Marital Coital Frequency: Unnoticed Outliers and Unspecified Interactions Lead to Erroneous Conclusions

Author(s): Joan R. Kahn and J. Richard Udry
Published by: American Sociological Association
Stable URL: http://www.jstor.org/stable/2095496
Accessed: 07/06/2008 11:49

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=asa.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We enable the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.
MARITAL COITAL FREQUENCY: UNNOTICED OUTLIERS AND UNSPECIFIED INTERACTIONS LEAD TO ERRONEOUS CONCLUSIONS*

JOAN R. KAHN  J. RICHARD UDRY
Carolina Population Center, The University of North Carolina at Chapel Hill

In a recent application of fixed-effects modeling, Jasso estimates age and period effects on marital coital frequency that take into account cohort effects. Although the application is an innovation for researchers interested in identifying net age and period effects, the results are in some respects problematic. Jasso claims that over the period 1970–1975 (1) net of cohort and age effects, average marital coital frequency declined significantly by 50 percent (p < .01), and (2) net of period and cohort effects, coital frequency increased with wife’s age (p < .05).

Both of these findings are difficult to reconcile with previous findings in the literature on sexual behavior. The negative period effect is inconsistent with findings for the 1965–1970 interval, which showed a significant increase in marital coital frequency (Westoff, 1974). It is surprising that in the 1970–1975 period this trend should reverse itself, especially since public opinion surveys at the time showed that women’s attitudes about sexual behavior became more liberal (The Virginia Slims American Women’s Opinion Poll, 1974). As far as the wife’s age effects are concerned, Jasso’s findings contradict past studies that show a negative age effect for both wives and husbands, as well as negative marital duration effects (James, 1974; Kinsey et al., 1948, 1953; Udry and Morris, 1978; Udry, 1979; Udry et al., 1982).

We believe that Jasso’s conclusions are misleading for two reasons: (1) she did not check her data for outliers and, more importantly, (2) she did not notice an important interaction with marital duration. In the pages that follow, we show that dropping eight of 2,063 observations wipes out the significance of the wife’s age effect and reverses the sign of the husband’s age effect. We also show that Jasso’s negative period effect and positive wife’s age effect are artifacts of the well-known fact that marital sexual activity declines dramatically during early marriage (James, 1981; Sen Gupta, 1969; Westoff and Westoff, 1971). Specifically, we show that at marital durations of greater than two years, neither period, wife’s age, husband’s age, nor marital duration has any effect on marital coital frequency.

Jasso combines a fixed-effects approach with nonlinear transformations to estimate age and period effects that are net of cohort effects. She claims that in past studies estimates of age and period effects were biased because they did not take into account the potentially compounding effect of cohort membership. (In light of the age-period-cohort [APC] identity, most studies simply ignored cohort membership.) Jasso proposes a model that both breaks the APC identity and controls for unobservable heterogeneity. The fixed-effects approach uses panel data to control for couple-specific, time-invariant (i.e., fixed) unobservable covariates. It is often referred to as dummy variable regression because the unobservable couple effects are captured by a series of couple dummies (i.e., one for each couple).

The model underlying Jasso’s analysis is the following:

\[
C_{ij} = \beta_1 \ln W_{ij} + \beta_2 H_{ij} + \beta_3 M_{ij} + \eta \text{PERIOD}_j + \gamma_1 \text{WCOH}_i + \gamma_2 \text{HCOH}_i + \gamma_3 \text{MARDATE}_i + \alpha_x X_{ijk} + \delta_3 D_i + \varepsilon_{ij}
\]

where \(i = 1,2,\ldots,N\) couples, \(j = 1,2,\ldots,K\) periods.

CF represents the coital frequency in the past four weeks for couple \(i\) in period \(j\). \(W_{ij}\) is wife’s age, \(H_{ij}\) is husband’s age. \(M_{ij}\) is marital duration, \(\text{PERIOD}_j\) is date of interview (either 1970 or 1975), \(\text{WCOH}_i\) is wife’s birth cohort, \(\text{HCOH}_i\) is husband’s birth cohort, \(\text{MARDATE}_i\) is the date of marriage, \(X_{ijk}\) is a series of control variables (see Jasso, 1985), and \(D_i\) represents the couple-specific dummy variables. (Note that all variables representing dates, including \(\text{PERIOD}_j\), have been coded in decimal years to reflect the actual month of occurrence.)

Equation (1) clearly cannot be estimated as written because of the age-period-cohort identity. That is, each of the age variables is a linear combination of \(\text{PERIOD}_j\) and the cohort variables (e.g., \(W_{ij} = \text{PERIOD}_j - \text{WCOH}_i\)). To break the identities in equation (1), Jasso transforms the age variables by taking the natural logarithm of each. Her model becomes:

\[
C_{ij} = \beta_1 \ln W_{ij} + \beta_2 \ln H_{ij}
\]

* Please address all correspondence to: Joan R. Kahn, Carolina Population Center, University of North Carolina, University Square 300A, Chapel Hill, NC 27514–3997.

The authors gratefully acknowledge the helpful comments of David K. Guilkey, William M. Mason, Michael J. Moore, Ronald R. Rindfuss, C. Suchindran and Amy Ong Tsui.
In theory, equation (2) is an estimable formulation of a fixed-effect age-period-cohort model. However, in practice, it is computationally intractable because of the large number of dummy variables required. It has been shown that, when using only two waves of panel data, equation (2) can be estimated using first differences (i.e., the model is specified for each wave and then one is subtracted from the other).

Taking first differences from equation (2), we have:

\[ \Delta CF_i = \beta_1 (\Delta \lnWAge_i) + \beta_2 (\Delta \lnHAge_i) + \eta (\Delta \text{PERIOD}) + \alpha_k (\Delta X_{ik}) + \Delta \epsilon_i \]

where \( \Delta \) signifies the change between time 1 and time 2.

Note that the cohort terms as well as the couple-specific dummies drop out of equation (3) because their values remain constant in the two waves. Each of the parameters in equation (3) can be interpreted as net of both the cohort effects and the effects of the unobservable couple-specific covariates. In breaking the age-period-cohort identity, Jasso is still unable to estimate cohort effects; rather, she estimates age and period effects that are not confounded with cohort effects. Specifically, she estimates equation (3) for a sample of continuously married women interviewed in both 1970 and 1975 as part of the National Fertility Survey. As described above, she found significant age and period effects that directly contradict past research. Column 1 of Table 1 presents Jasso’s results (Jasso, 1985, Table 4).

Troubled by her substantive findings, we re-estimated equation (3) using the same data. Except for a sample size difference of one observation, we almost perfectly replicated her results.

### Table 1. Fixed-Effects Estimates of the Determinants of Marital Coital Frequency

<table>
<thead>
<tr>
<th></th>
<th>(1) Jasso’s Results</th>
<th>(2) Our Replication</th>
<th>(3) Drop 4 Miscodes</th>
<th>(4) Drop 4 Miscodes &amp; 4 Outliers</th>
<th>(5) Marital Duration ≤ 2</th>
<th>(6) Marital Duration &gt; 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>-0.72***</td>
<td>-0.72***</td>
<td>-0.75***</td>
<td>-0.67***</td>
<td>-3.06**</td>
<td>-0.08</td>
</tr>
<tr>
<td>Log Wife’s Age</td>
<td>27.61**</td>
<td>27.50**</td>
<td>21.99*</td>
<td>13.56</td>
<td>29.49</td>
<td>1.62</td>
</tr>
<tr>
<td>Log Husband’s Age</td>
<td>-6.43</td>
<td>-6.38</td>
<td>1.87</td>
<td>7.87</td>
<td>57.89</td>
<td>5.23</td>
</tr>
<tr>
<td>Log Marital Duration</td>
<td>-1.50***</td>
<td>-1.51***</td>
<td>-1.61***</td>
<td>-1.56***</td>
<td>-1.51*</td>
<td>1.29</td>
</tr>
<tr>
<td>Wife Pregnant</td>
<td>-3.71***</td>
<td>-3.70***</td>
<td>-3.71***</td>
<td>-3.74***</td>
<td>-2.88***</td>
<td>-3.95*</td>
</tr>
<tr>
<td>Child under 6</td>
<td>-0.56**</td>
<td>-0.55**</td>
<td>-0.73**</td>
<td>-0.68**</td>
<td>-2.91***</td>
<td>-0.55**</td>
</tr>
<tr>
<td>Wife Employed</td>
<td>0.67</td>
<td>0.38</td>
<td>0.17</td>
<td>0.23</td>
<td>0.86</td>
<td>0.02</td>
</tr>
<tr>
<td>Husband Employed</td>
<td>-1.28**</td>
<td>-1.27**</td>
<td>-1.29**</td>
<td>-1.10**</td>
<td>-4.11***</td>
<td>-0.38</td>
</tr>
<tr>
<td>R²</td>
<td>.0475</td>
<td>.0474</td>
<td>.0568</td>
<td>.0612</td>
<td>.2172</td>
<td>.0411</td>
</tr>
<tr>
<td>N</td>
<td>2062</td>
<td>2063</td>
<td>2059</td>
<td>2055</td>
<td>243</td>
<td>1812</td>
</tr>
</tbody>
</table>

* p < .10.
** p < .05.
*** p < .01.
results (see column 2 of Table 1). However, after applying standard regression diagnostic techniques, we found serious problems with outliers and specification error. First, in the 1970 wave, four observations were coded as 88, even though the missing code was specified as 99 in the codebook. We are fairly certain that the 88s should have been coded as missing since no other values was greater than 63. In fact, 99.5 percent of the sample had values less than 40. Furthermore, all of the cases with 88 in 1970 had coital frequencies of less than or equal to 10 in 1975, indicating that these couples were probably not highly active.

Deleting the four 88-codes reduced the significance level on the wife's age effect from p < .05 to p < .10, and reversed the sign on the husband's age effect (see column 3 in Table 1). The fit of the model improved substantially from an $R^2$ of .047 to .057. This in itself is not damning evidence against Jasso's analysis, although it indicates that her results are sensitive to specific cases. To determine the potential influence of the other cases, we examined the studentized residuals (Belsley et al., 1980; Bollen and Jackman, 1985; Weisberg, 1980) and found a handful of cases with extremely large residuals. The 88s had by far the largest studentized residuals, ranging between 10 and 12. Four additional cases had studentized residuals between 3 and 7. An examination of the DFBETAs showed that the omission of these observations resulted in large changes in the parameter estimates. These observations correspond to the four observations with coital frequencies of greater than 40 (i.e., they are at the extreme high end of the coital frequency distribution).

Since in regression analysis the emphasis is on generalizability of results, one should be especially skeptical of results that are unduly sensitive to a small number of observations (Bollen and Jackman, 1985). We are convinced that the values of 88 are keypunching errors and should therefore be omitted. We would not argue that the other coital frequency values over 40 were necessarily inaccurately reported. However, since they give undue weight to the far extreme of the distribution, we feel justified in omitting them as well.

By deleting these four additional outliers, the results change significantly (see column 4 in Table 1). The wife's age effect is now longer significant, even with over 2,000 cases and at the generous .10 level. The husband's age effect switches sign from negative to positive, but remains insignificant. Finally, the fit improves from an $R^2$ of .0568 to .0613. Thus, we have shown that the inclusion or exclusion of as few as eight observations can have a profound impact on Jasso's results. However, the fact remains that the period effect is still positive (though the coefficient is not significantly different than zero).

Jasso rightly points out that coital frequency declines most rapidly early in marriage and then only moderately thereafter (p. 235). Figure 1 shows the plot of coital frequency by marital duration for the sample in 1970 and 1975. (Note that by 1975 all respondents had been married for at least five years.) The steep dropoff in the first few years of marriage has been well documented elsewhere (James, 1981; Sen Gupta, Westoff and Westoff, 1971). This leads us to believe that a different process may underlie coital behavior early in marriage than later on, indicating the possibility that the relationship between coital frequency and its determinants may also be different (i.e., suggesting the presence of interactions).

We estimated the model separately for couples with 1970 marital durations of two or fewer years and more than two years and found strikingly different results (see columns 5 and 6 in Table 1). The negative period effect and the positive wife's and husband's age effects are confined to the first two years of marriage. Note that the model fit is considerably better in the first two years of marriage ($R^2 = .2172$ vs. $R^2 = .0411$), though this may be due in part to the small sample size. The difference between the two models is highly significant.

---

1 The studentized residual for an observation is the OLS residual divided by the standard error estimated without that observation. Belsley suggests that the studentized residual is distributed closely to a t-distribution with $N - p - 1$ degrees to freedom (where $p$ is the number of variables in the model).

2 The DFBETA indicates how an individual coefficient changes when a case is omitted.

3 One additional observation had a value of coital frequency of greater than 40, but since its value in 1970 was 63 and in 1975 it was 60, the first difference (i.e., $63 - 60 = 3$) was not unusually large. Of the other four observations with frequencies over 40, all had very large first differences (which explains their excessive influences in the analysis).
significant as evidenced by strong interaction effects (not shown).

The interactions with marital duration indicate that a very different process underlies the behavior of recently married couples than couples married for longer durations. However, we are convinced that the model for those with marital durations of less than two years is not interpretable. Recall that this subsample is restricted to couples married 0–2 years in 1970 and, hence 5–7 years in 1975. The period effect for this group indicates the net effect of the passage of one year, holding all else (including marital duration) constant. However, when marital duration is held constant, there is no variability in period, other than the actual month of interview. (Note that period was coded in terms of the decimal-year of interview, so that an interview in March, 1970 would have a value of 0.7025.) Thus, the period coefficient captures the effect of being interviewed in January versus June of a year rather than the effect of being interviewed in 1970 versus 1975. With no a priori theory of monthly variation in marital sexual behavior, we have no explanation for the significant negative period effect.

We should note that these findings are in no way dependent on the omission of the 8 miscodes and outliers. We found that, whether based on 2055 or 2063 observations, the strong interactions with marital duration persist. We must conclude, therefore, that the application of the fixed-effects model to the present problem must be limited to the couples married for more than two years and not applied to newly married couples.

In summary, after reanalyzing the same data used by Jasso, we come to the following altered conclusions:

1. When outliers are removed, neither husband’s nor wife’s age is significantly related to frequency of intercourse, though period and marital duration effects remain.
2. When the model is respecified separately by duration of marriage, the age, duration, and period effects disappear for couples married more than two years in 1970 (88 percent of the sample).

Jasso’s only new finding to hold up after reexamination is the period effect, yet even this effect is limited to couples married for two or fewer years in 1970. Furthermore, as described above, the period effect for this group has no intuitive meaning.

We are glad that Jasso has introduced us and other sociologists to fixed-effects models, since they are likely to have many useful applications. In addition, Jasso’s use of nonlinear transformations clearly breaks the age-period-cohort identity, though not without making considerable assumptions about higher order terms. We feel, however, that researchers should be especially skeptical of counterintuitive results. As we have shown above, as few as eight observations can have a profound impact on results. Furthermore, the model specification can easily obscure the underlying behaviors. Thus, it is important to be extremely thorough before accepting new results that contradict all previous research.

REFERENCES


