
RESPONSE TO ERMISCH, MARTIN, AND WU*

JO ANNA GRAY, JEAN STOCKARD, AND JOE STONE

Our recent paper in Demography (Gray, Stockard, and Stone 2006) has attracted the close scrutiny of several prominent academics. Three sets of formal comments, authored independently by Ermisch, Martin, and Wu (EMW), appear in this issue of Demography. In this response we argue that the analysis and evidence of our 2006 paper have withstood the scrutiny of EMW. In particular, we find that a substantial part of the rising share of nonmarital births since 1970 is due to a selection effect associated with marriage. This same selection effect also explains how birth rates could rise in both groups, even though their combined birth rate did not. In sum, though we appreciate the opportunity to expand on several key aspects of our 2006 article, we see no reason to substantially revise any of our major conclusions based on the EMW comments.

We are flattered that our recent article in this journal (Gray, Stockard, and Stone 2006; hereafter referred to as *GSS*), has attracted such close attention from Ermisch, Martin, and Wu (hereafter *EMW*). While we appreciate the opportunity to expand on several key aspects of our article, we see no reason to substantially revise any of our major conclusions based on the EMW comments. Reading EMW, one might think that we had proposed the demographic equivalent of Newton's second law of thermodynamics—the existence of a universal phenomenon, manifest in identical form in all places, for all groups, during all time periods, regardless of circumstances. It will be helpful, then, to review briefly the central points in *GSS* before turning to the major EMW comments, along with our responses.

A major objective in *GSS* was to offer and test an explanation for an apparent paradox: among black women and white women aged 20 to 39, birth rates increased sharply for unmarried women during the period 1974–2000. But they also increased for *married* women, and yet the total birth rate for married and unmarried women combined was essentially unchanged. Because the total birth rate did not change, it seems obvious by inspection that the rises in unmarried and married birth rates could not have come from a general rise in fertility among women aged 20–39. We recognized these patterns as an example of a phenomenon called “Simpson's paradox,” often illustrated by a joke, as told at Harvard, that when a student transfers from Harvard to Yale, mean intelligence rises at both places. Both means rise *not* because the average intelligence of the combined student bodies changed, of course, but because the *composition* of the student body changed at each school. The implication of the joke is that the intelligence of a student who chooses to transfer from Harvard to Yale must be *below* the mean at Harvard but *above* the mean at Yale, so both means rise when the student transfers.

In the case of birth rates, *GSS* argued that between 1974 and 2000, sharp increases in the proportion of women who were single, which we term the *single share*, or *Su*, changed the composition of the pools of married and unmarried women. The rising single share had a selection effect on the pools of married and unmarried women akin to the hypothetical student transfer from Harvard to Yale. Women with target fertility below the average for married women but above the average for unmarried women became less likely to marry than previously; thus, mean birth rates for both groups rose over the period, even as the total birth rate was flat.

*Jo Anna Gray, Department of Economics, University of Oregon, Eugene, OR 97403-1285; e-mail: jgray@oregon.uoregon.edu. Jean Stockard, Department of Planning, Public Policy, and Management, University of Oregon. Joe Stone, Department of Economics, University of Oregon. The authors thank Stephen Haynes and Robert O'Brien for their comments on an earlier draft of this response.

The empirical tests reported in GSS focus on the implications of this selection effect for the ratio of unmarried births to total births—referred to as the *nonmarital fertility ratio*, *NFR*, in our article. The bit of algebra included in GSS was intended only to “highlight and illustrate” those implications, not to suggest that the effect is the only factor, or even a dominant one, in determining birth rates or *NFR* for all groups or time periods. Nevertheless, using age- and race-specific panel data, GSS found parameter values strikingly consistent with those predicted by our illustrative model, and a dominant role for the selection effect of the single share in determining *NFR* for the particular groups and period we studied.

ERMISCH AND STATISTICAL CHALLENGES

While the particulars vary, EMW share a common line of argument: (1) factors common to both *NFR* and *Su* caused the two measures to rise together, and consequently, (2) the selection effect of *Su* on *NFR* found by GSS is spurious. Ermisch supported this argument primarily with three challenges to the statistical validity of our estimates. We argue that all three are invalid.

Ermisch’s central argument was that *NFR* and Su^2 are nonstationary and, as a result, our estimates are inconsistent and especially vulnerable to spurious regression. This argument is invalid as applied to GSS. First, *NFR* and *Su* are shares, bounded between 0 and 1; neither can exhibit nonstationary behavior in a sufficiently long sample. Second, the unit root tests for *NFR* and Su^2 reported by Ermisch are based on only 23 years of annual data (1980–2002). As Ermisch himself acknowledged, standard unit-root tests of the sort he used have weak power. The problem is particularly marked for highly persistent series, and further compounded when sample periods are short.

Third, even when the variables in a regression *are* nonstationary, inconsistent estimates arise only if the variables are not cointegrated (Engle and Granger 1987; Hamilton 1994). Seven of the eight paired (*NFR* and Su^2) cross-section time series examined in GSS (2006) exhibited unit roots, even over the longer sample periods estimated in that article. For five of those seven, however, the data are consistent with the presence of a single cointegrating vector at significance levels of 5%. Unit roots are also present when the data *are* grouped by race, but, again, we found that *NFR* and Su^2 are cointegrated for both blacks and whites, which is broadly consistent with results reported in GSS (2005). Thus, Ermisch is simply wrong in asserting that the statistical tests reported in GSS (2006) are “invalid because the variables in the analysis are not stationary time series” (p. 193). Finally, if unit roots and spurious correlation were responsible for our regression results, one might expect the addition of a time trend or period effects to significantly alter the results. As discussed below, they do not.

Ermisch also objected to the estimation procedure employed in GSS, suggesting that our estimates of the effect of the single share on *NFR* are inconsistent and our test results are “highly suspect” because we did not use seemingly unrelated regression (SURE). On this point, Ermisch is certainly incorrect: SURE affects only estimates of standard errors, not of structural parameters as Ermisch implied, and in our case, the effect on standard errors was inconsequential. In GSS, we used panel-corrected standard errors (PCSE), which, depending on the choice of estimator, can incorporate adjustments for heteroskedasticity, for common shocks (as in SURE), and/or for autocorrelation. We emphasized only this last adjustment in GSS because it is the only one that made much difference to the standard errors.

These points are illustrated in Table 1 in this response. Column 1 presents “white period” estimates of the key relationship developed and tested in GSS over the longer of the two sample periods reported by Ermisch (1965–2000). These estimates correct for heteroskedasticity and autocorrelation. Column 2 presents “cross-section SURE” estimates, which account for heteroskedasticity, autocorrelation, and contemporaneous correlations across the age-race groups, as Ermisch suggested. Note that the coefficients are identical in

Table 1. Nonmarital Fertility Ratio for Women Aged 20–39: 1965–2002

Variable	White Period (1)	Cross- Section SURE (2)	Cross- Section SURE (3)	Cross- Section SURE (4)	Cross- Section SURE (5)
Constant	–0.0143 (0.0072)	–0.0143 (0.0082)	–0.0166 (0.0079)	0.0044 (0.0096)	0.0219 (0.0115)
Su^2	1.0099 (0.0260)	1.0099 (0.0281)	0.9271 (0.0353)	0.9428 (0.0344)	1.1461 (0.0437)
Su					–0.1533 (0.0370)
Time			0.0009 (0.0003)		
Age-Race Fixed Effects	yes	yes	yes	yes	yes
Period Effects	no	no	no	yes	no
Adjusted R^2	.9851	.9851	.9856	.9896	.9855
Number of Observations	285	285	285	285	285

Notes: Standard errors are in parentheses. The dependent variable is the nonmarital fertility ratio by race and five-year age interval. Our data are not available until 1968 for black women and 1969 for white women aged 35–39. See GSS (2006) for further explanation of the variables and data.

columns 1 and 2. SURE estimation has no effect on estimates of the coefficients; furthermore, the changes in standard errors are inconsequential.

Column 3 of Table 1 augments the baseline specification with a time trend, while column 4 adds period effects instead. Neither modification alters the conclusions presented in GSS: the coefficient on Su^2 is only slightly affected and remains near 1. Accordingly, the data do not support Ermisch's contention that the results reported in GSS are the result of spurious correlation produced by common trends in NFR and Su^2 .

A final statistical issue arose when Ermisch proposed and implemented his own tests of the GSS model. Ermisch tested prediction errors generated by our stylized *theoretical* model (not its estimated counterpart) for unit roots, failed to reject a unit root in most cases (see Table 1 of Ermisch), and interpreted this result as evidence against the GSS selection effect. The unit-root tests that Ermisch employed have notoriously weak power, especially in relatively short time series, such as those chosen by Ermisch. New, more powerful tests for panel data exploit both the cross-section and time-series structure of panel data.¹ Of course, taking advantage of longer sample periods increases power as well. Thus, whereas Ermisch failed to reject unit roots in all but one of the series reported in Table 1 of his comment, we reject unit roots at the 1% level using the more powerful panel Dickey-Fuller test over the period 1957–2000, the time period employed in GSS. (See Table 2 of this response.) Perhaps surprisingly, unit roots are also rejected at the 1% level if we shorten the sample to match Ermisch's (1980–2002).²

In view of the limited power of the unit-root tests he employed, Ermisch went on to assert that even if his constructed error term “does not have a unit root, its strong degree of persistency . . . contradicts the GSS theory” (p. 197). We disagree. Clearly, evidence of a selection effect does not imply the absence of other effects—a point addressed further in the next section. Social and economic factors may influence NFR through channels other

1. These tests are available in such widely used and relatively “friendly” software packages as E-views.

2. Our reservations regarding the test results reported in Table 1 of Ermisch also apply to those reported in Table 2 of Ermisch.

Table 2. Panel ADF Test Statistics

Null Hypothesis	Number of Observations	Lags	Chi-square Test Statistic
<i>(NFR - Su²)</i> Has a Unit Root			
Sample period: 1957–2000	274	0–7	49.40**
Sample period: 1980–2002	272	0–3	37.62**
<i>[UBR / MBR - Su / (1 + Su)]</i> Has a Unit Root			
Sample period: 1957–2000	173	0–7	60.72**
Sample period: 1980–2002	172	0–3	45.81**

Notes: See Table 1 notes. Lag length is based on the Schwartz Information Criterion.

** $p < .01$

than Su and its selection effect. These influences may be strong and persistent. Thus, we emphatically dispute Ermisch's assertion that persistent (serially correlated) deviations of NFR from the predictions of the GSS model contradict the model.

In conclusion, we find all three of Ermisch's statistical objections unpersuasive.

ALTERNATIVE HYPOTHESES AND TESTS

Both Ermisch and Martin suggested that factors other than the selection effect described in GSS are important—at times perhaps dominant—in explaining birth rates and birth shares over subsets of the time periods reported by GSS. Martin described Ermisch as showing “that the GSS model alone cannot explain all the variation in the ratio of nonmarital to marital birth rates from 1974 to 2000” (p. 203–204). We do not disagree; GSS does not claim otherwise, nor did we intend to imply otherwise. Our goal in GSS was to call attention to a selection effect that is commonly overlooked in studies of fertility behavior, illustrate its implications for measured birth rates and shares, and then demonstrate that the effect could be empirically important. Because we expected additional factors to be important in explaining birth rates and ratios at various points, we were surprised at the apparent power of the selection effect in both the relatively long samples examined in our article. Our conclusion in GSS (2006) was that valid tests of the importance of *other* factors in explaining birth rates and ratios should take account of this effect—a conclusion reaffirmed by the present exchange, in our view.

Ermisch went further, however, and asserted that the association between NFR and Su^2 documented in GSS is spurious. The implication is that the GSS results are due either to common trends in the data (addressed in columns 3 and 4 of Table 1 in this response) or to other factors that might cause both NFR and Su^2 to rise together. A challenge to alternative explanations of the joint behavior of NFR and Su^2 is that over the period 1974–2000, the marital birth rate rose along with the nonmarital birth rate, even though the total birth rate remained unchanged. While changes in social attitudes or other factors may or may not have had the effects claimed by Ermisch (and repeated by Martin), they do not explain how *both* married and unmarried birth rates could rise in the absence of a rise in the total birth rate. To reconcile these paradoxical patterns, we believe the selection effect GSS identified is required.

Martin took a tack similar in spirit to Ermisch's when he noted that the definition of NFR includes Su , so that NFR “will vary with Su^2 to some extent even if the GSS model is incorrect” (p. 203). This is correct, as far as it goes. As Eq. (1) below shows, NFR can be expressed as the product of Su and the ratio UBR / TBR —a definitional relationship. If the GSS selection effect is not present (UBR / TBR is independent of Su), the relationship between NFR and Su should be linear. If the GSS selection effect is present, the relationship should be nonlinear—indeed, quadratic if all the assumptions of the GSS illustration

hold. GSS found a strong, apparently quadratic, relationship between NFR and Su . But could it be simply the spurious result of the linear relationship between NFR and Su evident in Eq. (1)?

The answer is provided in the final column of Table 1, which reports the results of including both Su and Su^2 in a statistical model of NFR . If the selection effect is present in the form hypothesized in GSS, the estimated coefficient on Su^2 should be significantly positive and near unity, while the coefficient on Su should be near zero. On the other hand, if the selection effect is unimportant, the coefficient on Su should be 1, and the coefficient on Su^2 should be 0. The coefficient on Su^2 in column 5 remains significantly positive and near unity, even with Su accounted for separately in the regression. Furthermore, Su enters with a negative, not positive, coefficient alongside Su^2 , clearly refuting Martin's suggestion that Su^2 appears important in the GSS analysis only because it picks up the effects of a variable (Su) that we did not include in the regression.

Both Ermisch and Martin proposed alternative tests of the GSS selection effect that focus on the ability of the effect to explain movements in the ratio of the unmarried birth rate to the married birth rate, denoted UBR / MBR . We are puzzled by the focus on this measure for two reasons. First, it is not, as claimed by Ermisch, the "more fundamental" relationship in GSS. Indeed, the measure never arises in developing the simple model presented in our article. The "fundamental" relationship underlying the GSS selection effect is a relationship commonly used in demographic decompositions of NFR :

$$NFR = Su(UBR / TBR). \quad (1)$$

As Eq. (1) shows, NFR differs from Su only to the extent that the childbearing behavior of unmarried women as a subpopulation deviates from that of the population as a whole. Substituting our model's prediction for UBR / TBR into Eq. (1) produces the key equation in GSS: $NFR = Su^2$.

Given the claims of our article, we would have expected a skeptic to challenge the much cleaner GSS predictions for UBR / TBR and MBR / TBR individually:

$$UBR / TBR = Su \quad (2)$$

$$MBR / TBR = (1 + Su). \quad (3)$$

Had Ermisch or Martin chosen to focus on these more obvious implications, they might have found, as we did, that the predictions of our model, meant only as an illustration, hold up to the data remarkably well.

We are also puzzled by the focus on UBR / MBR because of its particular vulnerability to measurement error. Significant errors in estimating the size of the unmarried population (noted by Ermisch in his comment) mean that both UBR and MBR individually are subject to substantial measurement error. As the ratio of two ratios, each measured with substantial error, UBR / MBR is particularly volatile. Even so, formal statistical estimates of the model prediction for this measure—that is, $UBR / MBR = Su / (1 + Su)$ —over reasonably long sample periods (1957–2000 and 1968–2002) yield parameter estimates that are strikingly consistent with the predicted values of 0 for the constant term and 1 for the coefficient on $Su / (1 + Su)$.³ (See Table 3 in this response.) Furthermore, errors constructed by taking the difference between UBR / MBR and $Su / (1 + Su)$ do not exhibit unit roots in panel tests of nonstationarity applied over longer sample periods, contrary to the conclusions drawn by Ermisch. (See Table 2 in this response.)

3. The joint hypothesis that the constant and slope coefficient are within 5% of their predicted values is not rejected.

Table 3. *UBR / MBR, Women Aged 20–39*

Explanatory Variables	1957–2000	1965–2002
Constant	0.0014 (0.0254)	0.0034 (0.0325)
$Su / (1 + Su)$	0.9473 (0.0848)	0.9693 (0.1033)
Age-Race Fixed Effects	yes	yes
Period Effects	no	no
Adjusted R^2	.8917	.8711
Number of Observations	292	284

Notes: See Table 1 notes. The dependent variable is *UBR / TBR* by race and five-year age interval. Estimation is period SURE.

The issues raised by measurement error are particularly acute in the informal tests proposed by Ermisch and Martin, both of whom compared arithmetic changes in *NFR* and *UBR / MBR* over intervals much shorter than the time periods examined in GSS—in some cases as short as 10 years. Arithmetic comparisons over short intervals can be highly problematic when the data examined are subject to substantial measurement error because movements over short intervals can be easily dominated by these errors, rather than by fundamental behaviors. This problem is especially acute for *UBR / MBR* for the reasons we already discussed. An advantage of our formal statistical approach is that errors in the dependent variable do not bias estimated parameters as long as they are random.

OTHER ISSUES RAISED BY EMW

Regrettably, despite a reasonably generous allotment of journal space, we will not be able to address all of the remaining issues raised by EMW. In this section, we have selected several from among the most interesting for further discussion.

Ermisch and International Comparisons

In buttressing his argument of spurious correlation, Ermisch appealed to common international trends in *NFR*. Figure 4 of his comment presents data on *NFR*, but not Su^2 , for eight Organization for Economic Cooperation and Development (OECD) countries and U.S. whites. Ermisch interpreted the rising values of *NFR* across these countries as evidence of the influence of common factors other than the GSS selection effect. Our reading of the international data is quite different. While we would not necessarily expect the strength of the selection effect identified in GSS (2006) to be as strong in other countries or circumstances, we would be surprised if it played *no* significant role in determining *NFR* in European countries. Indeed, plots of both *NFR* and Su^2 for 19 European countries over two census years, 1991 and 2001, show a strikingly similar pattern to that found in the United States (see Figure 1).⁴ The strong visual correspondence between *NFR* and Su^2 is confirmed by the formal statistical estimates presented in Table 4. With fixed effects to account for idiosyncratic features of both the context and the data for individual countries, the intercept term reported in Table 4 is trivially small, and the slope coefficient on the selection term Su^2 is both significant and near unity. Again, though, our point is only that the selection effects appear to be an important, not exclusive, factor in determining *NFR*.

4. Data are from Eurostat. *NFR* is for the total population of women in each country, and Su^2 refers to women aged 15 to 44.

Figure 1. *NFR* and *Su*² for 19 European Countries, 1991 and 2001

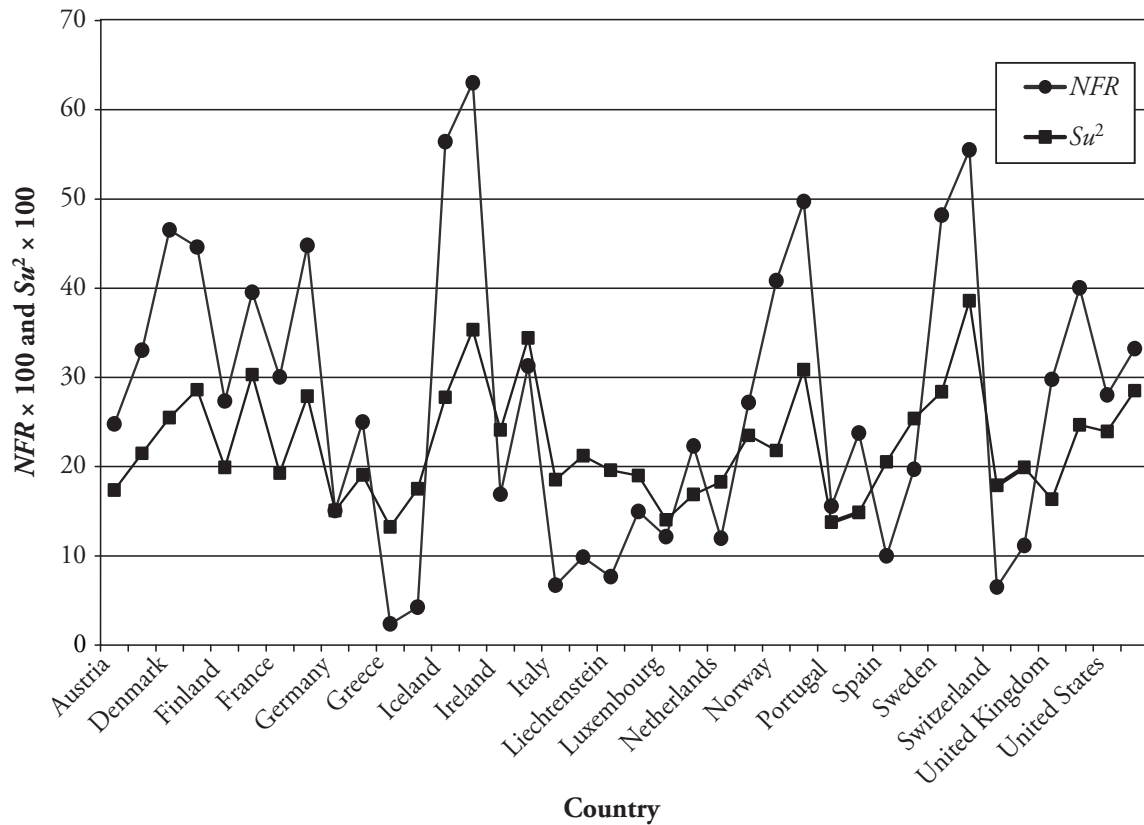


Table 4. International Evidence From 19 Countries on *NFR* and *Su*²

Variable	Estimates
Constant	-0.0176 (0.0538)
<i>Su</i> ²	1.2847 (0.2393)
Adjusted <i>R</i> ²	.9592
Number of Observations	38

Notes: The 19 countries are indicated in Figure 1. The dependent variable is the nonmarital fertility ratio for women of all ages in 1991 and 2001. *Su*² is the square of the single share of women aged 15–44 in 1991 and 2001. Estimation is period SURE with individual country effects.

Martin and Fertility Distributions

Aside from issues raised separately by Ermisch and treated in this response, Martin’s principal objection is that empirical distributions of time spent in marriage, conditional on fertility outcomes, are inconsistent with our model. In Table 3 of his comment, Martin presented figures for two cohorts of women, one 15 years older than the other, and argued

that “[t]he main discrepancy between Table 3 and the assumptions of the GSS model is in trends across cohorts” (p. 205). While we are impressed by Martin’s ingenuity, he has not quite gotten it right. His objection is based on the mistaken impression that the fertility distribution from which an individual’s target level of fertility is drawn, summarized by the parameter P in our stylized model, is “fixed” across cohorts. Although we do assume that P is “given” for a particular cohort (the population to which our model applies), we do not assume that P is invariant across cohorts. This was actually highlighted in GSS (2006) in the final of the four predictions drawn from the simple model we developed and in the subsequent discussion.

Wu and Model Properties

Wu argued that a satisfactory demographic relationship must satisfy Eq. (4) below, and that the GSS model fails this criterion. We argue that his criterion is not appropriate in many, if not most, models; furthermore, it is violated by relationships that are self-evidently correct. Wu’s proposed condition may be represented as

$$y = f(x) = f(x_1) + f(x_2), \quad (4)$$

where the subscripts represent mutually exclusive and exhaustive groups within a larger population (identified by the absence of a subscript). In Wu’s application of the condition, y and x are identified with NFR and Su for some larger population, such as blacks and whites combined. The subpopulation variables x_1 and x_2 are the single shares for blacks and whites, Su_1 and Su_2 . Finally, in the GSS illustration, $f(x)$, $f(x_1)$, and $f(x_2)$ are equal to Su^2 , Su_1^2 , and $Su_2^2(x)$, respectively. Substituting these values into Eq. (4) above gives Eq. (1) of Wu’s comment. The resulting condition is obviously violated. Thus, as Wu asserted, the GSS model fails Wu’s condition.

Wu’s condition will be violated by any nonlinear model, even a log-linear model, since the log of the mean is not the mean of the logs. Indeed, it may even be violated by the most elementary relationships in which $f(x)$ is linear and the relationship under consideration is self-evidently correct. For example, let y be the single share itself and U/N be the ratio of unmarried women (U) to the total population of women (N), so that the “model” under consideration is the widely used and indisputably correct identity $Su = U/N$. Equally “true” are the relationships describing the single shares for black and white women as subpopulations: $Su_1 = U_1/N_1$ and $Su_2 = U_2/N_2$. And yet this relationship fails to satisfy Wu’s requirement; Su , which is equal to U/N , is certainly not also equal to $U_1/N_1 + U_2/N_2$. Thus, even a self-evidently correct model may fail Wu’s condition.

WHAT MIGHT BE LEARNED?

For our own part, we acknowledge that we might have raised fewer hackles by being more explicit and generous with qualifications, and by taking more care with presentation. Our title, for example, implies a more expansive claim than any of the claims we actually made. Nonetheless, we maintain that our analysis of the birth rate paradox from 1974 to 2000 for adult women aged 20–39 strongly suggests that selection effects arising from changing marriage behavior can have powerful compositional effects on the pools of married and unmarried women. Certainly, effects may vary in importance across time and groups, and other factors may be more important over substantial periods of time. Still, it seems obvious by inspection that the rise in unmarried and married birth rates for women aged 20–39 over the final quarter of the past century could not have come from shared or idiosyncratic increases in desired fertility among unmarried and married women, simply because their combined birth rates did not rise.

It would, of course, be presumptuous of us to suggest what EMW might learn from this exchange. EMW reject our approach, but the best way to counter an idea is with a better

idea. If EMW have a better explanation for the central paradox motivating the GSS model, they didn't present it. In conclusion, we appreciate the close attention to our work, and hope readers and EMW find our responses useful.

REFERENCES

- Engle, R.F. and C.W.J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55:251–76.
- Gray, J., J. Stockard, and J. Stone. 2005. "A Tale of Two Shares: The Relationship Between the 'Illegitimacy' Ratio and the Share of Unmarried Women." *Economic Letters* 90:242–48.
- . 2006. "The Rising Share of Nonmarital Births: Fertility Choice or Marriage Behavior?" *Demography* 43:241–53.
- Hamilton, J. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.