AFDC and Births to Unwed Women

Abstract: Most work by economists has been inconclusive when seeking a consistent relationship between income-support programs (like Aid to Families with Dependent Children (AFDC)) and births to unwed women (or, as the literature traditionally terms it, illegitimacy). But a recent study (Kimenyi and Mbaku (1995)) reports a large, positive and statistically-significant relationship when data are weighted to reflect differences in propensities toward illegitimacy. We find that the 1995 study appears to rely upon erroneous data and irregular econometric technique. When these are remedied, the major results are reversed. We then suggest that a switching regimes methodology, with parameters influenced by other variables, is more appropriate to the issue. Our empirical results confirm the literature's consensus that AFDC and illegitimacy do not appear to be strongly related.

JEL classifications J13, I38
Keywords: Welfare programs, AFDC, illegitimacy, income-support programs, feminization of poverty

Kurt C. Schaefer*
Calvin College, 3201 Burton S.E., Grand Rapids, MI, USA 49546
616.957.6298, fax 616.957.8410, e-mail schk@calvin.edu

Sarah E. Hamersma
The University of Wisconsin, Madison, Wisconsin, USA

and

Thomas D.Vander Veen,
United States International Trade Commission, Washington, DC, USA

November 2001

---

* The views expressed here are the views of the authors and in no way represent the views of the United States International Trade Commission or any of its individual commissioners. The authors thank the first author's home institution for support of this project by funding summer research assistance and a brief reduction in teaching load. Correspondence should be directed to the first author.

AFDC and Births to Unwed Women

The relationship between income-support programs (like the Aid to Families with Dependent Children (AFDC) program) and rates of births to unwed women (or, as the literature traditionally terms it, rates of illegitimacy) has been a puzzling one for economists. When a large program offers cash benefits restricted to single parents with children, and simultaneously creates a large effective tax on earned income, one might expect a substantial increase in illegitimate births as a result. Yet the professional literature on the subject has been generally inconclusive. Some find a small positive relationship between benefit levels and illegitimacy rates (e.g., Caudill and Mixon 1993), or report mixed or statistically insignificant positive results (Duncan and Hoffman (1990), Lundberg and Plotnick (1990), Acs (1993)). Several (Ozawa (1989), Plotnick (1990), Clarke and Strauss (1998)) report a positive relationship, but these studies are limited to teenage illegitimacy, and teens constitute only 3-4% of AFDC household heads.[1] While Clarke and Strauss report a stronger relationship than other recent work, they counsel that their results are "unstable" enough to leave them "(less than) fully compelling" for white and black recipients, with no evidence of a relationship for hispanics (Clarke and Strauss, 735). Winegarden's (1988) time-series modeling reports an "AFDC effect" responsible for about half of the increase in non-white illegitimacy over the previous twenty years, but Moore and Caldwell (1977) and Ellwood and Bane (1985) find little or no evidence of a relationship between welfare benefit levels and illegitimacy, and even report some negative correlations (Ellwood and Bane (1985)). The major reviews of the literature (Moffitt (1992), Murray (1993), Acs (1993), Ellwood (1988)) conclude that there is only "mixed evidence of an effect of the welfare system on illegitimacy (Moffitt (1992)). "When the dust settles and researchers do everything they can to adjust for all these differences, they still find little relation between the level of benefits and the number of families headed by women" (Ellwood (1988)).

Kimenyi and Mbaku (1995) present a fresh approach to the issue. They argue that previous studies are fundamentally flawed, in that they treat all potential AFDC recipients as if they have identical propensities toward illegitimacy. Let us suggest an analogy:[2] We believe, for theoretical reasons, that among professional baseball teams better pitching should be associated with better team won-lost records. But will this relationship necessarily be evident in baseball statistics? Consider minor-league baseball, in which there are three divisions (A (the worst), AA, and AAA (the best)). Suppose that pitching and batting are relatively evenly matched in the lowest league, but that pitching improves more rapidly than batting as one moves up through the leagues. By Division AAA one could have such overpowering pitching that batting abilities would be swamped by even the poorest pitchers in the division. In that case, among Division A teams one would find a positive relationship between a team's pitching statistics and its won-lost record, but in other divisions pitching might appear to be uncorrelated to winning. Pitching would thus appear to be statistically unimportant, even though pitching would be the primary driving force in the division.

---

[1] Clarke and Strauss, 828. Rosenzweig (1995) reports a similar relationship for women younger than 23.

[2] Though this analogy is faithful to the logic of the original article, the analogy and labor-leisure diagrams that follow do not appear in the article; they represent our understanding of the original article.

In much the same way, Kimenyi and Mbaku argue that AFDC benefit levels might falsely appear to be uncorrelated with illegitimacy rates. Very low AFDC benefit levels might be enough to change the behavior of all of those who are predisposed toward forming single-parent families. If most states offer benefits that exceed these minimal levels, then regressions of benefit levels against illegitimacy rates might falsely indicate that AFDC benefits are of no consequence in encouraging illegitimacy. The authors propose an alternative methodology, and also rely upon state-level aggregated data to allow for situations in which individual-level data might not be available.

This argument can be presented using a standard labor-leisure diagram (Figure 1). Imagine that everyone in the relevant population has preferences represented by the indifference curves I_1, I_2, and I_3, representing the tradeoff between utility received from income and utility received from leisure. Say that eight women receive wage offer 1, three women receive wage offer 2, and one woman receives offer 3. Without an AFDC program (or other non-labor income) all women work at least some hours (L_1, L_2, L_3).

Now introduce a small AFDC benefit, represented by the height N in Figure 2. If this benefit is only available to single-parent households, all eleven women facing wage offers 1 and 2 would have an incentive to form single-parent households rather than accept their wage offers. Only a significantly larger AFDC benefit would induce women with wage offer 3 to form single-parent households. Such a situation would generate a paradox similar to that of our baseball metaphor: Benefit levels would strongly affect illegitimacy, but after a minimal benefit level is crossed, no correlation between benefit and illegitimacy would be apparent.

Kimenyi and Mbaku suggest that this potential problem with previous studies could be corrected by assigning weights to observed changes in illegitimate birth rates. When regressing changes in illegitimacy rates against changes in AFDC benefit levels in the United States between 1970 and 1980, the authors divide the states into low-, medium- and high-illegitimacy groups, based on 1970 illegitimate-birth rates. The authors place one-third of the states in each group. Changes in illegitimacy rates in the high-illegitimacy states are weighted three times more heavily than changes in low-illegitimacy states, and medium-illegitimacy states are weighted twice as heavily as low-illegitimacy states. The study reports a positive, large, statistically-significant coefficient for AFDC benefit level changes, which is taken to be evidence that generous AFDC programs are a major cause of the feminization of poverty. The study also reports a negative, significant coefficient for the South as a region, which is attributed to traditional, Bible-belt cultural conservatism.

Our paper aims to do two things. First we attempt to replicate the results of the 1995 paper, and extend them beyond the original 1970-80 data to test the dependence of the results on choice of era. Second, we suggest that the Kimenyi/Mbaku methodology may not generally be capable of answering the questions that have been set before it, and we present and estimate an alternative model that treats the problem as a switching regression with sample selection.

Replications of the 1995 Study

Tables 1a and 1b present the Kimenyi and Mbaku results in the four columns labeled "K/M." Column One of Table 1a reports a regression of each state's weighted 1970-80 changes in births to unwed women (their variable is named *WDBUW87*) against the following independent variables:

the state's average monthly family AFDC benefit in 1980 divided by the 1970 level (*AFDC87*);
the percent of the state's population that is African-American (*BLACK*);
the percent of the population that is urban (*URBAN*).

Kimenyi and Mbaku report positive, statistically significant coefficients on all three variables. These results are essentially stable as columns two, three, and four add additional explanatory variables. The addition of the ratio of state median incomes in 1980 and 1970 (*MEDY87*) is initially significant but loses its significance when the dummy variables *SOUTH* (for southern states) and *UTAH* are added. *SOUTH*'s coefficient is negative and significant at the 1% level in all of the K/M regressions. The addition of a male unemployment rate variable (*UNEMP*) in Column 4 has little effect on the regression results. Based on Column 4 (which we will call "Kimenyi and Mbaku's full first model"), Kimenyi and Mbaku conclude that race, geographic/cultural factors, and the change in AFDC payments all significantly affect changes in birth rates among unwed women. They suggest that the regression's high R-squared (.889) tends to confirm the reliability of these conclusions.

Table 1b reports a variation on the regressions of Table 1a, in which the AFDC and median-income variables are changed from 1970/1980 ratios to simple 1980 levels (*AFDC80* and *MEDY80*).[3] In column one, only the coefficient of the variable *BLACK* is significant. However, starting in column two the *AFDC80* coefficient becomes positive and significant, as it was throughout Table 1a. In columns three and four, the coefficient on *MEDY80* is negative and significant, indicating that states with higher incomes tended to experience smaller increases in illegitimacy. The sign on *SOUTH*, though not statistically significant, is positive rather than negative, and the coefficient on *UNEMP* in column four has become significant. Its positive sign indicates that states with higher male unemployment rates tended to experience larger increases in illegitimacy than those with lower unemployment rates.

Kimenyi and Mbaku conclude from these regressions that higher AFDC benefits (and larger increases in benefits) cause higher illegitimacy rates (and greater increases in illegitimacy rates). Since the authors use unwed-mother birth rates as an index of female headship, they reason that AFDC payments induce female headship among women who otherwise have low propensities to be female heads. Thus AFDC plays a significant role in the feminization of poverty.

[3] Since Kimenyi and Mbaku emphasize the results of the first full model, we will focus our attention on that model throughout most of our analysis.

Table 2 contains the variable means and standard deviations reported by Kimenyi and Mbaku for their 1970-1980 data, along with the same descriptive statistics for our data.[4] In replicating their study's results, we encountered serious difficulties in attempting to recreate the means and standard deviations reported by Kimenyi and Mbaku. In part this is because some data are not available in the places cited by their paper.[5] In other cases there are ambiguities regarding the identity of published statistics that they cite.[6] In still other cases different editions of the cited sources disagree about data values, or the citations are ambiguous.[7] Yet we believe that in every case we have selected the best available data to approximate the means and standard deviations reported by Kimenyi and Mbaku.[8] We are unable to locate data that approximate the statistics reported by the authors for the weighted change in births to unwed women from 1970 to 1980,[9] and the data that we cite differ significantly from the authors' data. We believe that their data for this variable are in error.

Even granting that the Kimenyi and Mbaku birth rate data are different from ours, it should still be possible to insert their mean independent variable values (Table 2, column 3) into each regression in Table 1a or 1b, and arrive at the same mean predicted dependent variable value (*WDBUW87*) for all four regressions in each table.[10] But this calculation yields different mean *WDBUW87* values for each regression equation.[11] Both the data and the estimations from those data appear to be in error.

Thus our results might be expected to differ from the Kimenyi and Mbaku study, and they do. Table 1a reports (in the columns headed "New") the regression results for Kimenyi and Mbaku's first weighted-birth-rate model, using our data. The sign on the

---

[4] After repeated phone conversations and e-mail correspondence spread over an 18-month period, the authors were unable to provide us with their data. In the 1995 paper, all of their data were cited from publicly-available, published sources.

[5] The 1970 data necessary to calculate *AFDC87* are not published in the *Social Security Bulletin* as cited. The 1970 and 1980 median income data are not published in *The Statistical Abstract of the U.S.* as cited. (The source contained only 1969 and 1979 data.) Finally, and most significantly, the birth rate data for the dependent variable *WDBUW87* do not appear in *Current Population Reports, Series P-60* or *The County and City Data Book, 1988* as cited.

[6] For example, state median income is not specified as median household, median family, or median family-of-four income, all of which are commonly-published, legitimate measures of state median income. Similarly, their "change in birth rates to unwed women between 1980 and 1970" could be a raw difference or a percentage change.

[7] The 1981 and 1985 editions of *The Statistical Abstract of the U.S.* report significantly different data for 1979 family-of-four median income. Some Kimenyi and Mbaku citations simply refer to "various issues" of the source.

[8] Our replication attempts and sources are discussed in detail in a fourteen-page, single-spaced appendix, available by request from the first author. Our data are also available by request.

[9] Several data sources were examined, several measures of illegitimate birth rate contemplated, and several possible data entry errors explored. None of these alternatives produced results comparable to those reported by the authors. See our Appendix A (available by request from the first author), pp. 6-15, for a detailed discussion of these procedures and their results.

[10] This assumes that we have correctly understood Kimenyi and Mbaku to have weighted only the dependent variable, rather than to have performed a "weighted regression" in which each observation counts for 1, 2, or 3 people.

[11] For example, in the case of Table 1a, the mean predicted *WDBUW87* values for the four regressions are 436.95, 81.02, 49.19 and 105.36.

AFDC benefit-change variable has been reversed, indicating that larger increases in AFDC benefits tend to be associated with *smaller* increases in illegitimacy rates. This AFDC coefficient is now statistically insignificant in every regression. The sign on the dummy variable representing the South as a region is also the opposite of the 1995 study, and is now statistically insignificant. The *INCOME* coefficient is much larger and significant. The *URBAN* coefficient is now insignificant, though unchanged in magnitude. Thus our results contradict all of the results that are salient to the major findings of the Kimenyi and Mbaku paper.

Table 1b reports results from the second model, using our data. The comparison between the two sets of results for this second model is similar to the earlier case: The AFDC benefit coefficient has changed sign and is now insignificant. In addition, the coefficient on *URBAN* is larger and is now significant, and median income has become insignificant.

Some Refinements

Following Kimenyi and Mbaku's practice, all of the regressions thus far have used nominal AFDC benefit levels and nominal median income levels. This overstates the changes in both benefits and income between 1970 and 1980. Average nominal AFDC benefits rose by 54.76% during the period, but inflation-adjusted benefits actually *fell* by 27.13%. Average nominal median incomes rose by 116.93%, but the corresponding real increase was only 9.66%. When one reestimates the full first model using real incomes and benefits, the coefficient on AFDC changes from –17.13 (nominal case) to –36.38 (inflation-adjusted case). Median income's coefficient changes from –152.51 to –301.70.[12] Thus our major results are strengthened by the use of inflation-adjusted data.

One might argue that it makes more sense to weight the changes in birth rates by initial level of *AFDC benefit* rather than by initial illegitimacy rate, since it is the welfare benefit that is presumed to affect the propensity toward illegitimacy. Following this instinct yields a negative, significant coefficient on the *SOUTH* variable, a result similar to Kimenyi and Mbaku's original report. However, the sign on the AFDC coefficient remains negative, and is now significant at the 1% level in three of the four regressions.

We considered the extent to which the K/M results depend on choice of time period by constructing datasets for the decades 1975-85 and 1980-90.[13] Using nominal pooled data from all three datasets, the regression yields the same signs and approximately same significance levels as the 1970-80 regression, with the following relatively minor changes: 1) The sign on the unemployment variable changes (but is insignificant in both cases) 2) The significance of the *URBAN* variable and the *SOUTH* dummy changes somewhat, though both coefficients are approximately the same magnitude in both cases.

[12] All levels of significance remain unchanged. Nominal data were deflated using the annual consumer price index.

[13] References for the sources of these data are included in the relevant sections of the data appendix, available from the first author.

3) The magnitude of the coefficients shrinks in the pooled-data regression for the AFDC, *BLACK*, and *INCOME* variables, and grows for the *SOUTH* and *UTAH* variables.

A Switching Regimes Model

Let us assume for the moment that Tables 1a and 1b report a reasonable set of estimates for the original Kimenyi and Mbaku models. Do we now know something about the relationship between AFDC benefits and illegitimacy? No. The size and signs of the coefficients in these regressions are the result of an arbitrary 1-2-3 weighting scheme; a different scheme would yield different, potentially contradictory results, and be no less justified theoretically. No evidence has been presented to support the idea that residents of high-illegitimacy states have one-third the propensity toward illegitimacy—not, say, one-tenth or one-half the propensity-- of residents in low-illegitimacy states. Nor is there a rationale for grouping states into thirds rather than, say, quintiles or deciles.

We suggest that the situation instead be modeled as a switching regimes problem with parameters influenced by other variables. Figure 3 presents a simple bivariate case in which modest AFDC benefit increases may initially bring forth higher illegitimacy rates, exhausting most of the pool of those with high illegitimacy propensities; thereafter benefit increases would appear to be unrelated to illegitimacy. If most of the data were generated in states to the right of the switch in the relationship, a regression of illegitimacy-rate changes on AFDC-benefit-level changes might yield no evidence of significance.[14] In such a case Kennedy (1998, 100) suggests that one should respecify the regression equation such that the derivative of birth rates with respect to AFDC benefit changes is a function of the AFDC level (larger at low benefits than at high benefits), then use unweighted birth-rate data to directly test whether the function has statistical significance. Let

$$\Delta BUW_i = a + b_1 (\Delta AFDC_i) + (\blacklozenge_i) + \varepsilon_i \ , \qquad (1)$$

where ΔBUW_i is the change in birth rate, $\Delta AFDC_i$ represents the change in benefit levels, ($\blacklozenge_i$) represents other independent variables, and b_1 is a function of initial AFDC levels. Taking the simple linear case, we have

$$\Delta BUW_i = a + [c_1 + c_2 (AFDC)]*(\Delta AFDC_i) + (\blacklozenge_i) + \varepsilon_i \qquad (2)$$

$$= a + c_1 (\Delta AFDC_i) + c_2 (AFDC)*(\Delta AFDC_i) + (\blacklozenge_i) + \varepsilon_i \ . \qquad (3)$$

[14] In effect, the Kimenyi/Mbaku model 1) assumes that the first one-third of the observations fall to the left of the switch-point; 2) assumes that if this relationship were extrapolated rightward, the mean of the next one-third of the observations lies half-way between the x-axis and the extrapolated values; 3) assumes that the mean of the final one-third of the observations lies one-third of the way upward from the x-axis toward extrapolated values above them; and 4) assumes a monotonic relationship between observed illegitimacy rates and observed AFDC-benefit levels, so that birth-rate changes can be weighted by birth-rate levels rather than AFDC benefit levels. Thus if the middle observations were doubled and the rightward observations were tripled, all observations would lie along a regression line extrapolated from the first one-third of the observations.

Following the Kimenyi and Mbaku hypothesis, we would expect $c_1>0$ and $c_2<0$.

Table 3 presents our estimation of this equation. We use nominal 1970-80 data, to allow a direct comparison to the original paper's results. The ΔAFDC coefficient (c_1) is still negative and insignificant. The interaction-term coefficient (c_2) is tiny and insignificant. The only other noteworthy difference between this regression and its earlier siblings is that the coefficients on *URBAN*, *SOUTH* and *UTAH* have all lost statistical significance.

Conclusion

It seems that we have returned to the place where this discussion started. The attempt to find a positive relationship between AFDC benefits and births to unwed women by appealing to variations in propensity toward illegitimacy appears to instead yield iconoclastic results. We are left with coefficients of the "wrong" sign. This might indicate a misspecification of the model, though no one appears to have suggested a more compelling specification that consistently yields contrary results among the general AFDC population. One is otherwise left to conclude that AFDC benefits have not exerted a strong positive influence on births to unwed women.

References

Acs, Gregory, 1993, The impact of AFDC on young women's childbearing decisions, Unpublished paper, The Urban Institute.

Clarke, George R. G. and Robert P. Strauss, 1998, Children as Income-Producing Assets: The Case of Teen Illegitimacy and Government Transfers, *Southern Economic Journal* 64, 827-856.

Caudill, Steven B., and Franklin G. Mixon, 1993, A Note on the effects of AFDC payments on birthrates, *Rivista Internazionale di Scienze Economiche e Commercali* 40, 379-84.

Duncan, Greg J., and Saul D. Hoffman, 1990, Welfare benefits, economic opportunities, and out-of-wedlock births among black teenage girls, *Demography* 27, 519-35.

Ellwood, David T., 1988, *Poor Support: Poverty in the American Family* (Basic Books, Inc., New York).

Ellwood, David T. and Mary Jo Bane, 1985, The Impact of AFDC on Family Structure and Living Arrangements, *Research in Labor Economics* 7, 137-207.

Kennedy, Peter, 1998, *A Guide to Econometrics*, fourth edition (The MIT Press, Cambridge).

Kimenyi, Mwangi S. and John Mukum Mbaku, 1995, Female Headship, Feminization of Poverty and Welfare, *Southern Economic Journal* 62, 44-52.

Lundberg, Shelly, and Robert D. Plotnick, 1990, Effects of state welfare, abortion and family planning policies on premarital childbearing among white adolescents, *Family Planning Perspectives* 22, 246-51.

Moffitt, Robert, 1992, Incentive effects of the U.S. welfare system: A review, *Journal of Economic Literature* 30, 1-61.

Moore, Kristen A and Steven B. Caldwell, 1977, The Effect of Government Policies on Out-of-Wedlock Sex and Pregnancy, *Family Planning Perspectives* 9, 164-69.

Murray, Charles, 1993, Welfare and the Family: The U.S. experience, *Journal of Labor Economics* 11, 224-62.

Ozawa, N. Martha, 1989, Welfare policies and Illegitimate Birth Rates Among Adolescents: Analysis of State-by-State Data, *Social Work Research and Abstracts* 14, 5-11.

Plotnick, Robert D., 1990, Welfare and Out-of-Wedlock Childbearing: Evidence from the 1980s, *Journal of Marriage and the Family* 52, 735-46.

Rosenzweig, Mark, 1995, Welfare, marital prospects and nonmarital childbearing, Unpublished paper, University of Pennsylvania.

United States Bureau of the Census, 1970, 1972, 1983, *County and City Data Book* (Government Printing Office, Washington).

United States Bureau of the Census, 1989, *Current Population Reports, Series P-25, #1040-RD-1, "Population Estimates by Race and Hispanic Origin for States, Metropolitan Areas, and Selected Counties: 1980 to 1985,"* by David L. Word (Government Printing Office, Washington).

United States Bureau of the Census, 1991, *Current Population Reports, Series P-60* (Government Printing Office, Washington).

United States Bureau of the Census, 1979, 1982, 1986, *State and Metropolitan Data Book* (Government Printing Office, Washington).

United States Bureau of the Census, 1979, 1980, 1981, 1984, 1985, 1987, 1991, 1993, 1995, *Statistical Abstract of the United States* (Government Printing Office, Washington).

United States Bureau of the Census, World Wide Web Page, www.census.gov/population/www/socdemo/birth.html

United States Department of Health and Human Services (Department of Health, Education, and Welfare), 1970, 1975, 1980, 1985, 1990, *Vital Statistics of the United States* (Government Printing Office, Washington).

United States Department of Labor, June 1991, *Bureau of Labor Statistics Bulletin #2381, "Geographic Profile of Employment and Unemployment, 1990"* (Government Printing Office, Washington).

United States Social Security Administration, 1970, 1971, 1981, 1987, 1992, *Social Security Bulletin, Annual Statistical Supplement* (Government Printing Office, Washington).

United States Social Security Administration, December 1970, January 1976, *Social Security Bulletin* (Government Printing Office, Washington).

Winegarden, Calmon R., 1988, AFDC and the Illegitimacy Ratios: A Vector-Autoregressive Model, *Applied Economics* 20, 1589-1601.

Table 1a: Comparison of Regression Results
(Based on regression reported in Table II in K/M paper)
(Dependent variable: *WDBUW*)

Variable	1		2		3	
	K/M	New	K/M	New	K/M	New
Intercept	179.01 (3.48)	38.57 (0.70)	35.19 (0.27)	368.67** (2.24)	-122.12 (-1.08)	399.27** (2.40)
AFDC87	62.34** (2.20)	-13.49 (-0.49)	58.18** (2.11)	-15.97 (-0.61)	41.50* (1.76)	-11.98 (-0.46)
BLACK	6.91*** (12.50)	6.88*** (11.89)	6.93*** (12.89)	6.84*** (12.39)	8.09*** (15.20)	6.27*** (9.91)
URBAN	1.34** (2.58)	0.80 (1.52)	1.33** (2.65)	0.45 (0.86)	0.85* (1.95)	0.79 (1.45)
MEDY87			-95.99* (-1.75)	-139.58** (-2.12)	-8.04 (-0.16)	-166.49** (-2.47)
SOUTH					-74.73*** (-4.25)	28.63 (1.46)
UTAH					-44.83 (-1.17)	-62.84 (-1.34)
UNEMP						
ADJ. R2	.825	.8023	.832	.820	.890	.829
F	62.621***	53.75***	50.41***	45.35***	54.71***	32.57***

Note: *t*-statistics are given in parentheses below the coefficients. *F* is the *F*-ratio. Asterisks denote levels of significance at the (*) 10%, (**) 5

Table 1b: Comparison of Regression Results
(Based on regression reported in Table III in K/M paper)
(Dependent variable: *WDBUW*)

Variable	1		2		3	
	K/M	New	K/M	New	K/M	New
Intercept	-1.95 (-0.08)	33.61 (0.91)	67.82 (1.33)	36.28 (0.52)	58.65 (1.12)	58.02 (0.77)
AFDC80	0.12 (1.62)	-0.14 (-1.26)	0.20** (2.26)	-0.14 (-1.10)	0.25** (2.49)	-0.12 (-0.80)
BLACK	5.13*** (10.59)	6.47*** (9.75)	5.18*** (10.89)	6.47*** (9.39)	5.01*** (10.39)	6.25*** (8.76)
URBAN	0.29 (0.73)	1.13* (1.94)	0.27 (0.69)	1.15* (1.71)	0.42 (1.06)	1.49** (2.11)
MEDY80			-0.004 (-1.58)	-0.0002 (-0.05)	-0.005* (-1.81)	-0.002 (-0.51)
SOUTH					12.60 (0.83)	-1.99 (-0.08)
UTAH					-49.31 (-1.52)	-77.20 (-1.51)
UNEMP						
ADJ. R2	.803	.809	.810	.804	.816	.806
F	54.12***	56.18***	42.92***	40.97***	29.90***	27.92***

Note: *t*-statistics are given in parentheses below the coefficients. *F* is the *F*-ratio. Asterisks denote levels of significance at the (*) 10%, (**) 5
-- The authors reported this coefficient with just one asterisk in the original table, but its *t*-statistic indicates that it should have two asterisks,

Table 2. Descriptive Statistics: Means and Standard Deviations

Variable	Our Data	Kimenyi/Mbaku Data
WDBUW87 *(weighted difference in births to unwed women, 1980 to 1970)*	145.73 (105.40)	68.51 (30.71)
DBUW87 *(difference in births to unwed women, 1980 to 1970)*	66.83 (28.82)	
BLACK *(% of population that is black, 1980)*	10.34 (12.52)	10.33 (12.50)
URBAN *(% of population that is urban, 1980)*	67.58 (14.99)	67.58 (14.98)
MEDY80 *(median family income, 1980)*	22015.27 (2489.08)	22013.73 (2484.43)
MEDY87 *(median family in income in 1980 / median family income in 1970)*	2.15 (0.13)	2.14 (0.12)
AFDC80 *(average AFDC payment per family per month, 1980)*	249.71 (84.26)	249.72 (84.32)
AFDC87 *(average AFDC payment per family per month in 1980 / average AFDC payment per family per month in 1970)*	1.54 (0.27)	1.54 (0.26)
UNEMP *(male unemployment rate, 1980)*	6.50 (1.85)	6.50 (1.85)

Note: Standard deviations are given in parentheses below means.

Table 3. Effect of ΔAFDC allowed to depend on initial AFDC level
First Full Model, 1970-80 data
(Dependent variable: *unweighted DBUW*)

Variable	Coefficient and t-statistic
Intercept	156.04** (2.39)
ΔAFDC	-0.21 (-1.28)
*ΔAFDC*AFDC70*	0.0004 (0.50)
BLACK	1.33*** (4.95)
URBAN	0.27 (1.23)
MEDY87	-57.21** (-2.16)
SOUTH	-3.25 (-0.40)
UTAH	-21.80 (-1.18)
UNEMP	2.31 (1.55)
ADJ. R-SQUARED	0.659
F	10.41***

Note: *t*-statistics are given in parentheses below the coefficients. *F* is the *F*-ratio. Asterisks denote significance at the (*) 10%, (**) 5%, and (***) 1% levels.

Figure 1

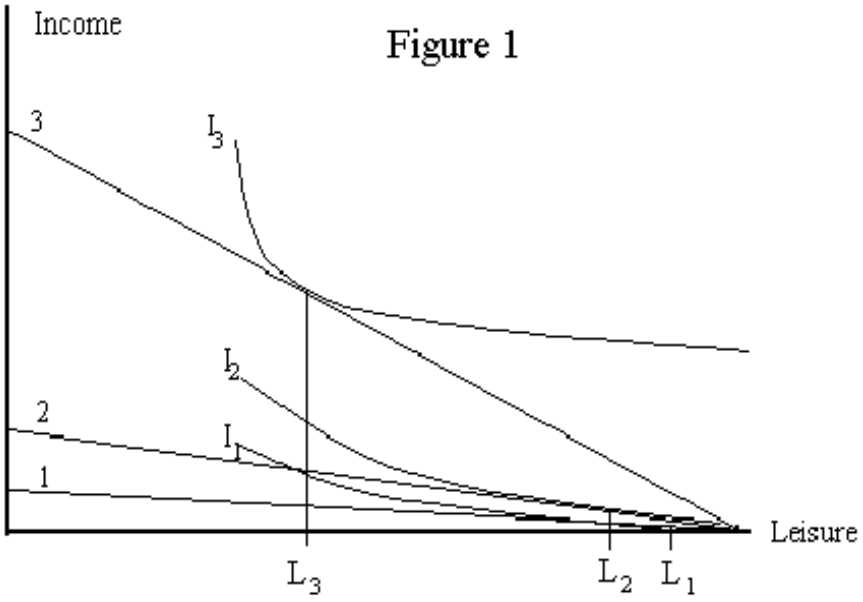
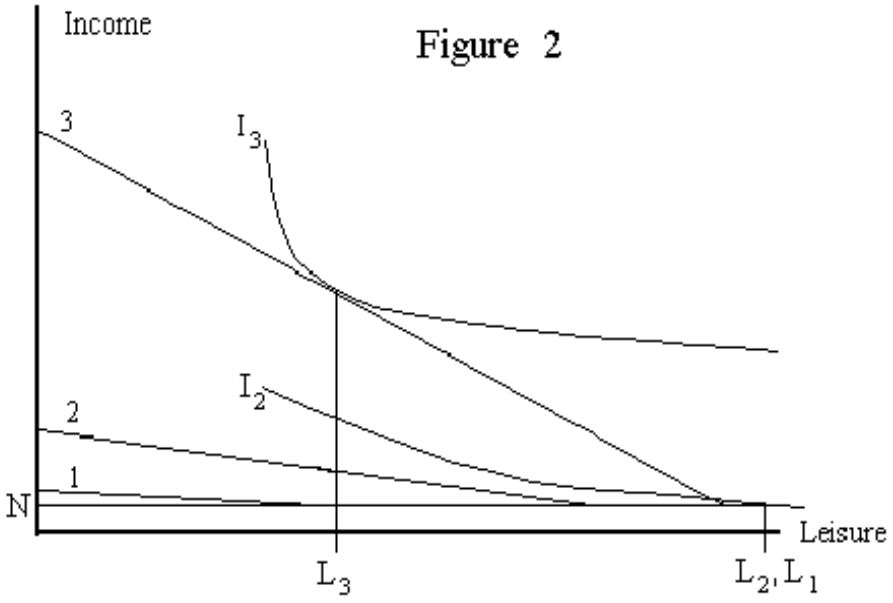


Figure 2



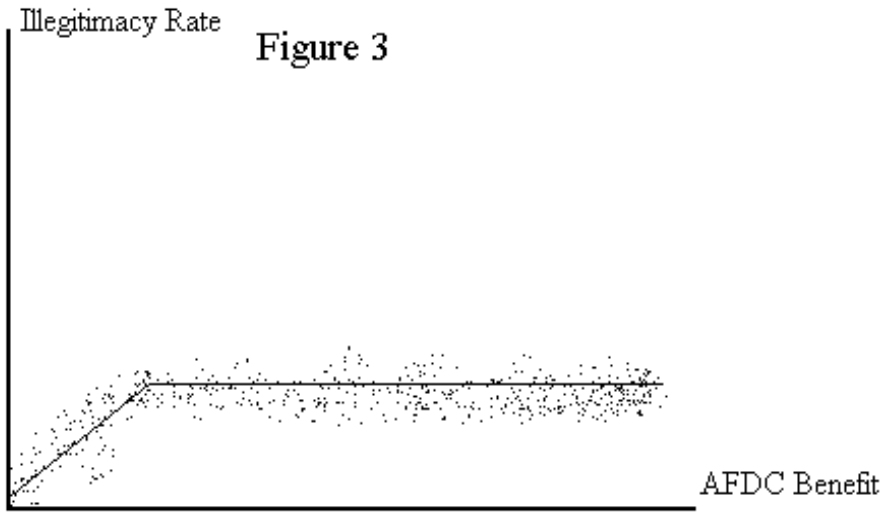


Figure 3

Data Collection Appendix

This appendix documents in detail the steps taken in gathering the 1970-1980 data necessary to replicate the work done by Kimenyi and Mbaku (abbreviated hereafter K/M (1995)). A total of 10 variables are used in their regressions. For each variable, this appendix gives the definition, data source(s) cited by Kimenyi and Mbaku, a description of our attempts to replicate the data used, and our final data source.

At the end of each section we also give our source for the 1975-1985 and 1980-1990 data with which we test the original estimations' stability over different periods of time.

1. AFDC80

Definition: AFDC80 = average monthly AFDC benefits per family in 1980 by state (K/M pp. 48-49)

Citation: *Social Security Bulletin* (various issues)

Replication Attempt: The *Social Security Bulletin Annual Statistical Supplement, 1981* contains a table with monthly average AFDC payment per family by state. Use of these data gives means and standard deviations almost identical to those reported by Kimenyi and Mbaku (see our Table 2). All of our data entry was double-checked in detail for possible typographical errors. Thus we concluded that these were the appropriate data for the replication.

Final Data Source: U.S. Social Security Administration. *Social Security Bulletin, Annual Statistical Supplement, 1981*. Table 191, p. 250.

Source for the 1975 data: U.S. Social Security Administration. *Social Security Bulletin,* January 1976, p. 72, Table M-33, col. 5 ("Per Family" data).

Source for the 1985 data: U.S. Social Security Administration. *Social Security Bulletin, Annual Statistical Supplement, 1987*, p. 295, Table 218, column 6 ("Monthly Average per Family" column).

Source for the 1990 data: U.S. Social Security Administration. *Social Security Bulletin, Annual Statistical Supplement, 1992*, p. 320, Table 9.G2, column 6 ("Monthly Average per Family" column).

2. AFDC87

Definition: AFDC87 = ratio of average monthly AFDC benefits per family in 1980 to average monthly AFDC benefits per family in 1970 (K/M pp. 48-49)

Citation: *Social Security Bulletin* (various issues)

Replication Attempt: Replication requires data for average monthly benefits per family in 1970, divided into AFDC80 (above). The *Social Security Bulletin Annual Statistical*

Supplement, 1970 (Table 140, p.134) contains state level data for AFDC benefits, but gives only the "per recipient" rather than the "per family" benefit amounts. It also gives only the average benefit in the month of December rather than over the course of the entire year. The December 1970 issue of the *Social Security Bulletin* (Table M-26, p.41) gives similar data for the month of July, along with the number of families receiving AFDC and number of recipients of AFDC for July. While per-family benefits are still not given in this table, an average per-family benefit for the month of July 1970 can be calculated from the table ([# total recipients/ # families] * average benefit per recipient].

The same formula would yield average per-family benefit in December using the December benefit data, but the data on the number of families and recipients are available only for July, so using the December data creates a less accurate picture of family benefits. In hopes of finding published average per-month per-family data, the *Annual Statistical Supplement* for 1971 and all of the 1971 issues of the *Bulletin* were thoroughly checked. Unfortunately, the 1971 *Supplement* contains only December 1971 data, and the only 1971 *Bulletin* that contains social welfare expenditure information does not contain state-level data. Thus we were left with two less-than-ideal measures of average per-family AFDC benefits in 1970. Since Kimenyi and Mbaku do not report any descriptive statistics for the 1970 data, each of these two potential measures were used to calculate an AFDC87 ratio. The means and standard deviations were as follows:

K/M AFDC87	AFDC87 using July data	AFDC87 using December data
1.54	1.51	1.49
(0.26)	(0.29)	(0.30)

Neither of our measures appeared close enough to indicate we had replicated the data used by Kimenyi and Mbaku. Furthermore, Kimenyi and Mbaku did not report doing any of these calculations to measure monthly average family benefits for 1970. However, we had exhausted their reported source. We consulted the *State and Metropolitan Data Book* (SMDB) as a likely potential source. The 1979 SMDB contains a table of average monthly AFDC payments per family for 1970. These data yield an AFDC87 variable whose means and standard deviations are nearly identical to those reported by Kimenyi and Mbaku:

K/M AFDC87	AFDC87 using SMDB data
1.54	1.54
(0.26)	(0.27)

Because these new values were by far the closest to replicating the descriptive statistics given by Kimenyi and Mbaku, and because the data source the authors cite did not contain data for average benefits per month per family for 1970, we rely upon the data from SMDB 1979 in our calculation of AFDC87.

Final Data Source: For 1980 data: same as for AFDC80
For 1970 data: U.S. Bureau of the Census. *State and Metropolitan Data Book, 1979*. Table A, Column 952, p. 59.

3. BLACK

Definition: BLACK = percentage of state population that is black, 1980 (Kimenyi and Mbaku p. 49)

Citation: *Statistical Abstract of the U.S.* (no specific year cited)

Replication Attempt: The 1984 *Statistical Abstract of the U.S.* contains a table listing the total population by state (in thousands) and the total black population by state (in thousands). Dividing the black population by the total population for each state and multiplying by 100 gives the percentage of each state that is black. The means and standard deviations of these data match those reported by Kimenyi and Mbaku almost exactly (see Table I).

Final Source: U.S. Bureau of the Census. *Statistical Abstract of the U. S., 1984*. Table 37, p. 36.

Source for the 1985 data: U.S. Bureau of the Census, *Current Population Reports, Series P-25, #1040-RD-1*, 1989, Table 1B, column 9 ("Percent Black 1985" column).
Source for the 1990 data: U.S. Bureau of the Census. *Statistical Abstract of the United States, 1993*, p. 30-31, Table 32, column 4.

4. URBAN

Definition: URBAN = percentage of state population that is urban, 1980

Citation: *Statistical Abstract of the U.S.* (no specific year cited)

Replication Attempt: The 1981 *Statistical Abstract of the U.S.* contains tables giving the percentage of each state that is urban as well as the percentage of each state that is metropolitan (Table 10, p. 12 and Table 22, p. 19, respectively). The descriptive statistics for "metropolitan" were not very close to the Kimenyi and Mbaku values, but the mean for "urban" matched up exactly and the standard deviation almost exactly (see Table 2). Thus we concluded that these values for "percent urban" were most likely the same as those used by Kimenyi and Mbaku.

Final Source: *Statistical Abstract of the U.S., 1981*. Table 10, p. 12.

Source for 1985 data: The percent urban for each state in 1985 was not available, as 1985 is not a census year. Data for the percentage in metropolitan areas are available, but were clearly incompatible with the 1980 and 1990 percent urban data, indicating a dramatic rise and then fall in urbanization at the 1980-1985-1990 checkpoints for some states, a fall and then rise for others. In the end, we have simply averaged the 1980 and 1990 percent urban figures to arrive at the 1985 figures. It seems justifiable that changes in

urbanization during the 1980s should have been (approximately) a linear function of time.
Source for the 1990 data: U.S. Bureau of the Census. *Statistical Abstract of the United States, 1993*, page 34, Table 37, column 4.

5. MEDY80

Definition: MEDY80 = median income of state, 1980

Citation: *Statistical Abstract of the U.S.* (no specific year cited)

Replication Attempt: This variable was very difficult to replicate. The *Statistical Abstract of the U.S.* contains only 1979 median income data, citing the 1980 U.S. Census is one of its sources. Every *Statistical Abstract* from 1980 to the present was searched for 1980 median income data, but 1980 data were simply never reported. We presume Kimenyi and Mbaku used 1979 median income data.

Kimenyi and Mbaku do not report which unit of income was used: median income per capita, median income per family, or median income per family of four. As their mean for 1980 ($22013.73) was clearly too high to be a per-capita measure, we considered both the per-family and per-family-of-four data published in The 1985 *Statistical Abstract of the U.S.* (Tables 750 and 751, p. 450). The descriptive statistics follow:

K/M MEDY80	MEDY80 per family	MEDY80 per family of 4
22013.73	19492.06	22229.33
(2484.43)	(2591.52)	(2501.84)

Neither option seemed to replicate those the authors report. However, we found that the 1981 *Statistical Abstract* reports data for 1979 median income for families of four that differ from those in the 1985 *Statistical Abstract*. The descriptive statistics for the 1979 family-of-four data reported in the 1981 *Statistical Abstract* are very close to those reported by Kimenyi and Mbaku:

K/M MEDY80	MEDY80 using ’79 data from 1981 SA
22013.73	22015.27
(2484.43)	(2489.08)

Since these data were clearly the best match to those reported by the authors, and since they were found in a source Kimenyi and Mbaku cite, we chose this source as the best replication of the Kimenyi and Mbaku data. Thus we suggest that that Kimenyi and Mbaku’s “median income in 1980” actually translates to “median family-of-four income in 1979,” as reported in the 1981 *Statistical Abstract.*

Final Source: U.S. Bureau of the Census. *Statistical Abstract of the U.S., 1981*. Table 735, p. 441.

Source for the 1975 data: U.S. Bureau of the Census. *Statistical Abstract of the United States, 1980,* page 455, Table 754, column 3.
Source for the 1985 and 1990 data: U.S. Bureau of the Census. *Statistical Abstract of the United States, 1995,* page 473, Table 730, columns 2 and 7.

6. MEDY87

Definition: MEDY87 = ratio of median income in 1980 to median income in 1970

Citation: *Statistical Abstract of the U.S.* (no specific year cited)

Replication Attempt: Clearly, this variable depends on both MEDY80 and median income data for 1970. As was the case with 1980 data, true 1970 median income data were simply not available. All published median income data were given for 1969 rather than 1970, again because of the timing of the decennial U.S. Census. Thus we again assume that Kimenyi and Mbaku report 1969 data as 1970 data.

We again tried both per-family and per-family-of-four data to replicate the Kimenyi and Mbaku MEDY70. In comparing the 1985 *Statistical Abstract* (Tables 750 and 751, p. 450) with the 1981 edition, we detected a presumed typographical error in the 1985 *Statistical Abstract*: it lists median family-of-four income for Alabama as $3,825. This is less than half that of any other state. The 1981 *Statistical Abstract* reports $8,825, which is much more consistent with the other states. The 1980 *Statistical Abstract* (Table 754, p. 455) reports an Alabama value of $8,825 as well, which we presume to be correct.

The only other difference between the 1985 and 1981 editions was in the median income of Vermont: the 1985 edition reports $9,746, while the 1981 edition reports $9,742. The 1980 edition reports $9,746, which suggests an error in the 1981 edition. Thus we presume a $9,746 value for Vermont, though this does not materially affect the means, standard deviations, or regression results.

The remains of all three data sets for 1969 median income for a family of four were identical: only typographical errors had made the lists distinct. Thus we could double-check our choice of MEDY80 data by comparing AFDC87 ratios calculated from each AFDC80 option to those cited by Kimenyi and Mbaku. We double-checked our decision to use family-of-four data in the same way:

K/M MEDY87	MEDY87(fam)	MEDY87 (fam-4), SA 85	MEDY87 (fam-4), SA 81
2.14	2.14	2.17	2.15
(0.12)	(0.14)	(0.11)	(0.13)

While all of our descriptive statistics are very close to those reported by Kimenyi and Mbaku, it seems most appropriate to continue to use the data for family-of-four presented in the 1981 *Statistical Abstract* as the best replication of MEDY80.

Final Source: *Statistical Abstract of the U.S., 1981*. Table 735*, p. 441.
*with corrected Vermont statistic of "$9,746"

7. SOUTH

Definition: SOUTH = 1 if a southern state, and zero (0) otherwise (Kimenyi and Mbaku p. 49)

Final Source: List of states as given in Kimenyi and Mbaku footnote #7, p.49

8. UTAH

Definition: UTAH: 1 if state is Utah, and zero (0) otherwise (Kimenyi and Mbaku p. 49)

9. UNEMP

Definition: UNEMP = male unemployment rate, 1980 (Kimenyi and Mbaku p. 49)

Citation: *Statistical Abstract of the U.S.* (no specific year cited)

Replication Attempt: The 1981 *Statistical Abstract* of the U.S. gives a by-state listing of unemployment rates by gender for 1980. The descriptive statistics of this data set are identical to those reported by Kimenyi and Mbaku.

Final Source: U.S. Bureau of the Census. *Statistical Abstract of the U.S., 1981*, Table 638, p. 382.

Source for the 1975 data: U.S. Bureau of the Census. *Statistical Abstract of the U.S., 1979*, Table 672, p.405.
Source for the 1985 data: U.S. Bureau of the Census. *Statistical Abstract of the U.S., 1987*, Table 641, p.377.
Source for the 1990 data: U.S. Department of Labor, Bureau of Labor Statistics Bulletin #2381, June 1991, *Geographic Profile of Employment and Unemployment, 1990*, Table 12, pp. 35-48.

10. WDBUW

Definition: WDBUW = weighted change in birth rate to unmarried women between 1970 and 1980 (Kimenyi and Mbaku p. 48)

Citations: *Current Population Reports, Series P-60*, 1991
County and City Data Book, 1988

Replication Attempts: This variable involves both 1970 and 1980 rates of birth to unmarried women per 1000 live births, by state. (Kimenyi and Mbaku p. 49). We took the following steps to locate these data:

A. Looking in the Cited Sources

1. *Current Population Reports, Series P-60* are generally published in short booklets. Though the *P-60* series as a whole is entitled "Consumer Income," which seems an unlikely source for birthrate data, we consulted each report available from 1991. While some P-60 Series reports contain data tangentially related to birth rates, we are unable to locate the appropriate data anywhere in the series. We consulted CPR publications in series other than the *P-60* series, to explore the possibility of slight errors in the original citation. This search was also unproductive in locating data on illegitimate births. Interestingly, the Census Bureau (which publishes the *Current Population Reports*) directs researchers seeking birth rate information (including information on out-of-wedlock births) to *Vital Statistics of the U.*S., not to the *Current Population Reports*. (For example, consult the Census Bureau web site: www.census.gov/population/www/socdemo/birth.html.

2. We considered the other source cited, the *City and County Data Book, 1988*. While this source has only limited state-level data, one table (Table A, p. 4) reports birth rate data. However, the table reports only the general birth rate (regardless of marital status), the total number of births, and the percentage of births that were to mothers under 20 years old (Table A, Columns 32-34, p. 4). Furthermore, these data are all for the year 1984, not 1980 or 1970. As with the *Current Population Reports* citation, we explored the possibility of a typographical error or mistake in the year of the citation. We consulted several editions of the *County and City Data Book* (CCDB), along with the nearly identical publication *The State and Metropolitan Data Book* (SMDB). The 1986 SMDB lists general birth rates for 1983 and the percentage of births to mothers under 20 years old for 1982, and ranks each of them by state (Table 1, Columns 19-22, p. XXVII). The 1983 CCDB reports 1980 data (one of the years being sought), but again lists only total number of births, percentage of births that are to mothers under 20 years old, and general birth rate (Table A, Columns 30-32, p. 4). The 1977 CCDB contains birth rate and total number of births for 1975 and general birth rate for 1970 (Table 1, Columns 16-18, p. 3). No illegitimate or teenage birth data are reported. The same is true for the 1972 CCDB. We could only conclude at this point that the authors either did not use the cited sources or used some proxy for illegitimate birth rate.

B. Using Other Possible Sources

1. Before examining possible proxies (such as teen birth rate or a female-headed-households rate), we considered illegitimate birth rate data from other sources that appear in K/M citations. The 1991 *Statistical Abstract of the U.S.* contains state-level data for the 1980 percentage of births to unmarried women, in this case, per 100 live births rather than per 1000 (Table 93, p. 67). However, these statistics are not given by state for 1970

in any edition of the *Statistical Abstract of the U.S.* In reporting national-level data (for example, 1984, Table 97, p. 70) and state-level general birth rate data (for example, 1985, Table 84, p. 59), the *Statistical Abstract* generally cites *Vital Statistics of the U.S.*, an annual publication of the U.S. National Center for Health Statistics. (Note the Census Bureau reference to the same publication in 10.A.1 above.) The *1970-Natality* volume of *Vital Statistics* contains data for the ratio of illegitimate live births per 1000 total live births by state (Vol.1, Table 1-34, p. 1-31). Further, this table contains data from only 39 states and the District of Columbia, which (allowing some poetic license) compares favorably with Kimenyi and Mbaku's comment that 1970 data was available only for 40 states and the District of Columbia (p. 50). The *1980-Natality* volume reports the same information for 1980 (Vol. 1, Table 1-36, p. 1-60).

2. We wished to compare the means and standard deviations of these data with those reported by Kimenyi and Mbaku. Unfortunately, the authors report descriptive statistics only for the weighted changes in birth rates, not for the birth rates themselves or their unweighted differences. Thus we calculated the difference in birth rates for each state (illegitimate birth rate per 1000 births in 1980 – illegitimate birth rate per 1000 births in 1970), then (as described by Kimenyi and Mbaku) divided the states into three approximately equal groups according to 1970 birth rate to unmarried women. We then weighted each group's differences in birth rates according to the model prescribed by the authors (p. 48): States with the lowest 1970 illegitimate birth rates receive a weight of 1 (13 states), those with highest 1970 illegitimate birth rates a weight of 3 (13 states), and remaining states a weight of 2 (14 states). The results surprised us:

K/M WDBUW87	WDBUW87 (using Vital Statistics Data)
68.51	145.73
(30.71)	(105.40)

The descriptive statistics for our unweighted data follow:

DBUW87 (unweighted)
66.83
(28.82)

Our unweighted mean and standard deviation were actually quite close to the authors' *weighted* mean and standard deviation. We double-checked our data entry and consulted the *Statistical Abstract* tables to consider potential errors on our part. The 1991 *Statistical Abstract* reports national illegitimate birth rates in 1970 and 1980 as 10.7% (107 per thousand) and 18.4% (184 per thousand) respectively. These indicate that our calculated mean difference in birth rates of about 67 is not unreasonable. In fact, to generate the Kimenyi and Mbaku *weighted* mean difference of 68.51, the unweighted mean difference would have to be somewhere near 20-40, which is not supported by the national level data. We suspected Kimenyi and Mbaku might have inadvertently reported unweighted mean differences rather than weighted mean difference. However, this is unlikely because, if it were the case, our regression results should be comparable to those in the Kimenyi and Mbaku paper.

3. This left us wondering how else the authors may have calculated the change in birth rate. We had assumed that the change in birth rates would be measured as a difference (1980 rate – 1970 rate). However, since the authors refer once in the paper to the "rate of change of births to unmarried women" (p. 48), we considered measuring the percentage change rather than the simple difference, using 1970 as the base:

K/M WDBUW87	DBUW, %, base '70	WDBUW, weighted, %, base '70
68.51	66.84	127.56
(30.71)	(22.87)	(51.92)

The results are similar to those involving simple raw differences, because many of the initial birth rates (1970) are close to 100, which makes a percentage change calculation very similar to a raw difference.

We also tried calculated percentage changes using a 1980 base:

K/M WDBUW87	DBUW87, %, base '80	WDBUW87, weighted, %, base '80
68.51	38.97	76.23
(30.71)	(8.32)	(30.50)

While this weighted mean and standard deviation are the closest so far to the K/M report, regressions using this measure of WDBUW87 yield results significantly different from theirs, indicating a poor replication of their data.

4. *Vital Statistics* contains illegitimate birth rate data in two forms: total number of births to unwed mothers (1980, Vol. 1, Table 1-72, pp. 1-228 – 1-243 and 1970, Vol. 1, Table 1-69, pp. 1-177 – 1-191) and illegitimate births "per 1000 live births" (the form we had been using). Since this source also contains information on total number of births per state (1980, Vol. 1, Table 1-49, p. 1-75 and 1970, Vol. 1, Table 1-47, p. 1-44) it seemed wise to try using these two raw data sets to see if some combination of them would yield appropriate statistics. We calculated:

births to unwed F per 1000 live births = (# births to unwed F/ # total live births) * 1000

These data yielded the following descriptive statistics:

K/M WDBUW87	DBUW87 w/ *V.S.* raw data	WDBUW87, wtd, w/ *V.S.* raw data
68.51	64.85	138.96
(30.71)	(23.14)	(86.84)

These results are close those obtained using the earlier table, and are again not at all similar to the Kimenyi-Mbaku statistics.

5. At this point, it appeared that Kimenyi and Mbaku could not have used the source we were using for birth rate statistics. However, before moving on to other sources, it should

be mentioned that, aside from the mean and standard deviation reported for birth rates by Kimenyi and Mbaku, the statistics cited occasionally in the text of the Kimenyi-Mbaku paper are actually consistent with the *Vital Statistics* data that we were using. In fact, though they cite the *Statistical Abstract of the U.S.* (*SA*), their data differ in three ways from those presented in *SA*:

a. The *SA* does not have the precision of the numbers given in the paper. (Those in the paper have one more decimal place.) (See 1991 *SA*, Table 93, p. 67)

b. The number given for illegitimate births to whites is less precise in the *SA* than in the Kimenyi-Mbaku paper, and the number given for nonwhites in 1950 is simply different. Specifically, the *SA* lists its white and nonwhite percentages as 1.7% and 16.8% respectively, while Kimenyi and Mbaku report 17.5 per 1000 births (1.75%) for whites and 179.6 per 1000 births (17.96%) for nonwhites.

c. The *SA* does not give a statistic for the 1988 nonwhite illegitimate birth rate, which Kimenyi and Mbaku report as 539 per thousand. The *SA* does give a statistics for blacks, but this is significantly different (63.5%) (see *SA* 91, Table 92, p. 67).

Since the data reported in the paper could not have come from the *Statistical Abstract* as cited, it seemed appropriate to compare the reported statistics to those we found in *Vital Statistics of the U.S.*. The 1980 statistics given by Kimenyi and Mbaku are identical to those reported in *Vital Statistics,* except that they report the average for nonwhites as 485.5, while *Vital Statistics* lists 484.5. The numbers for 1988 are also the same, except that Kimenyi and Mbaku have one less decimal place than the *Vital Statistics* reports. This finding reinforces our decision to use the statistics that we found in *Vital Statistics*, since it appears that these data are consistent with those used early in the paper by Kimenyi and Mbaku. However, the data source for their regressions still appears to not be *Vital Statistics*.

C. Possible Proxy Use or Alternative Measurement Methods

1. It is possible that, instead of directly using "birth rate to unwed mothers per 1000 total births," the authors used some other measure as a proxy for this statistic. This possibility was especially interesting to us because the direct measures were not present in the cited sources. We first tried using the teenage birth rates given in CCDB, since they may in some loose sense reflect the general birth rate to unmarried women. However, this statistic was only available for 1980 (1983 CCDB, Table A, Column 31, p. 4). So this could not be the authors' proxy.

The CCDB and SMDB contain data related to female-headed households. Since the Kimenyi-Mbaku paper consistently relates increases in illegitimate birth rates to increases in female household headedness (p. 48), we checked to see if female-headed households' births were used as a proxy for illegitimate births. The number of female-headed households per 1000 households was computed for 1970 and 1980,

FHH per 1000 HH = (total # FHH / total # HH) * 1000 ,

using data from the 1982 SMDB (Table C, Columns 104, 105, 116, and 122; pp. 454-455). We then assigned weights based on the number of FHH per 1000 HH in 1970 (the proxy for births to unmarried women in 1970). This yielded the following descriptive statistics:

K/M WDBUW87	DFHH87	WDFHH87
68.51	16.01	34.03
(30.71)	(7.15)	(23.68)

These statistics were clearly not those used by Kimenyi and Mbaku. These results make the possibility of proxy use seem quite unlikely. While this is not surprising, it leaves us at a loss as to the source of the data used by Kimenyi and Mbaku.

2. Since we had exhausted the cited sources, and since the data we did find were not consistent with Kimenyi and Mbaku's calculations, we considered other ways they could have measured births to unwed mothers (without using a proxy) that would give different results than ours. While Kimenyi and Mbaku report that they used births to unwed mothers "per 1000 live births," some sources document illegitimate birth rates "per 1000 women 15-44." This is actually the more traditional measure when describing the overall birth rate. If Kimenyi and Mbaku used this measure of illegitimate birth rate, then it would make sense that our data from *Vital Statistics* would not compare to theirs. Both the *Statistical Abstract* (1995, Table 94, p. 77) and the *Vital Statistics* (1980, Table 1-32, p. 1-56) report national (but not state-level) illegitimacy statistics in this traditional way. Furthermore, the two sources agree exactly in their reported statistics: 26.4 illegitimate births per 1000 women ages 15-44 in 1970, and 29.4 in 1980. While the national average is not as useful to us as state-level data, this national average change of only 3 (11%) is clearly not large enough to produce a weighted mean difference of 68.51 at the state level. It follows that this method of measurement could not have been the one used by Kimenyi and Mbaku.

D. Possible Data Entry Errors

1. Having concluded that differences in measurement do not seem to account for the differences between the Kimenyi-Mbaku birth rate data and the *Vital Statistics* data, we considered ways in which errors in data entry could cause such differences. We first considered misreadings of the data tables in *Vital Statistics*, examining the possibility that a research assistant entered one of the columns "White," "All Other," or "Black" instead of the column for "All Races" when dealing with birth rates. A look at the national-level changes eliminated the possibility that "All Other" or "Black" could have been entered, because those statistics have a national average change (from 1970 to 1980) of 132.9 and 177.7 respectively. We are looking for something with a change between about 20 and 40 to get a weighted mean change of 68.51. This left us with the possibility that the authors could have used the "White" data from the tables, which indicates a national

average difference of 55.9. While this is higher than what we expect to find in the correct replication data, we entered all of the state-level data to consider this possibility. Our results:

K/M WDBUW87	DBUW87 for whites	WDBUW87 weighted, for whites
68.51	38.95	77.10
(30.71)	(21.63)	(65.72)

While these values were certainly closer to the Kimenyi-Mbaku results than our other attempts, it was clear (especially from the standard deviation) that this error was also not the source of the differences in data.

2. The *Vital Statistics* data table for 1970 is set up in a way that encourages a different mistake. Although the list of states is alphabetical, the missing states (of which there are 11) are completely absent. Instead of keeping these states in the list and reporting their statistics as missing, the list simply deletes them. If one recorded the statistics straight through, without skipping the appropriate states, one would have states and data mismatched. The assignment of weights would then be based on this faulty data entry. We entering the data in this way, and the resulting descriptive statistics follow:

K/M WDBUW87	DBUW87 w/ mismatched #s	WDBUW87 wtd, w/ mismatched #s
68.51	76.24	141.56
(30.71)	(89.02)	(206.89)

Clearly, this approach does not yield the Kimenyi-Mbaku results.

3. It is possible that the District of Columbia could be listed alphabetically in some data sets but listed separately (such as with Guam, Puerto Rico, and/or the U.S. Virgin Islands) in others. If this caused errors in the data entry, an entire set of data could be misassigned. However, upon looking through all of the sources that we considered, there are only two ways in which the District of Columbia is listed: alphabetically (after Delaware) and geographically (in the S. Atlantic category). Since alphabetical order is straightforward, and since geographic order forces one to make sure *each* state is numbered correctly in data entry, it seems unlikely that the District of Columbia could have caused confusion.

Hence we found no way to replicate an entry error.

E. Differences in Weighting

1. While the method of weighting in the Kimenyi-Mbaku paper appears to be straightforward (p. 48), our inability to reproduce their weighted-change-in-birth-rate mean drove us to reconsider our understanding of the weights. It would be consistent with their theoretical framework to weight according to AFDC benefit levels in 1970 rather than birth rates in 1970. However, this presents a methodological problem, in that there is a full data set (51 observations) on which to base the weighting, but (due to

missing 1970 birth rate data) there are only 40 observations for which weighted birth rates can be calculated. This weighting scheme would also not resolve the basic problem: the unweighted change in birth rates appears to be too high to start with. This would be true for any alternative scheme that assigns equally-sized groups weights of one, two, and three. This also eliminates the possibility that the weights were assigned backwards (*largest* weight to *lowest* 1970 birth rate).

2. One reviewer suggested that the entire problem revolves not around erroneous data, but rather the meaning of "weighted" in the regression results. Perhaps instead of creating a weighted dependent variable, as the original paper seems to suggest, the paper reports weighted regressions in the more traditional sense—each observation of all variables weighted for importance, rather than weighting only the dependent variable. Unfortunately, when we replicate the regressions weighting each observation (with the 1, 2, 3 weights suggested in the paper) we continue to get negative coefficients on the AFDC variable: a coefficient of –10.1 in the smallest model, with coefficients of –12.46, -10.28 and –11.42 in the succeeding models.

3. Our final idea for a potential weighting scheme was weights of 1.01, 1.02, and 1.03 (rather than 1, 2 and 3). Although this seemed unlikely, it had the potential to convert the *Vital Statistics* data into the weighted mean difference reported by Kimenyi and Mbaku. The results of using this weighting scheme (replacing weights of 1 with 1.01, etc.) were as follows:

K/M WDBUW87	WDBUW87, wt =1.01, etc.	WDBUW87, % change, wt =1.01, etc.
68.51	68.29	68.11
(30.71)	(29.75)	(23.04)

While these values are the closest yet, they are not close enough for us to conclude that Kimenyi and Mbaku used these weights. This weighting scheme also yields regression results dramatically different from those reported by Kimenyi and Mbaku, reversing the sign on the AFDC variable.

F. The Final Decision

1. Having exhausted all of the available sources and all of our ideas regarding possible data entry errors or weighting scheme differences, it was necessary to choose data for the variable WDBUW87. In our judgement, the best source is *Vital Statistics of the U.S.* Since both the U.S. Census Bureau and the *Statistical Abstract of the U.S.*, the sources emphasized by the original authors, cite this publication as the primary source for birth rate information, it seemed a reasonable choice. Furthermore, it was the only primary source with detailed statistics regarding the number of births to unmarried women per 1000 live births. Since this is the statistic Kimenyi and Mbaku say they are using, we have no reason to think their data should differ from those we found in *Vital Statistics*. The fact that their reported mean and standard deviation differ substantially from our own remains a mystery.

Final Sources: *Vital Statistics of the U.S.*, 1970, Vol. 1, Table 1-34, p. 1-31
Vital Statistics of the U.S., 1980, Vol. 1, Table 1-36, p. 1-60

Source for the 1975 data: U.S. Department of Health and Human Services (Department of Health, Education, and Welfare). *Vital Statistics of the U.S.,* 1975, Vol. 1, Table 1-34, page 1-50.

Source for the 1985 data: U.S. Department of Health and Human Services (Department of Health, Education, and Welfare). *Vital Statistics of the U.S.,* 1985, Vol. 1, Table 1-77, pp. 205-219.

Source for the 1990 data: U.S. Department of Health and Human Services (Department of Health, Education, and Welfare). *Vital Statistics of the U.S.,* 1990, Vol. 1, Table 1-83, p. 210.