, H. 1996. "Change in Male:Female Ratio Among Newborn Infants in Denmark." *Lancet* 348: 828-829.

Shyrock, Henry S., Jacob S. Seigel and Associates. 1976. *The Methods and Materials of Demography*. Condensed Edition by Edward G. Stockwell. San Diego: Academic Press.

Table 1. Relative Probability of Having a Male Baby By Sex(es) of Preceding Siblings, Second to Fourth Order Births (RPB = Relative Probability of a Boy) (CPS 1980, 1985, 1990, and 1995)

	2nd Births		3rd Births		4th Births	
	# births	RPB	# births	RPB	# births	RPB
Sex of immediately preceding sibling						
Male (reference)	40647 (51.01)	1.00	20369 (51.37)	1.00	9326 (51.28)	1.00
Female	37503 (51.01)	0.98	19805 (50.21)	0.95 *	9028 (49.89)	0.95
	chi square 2.76 (1df) (p=.097)		chi square 5.43 (1df) (p=.020)		chi square 3.53 (1df) (p=.060)	
Genders of all preceding siblings						
All Male (reference)			11154 (51.68)	1.00	2674 (50.75)	1.00
All Female			9990 (51.21)	0.98	2287 (50.02)	0.97
Mixed			18920 (50.08)	0.94 **	13283 (50.7)	1.00
			chi square 8.04 (2 df) (p=.018)		chi square .376 (2 df) (p=.829)	

Table 2. Probability of Having a Male Child Among First to Fourth Order Births, By Selected Characteristics of the Birth

	# Births	% male	Sex Ratio	Odds Ratio
Birth Order				
1 (reference)	104712	51.75	103.5	1.00
2	78363	50.74	101.5	0.96 ***
3	40323	50.82	101.6	0.96 **
4	18533	50.6	101.2	0.95 **
				chi square 24.3 (3df) (p=.0001)
Maternal age (years)				
< 20	42166	51.57	103.1	1.01
20-24 (reference)	92884	51.32	102.6	1.00
25-29	68889	50.91	101.8	0.98
30-34	28976	51.03	102.1	0.99
>=35	9016	50.53	101.1	0.97
				chi square 7.1 (3df) (p=.1323)
Year of birth				
1960-1964	30176	51.24	102.5	1.00
1965-1969	32359	51.31	102.6	1.00
1970-1974	35284	51.05	102.1	0.99
1975-1979	36729	50.83	101.7	0.98
1980-1984 (reference)	29207	50.95	101.9	1.00
1985-1989	19973	51.35	102.7	1.00
1990-1994	9586	50.72	101.4	0.98
				chi square 4.2 (6df) (p=.648)
Season of Birth				
Spring (Mar-May) (Reference)	59922	51.2	102.4	1.00
Summer (Jun-Aug)	62588	51.3	102.6	1.00
Autumn (Sept-Nov)	61494	50.94	101.9	0.99
Winter (Dec-Feb)	57865	51.3	102.6	1.00
				chi square 2.1 (3df) (p=.55)
Race				
White (reference)	188219	51.23	102.5	1.00
Black	26928	50.56	101.1	0.97 *
Hispanic	17688	51.23	102.5	1.00
Other	9096	52.02	104.0	1.03
				chi square 6.84 (3df) (p=.077)
Region				
Northeast (reference)	52837	51.29	102.6	1.00
Midwest	60053	51.08	102.2	0.99
South	75869	51.04	102.1	0.99
West	53172	51.4	102.8	1.00
				chi square 2.04 (3df) (p=.56)
Multiple Birth				
Singleton (reference)	237940	51.18	102.4	1.00
Multiple	3991	51.06	102.1	1.00
				chi square .022 (1df) (p=.881)
Education				
Less than Highschool	44046	51.03	102.1	1.00
Highschool (reference)	60618	51.01	102.0	1.00
Some College	99241	51.25	102.5	1.01
Degree	17197	51.57	103.1	1.02
Post-Graduate	20829	51.38	102.8	1.02
				chi square 2.65 (4df) (p=.619)

Table 3. Percent of Female Respondents with Missing Sex of Child (1-3) Information By Parity: CPS 1985, 1990, 1995

1985 CPS

	Sex of Child		
PARITY	1st Child	2nd Child	3rd Child
1	17.22	--	--
2	16.82	18.27	--
3	16.98	17.21	19.37
4	18.04	18.44	19.16
5	19.18	19.49	20.17
6	19.29	19.8	21.71
7	18.85	18.65	20.38
8	21.09	22.18	23.64
9	24.62	25.64	27.69
10	18.18	19.7	21.21

1990 CPS

	Sex of Child		
PARITY	1st Child	2nd Child	3rd Child
1	13.98	--	--
2	12.83	14.61	--
3	14.14	14.34	16.02
4	15.02	15.62	16.4
5	18.94	20.32	21.71
6	16.82	16.98	17.92
7	17.82	18.10	18.1
8	18.37	20.41	20.92
9	21.05	21.93	22.81
10	16.33	17.01	19.73

1995 CPS

	Sex of Child		
PARITY	1st Child	2nd Child	3rd Child
1	16.9	--	--
2	13.76	13.9	--
3	16.11	16.4	16.41
4	17.8	18.21	18.44
5	19.67	20.57	20.8
6	20.91	21.79	22.32
7	23.84	25.62	26.69
8	21.32	23.53	25
9	17.65	20.59	20.59
10	28.13	29.69	31.25

Table 4. Logistic Coefficients Predicting Missing "Sex of Child" Information, CPS 1990, 1995 (Unless Noted, Coefficients are Significant at *p* < .001)

	[1] First Child Missing	[2] Second Child Missing	[3] Third Child Missing	[4] Any Child Missing	[5] Any Child Missing	[6] Any Child Missing (age<41)	[7] Any Child Missing (age<41)	[8] Any Child Missing (age<41)
Race								
White (reference)								
Black	0.355	.340 *	not sig	0.357	0.335	0.178	0.524	0.357
Hispanic	0.110 *	not sig	not sig	0.089 *	.113 **	not sig	not sig	not sig
Other	0.432	not sig	not sig	0.412	0.443	0.448	0.404	0.403
Marital Status								
Married (reference)								
Divorced	0.313	not sig	not sig	0.267	0.259	0.317	---	---
Widowed	0.203	not sig	not sig	0.203	0.184	.344 *	---	---
Never Married	0.590	not sig	not sig	0.545	0.465	0.716	---	---
Number of Marriages	not sig	not sig	not sig	not sig	not sig	0.144	0.163	-0.126 *
Region								
Northeast (reference)								
Midwest	-0.124	not sig	not sig	-0.126	-0.147	-0.185	not sig	-0.133 *
South	-0.237	-0.387 **	not sig	-0.249	-0.277	-0.293	-0.222	-0.257
West	-.104 **	not sig	not sig	-0.101 **	-0.121	not sig	not sig	not sig
Educational Attainment	-0.023	not sig	not sig	-0.026	not sig	0.052	-0.030 *	not sig
Parity	not sig	not sig	not sig	0.034	not sig	not sig	0.063	not sig
Age	0.017	not sig	not sig	0.017	-0.009	-0.052	0.028	not sig
Age of Oldest Child	---	---	---	---	0.028	0.072	---	0.082
Years in Current Marriage	---	---	---	---	---	---	-0.015 **	-0.046
First Child Missing	---	9.070	3.891	---	---	---	---	---
Second Child Missing	---	---	5.875	---	---	---	---	---
Intercept	-2.423	-4.379	-4.498	-2.332	-1.815	-1.552	-2.845	-1.962
N	66,084	50,557	26,475	66,084	66,084	31,848	27,603	27,603

* *p* < .05 ** *p* <.01