Model Sensitivity in Panel Data Analysis: Some Caveats About the Interpretation of Fixed Effects and Differences Estimators

by

Terra G. McKinnish
Department of Economics
Campus Box 256
University of Colorado
Boulder, CO 80309-0256
(303) 492-6770
terra.mckinnish@colorado.edu

Generous comments by Dan Black and Seth Sanders are gratefully acknowledged, as is financial support from the National Science Foundation. Keenan Dworak-Fisher, Stephen Fienberg, E. Michael Foster, Jacob Klerman, Warren Sanderson, Jeffrey Smith, Lowell Taylor, and seminar participants at the Economic Demography Workshop at the 1998 Population Association of American Meetings, University of Pittsburgh, University of Western Ontario, and IBS seminar at University of Colorado, also provided useful comments.

Abstract

This paper shows that first-differences or fixed-effects models may understate the effect of interest because of the variation used to identify the model. For example, much of the recent research estimating the relationship between AFDC benefits and fertility has used fixed-effects models. If the cross-sectional variation in AFDC benefits is removed, then the variation over time within states left for identification will largely reflect transitory, idiosyncratic changes in welfare benefits, which are unlikely to affect fertility. In some sense, the independent variable as being measured with error, since it is contaminated with these temporary fluctuations. Estimates obtained using fixed-effects models are small because the "signal" of permanent changes is being overwhelmed by the "noise" of transitory changes.

Two empirical examples are presented: one on the relationship between AFDC and fertility and the other on the relationship between local economic conditions and AFDC participation. Coefficient estimates from first-differences, long-differences, and fixed-effects models are compared. These estimates differ in ways that are consistent with the presence of measurement error. Attempts are made to obtain consistent estimates using instrument variables. Because no valid external instrument is apparent in the first example, the lagged internal instruments suggested by Griliches and Hausman (1986) are considered, but the analysis suggests that most of these instruments are "weak." In the second example, the decline of the steel industry is used as an instrument that is correlated with the permanent change in the prospects of low-skill workers. Estimates from the 2SLS model are substantially larger in magnitude than those obtained with first-differences or fixed-effects.

Keywords: panel data, measurement error, model misspecification, welfare programs

I. Introduction

Researchers often want to predict the effect of a policy intervention, or some other sustained change in economic or social conditions, on individual behavior. For example, in *Losing Ground* (1984), Charles Murray discusses structural changes made to the Aid to Families with Dependent Children (AFDC) program during the 1960's. These include the elimination of the man-in-house rules that denied welfare benefits to women who cohabitated with a man, and the large and sustained increase in welfare benefits that occurred during this decade. Murray blames these changes for the subsequent increase in out-of-wedlock fertility in the US. Another example is *When Work Disappears* (1996), in which William Julius Wilson argues that many of the problems plaguing today's inner cities, such as welfare dependency and crime are "fundamentally a consequence of the disappearance of work" (p.xiii). He argues that the decline of the manufacturing sector, which provided relatively high wages to low-skill workers, created a climate of joblessness in which these problems are prone to develop. Both of these substantive arguments deal with the effect of some sustained change in conditions on individual behavior.

Unfortunately, the data used by researchers often contain more information on transitory fluctuations in conditions than on permanent policy interventions or long-term changes in the economic or social environment. In order to study how welfare benefit cuts would affect fertility rates, researchers typically use all of the available variation in welfare benefits. Some of this variation reflects persisting differences across states and over time in benefit level, but some merely reflects temporary, idiosyncratic differences that are unlikely to affect fertility. Similarly, Wilson and many others have theorized that the decline in opportunities for low-skill workers over the past several decades is responsible for increasing a variety of socially detrimental behaviors within the low-income population. The empirical research on this question has typically used variation in economic indicators that reflects both the overall structural decline of economic opportunities for low-skill workers as well as transitory, business cycle fluctuations in these opportunities.

If a researcher is specifically interested in the effects of sustained changes in the independent variable rather than transitory ones, the independent variable is often an imperfect measure of the true policy-relevant phenomenon. Well-known analytical results, discussed below, show that when the independent variable is measured with error, estimates of the relationship between the dependent and independent variable are inconsistent. Under certain assumptions, the coefficient estimate is attenuated towards zero. In addition, researchers often use methods of analysis that eliminate much of the most relevant variation in the data in order to cope with another common empirical problem: omitted unobserved characteristics that are potentially correlated with both the dependent variable and the independent variable of interest. Under the assumption that these confounding unobserved characteristics do not vary over time, this omitted-variable bias can be eliminated by estimating a fixed-effects or differences model on panel data.

Fixed-effects and first-differences models are becoming increasingly popular in a wide variety of literatures, such as the effect of AFDC benefits on fertility (Ellwood and Bane, 1985; Jackson and Klerman, 1994; Clark and Strauss, 1998; Matthews, Ribar, and Wilhelm, 1997; Argys, Averett, and Rees, forthcoming), the effect of local labor market conditions on AFDC participation (Fitzgerald 1994; Hoynes, forthcoming), the effect of AFDC on female headship (Moffitt, 1994; Hoynes, 1997), the effect of state characteristics on AFDC expenditures (Ribar and Wilhelm, 1999), the effect of abortion access and funding on abortion rates (Blank, George and London, 1996; Levine, Trainor and Zimmerman, 1996), the effect of school quality on student outcomes (Card and Krueger, 1992; Murray, Evans, and Schwab, 1995), the effect of prison populations on crime rates (Levitt, 1996), the effect of state alcohol policies on traffic fatalities (Dee, 1999), and the effect of domestic violence resources on intimate partner homicide (Dugan, Nagin, and Rosenfeld, 1999). It is relatively common for researchers to find that once they control for individual fixed effects, the coefficient estimate of interest becomes considerably smaller, and often insignificant. This is often interpreted as a finding of no effect.

Fixed-effects or first-differences models can correct the problem of unobserved (time-constant) characteristics, but in doing so eliminate a large amount of variation in the data. This tends to aggravate problems with measurement error bias. In the simple regression context, measurement error in the independent variable typically causes coefficient estimates to be attenuated towards zero. Fixed-effects and differences estimators exacerbate this attenuation bias. Unfortunately, the attenuated estimate might be misinterpreted as a finding of little or no effect. Because independent variables rarely capture exactly the factor that actually influences behavior, it is likely that results from much of the research employing panel data estimation are being given precisely this misinterpretation.

This paper provides evidence of the sensitivity of panel data estimates to models specification, and demonstrates that this sensitivity arises from the fact that different specifications make use of the long-term and short-term variation in the data to different degrees. In doing so, this paper joins a small literature that has demonstrated the potential costs associated with using longitudinal data to correct for unobserved heterogeneity. Freeman (1984) shows that estimates of the effects of unions on wages obtained using longitudinal data are generally biased downward due to misclassification of union status. Baker, Benjamin and Stanger (1999) reconcile diverse findings in the minimum wage literature by demonstrating that models using low frequency variation, such a first-differences models, tend to estimate very small effects.

In this paper, two empirical examples are used to illustrate three general methodological points. First, even though the independent variable may not be measured with error, per se, it still may mismeasure the factor that we consider relevant. For example, we may know the exact value of state welfare benefits from administrative records, but the variation in these benefit levels may reflect more than just the structural changes in generosity about which we would like to make inferences. Second, simply estimating and comparing first-differences, long-differences, and fixed-effects models is a useful diagnostic for detecting measurement error in the independent variable. Third, the measurement error literature provides considerable guidance on how to correct measurement error bias, even in the case of differenced data. Unfortunately, without external instruments, any method of obtaining a consistent estimate of the true coefficient requires very strong assumptions about the form of the measurement error.

This paper presents two empirical examples related to research on welfare programs that illustrate the sensitivity of estimates to whether the statistical model uses short-term or long-term variation to identify the relationship of interest. In both of these examples, analysis using fixed-effects or differences specifications is likely to understate the effect of interest due to transitory variance in the variable of interest. The first empirical example examines the relationship between AFDC benefits and the fertility rates of young women. The second example looks at the relationship between county-level earnings and AFDC expenditures. Two separate empirical examples using different data sets are presented in order to emphasize that the potential misinterpretation of panel data estimates is a widespread problem in social science research. In both of examples, differences in the coefficient estimates obtained from fixed-effects, first-differences and long-differences models are consistent with the presence of measurement error. The results suggest that current research using these models could substantially understate the relationships of interest.

In both examples, Instrumental Variables (IV) are used in an attempt to correct the measurement error. In the first empirical example, having no external instrument, the method suggested by Griliches and Hausman (1986), using lagged values of the independent variable as instruments, is investigated. Under appropriate assumptions this method does produce consistent estimates, but careful analysis demonstrates most of these lagged instruments are in fact "weak" instruments. In the second example, the decline of the steel industry is used as an instrument that captures a long-term change in the economic prospects of low-skill men. The resulting IV estimates confirm the belief that fixed-effect estimators understate the effect of long-term changes in economic conditions on AFDC participation.

Finally, the analysis in the second example is repeated using Unemployment Insurance (UI) expenditures as the dependent variable rather than AFDC expenditures. Because the UI program is designed to respond to transitory fluctuations in economic conditions, the results from the fixed-effects, first-differences, long-differences, and IV analysis should reflect the fact that short-term changes in economic conditions have a larger effect on UI expenditures than long-term changes. The coefficient estimates from the analysis of UI expenditures exhibit exactly the opposite patterns of those obtained analyzing AFDC expenditures, consistent with the prediction that UI expenditures are much more responsive to transitory variation in economic conditions than AFDC expenditures.

The remainder of the paper is organized as follows. Section 2 reviews basic measurement error results for panel data. Section 3 presents the findings of the first empirical example. Section 4 discusses the difficulties with using lagged internal instruments to correct measurement error. Section 5 presents the findings of the second empirical example, and Section 6 offers conclusions.

II. Measurement Error Results for Panel Data

This section reviews the basic analytical measurement error results for panel data, which will be used in the later empirical examples.

Consider the following model:

(1) $$Y_{it} = a_i + b\, Z_{it} + e_{it}$$

(2) $$X_{it} = Z_{it} + n_{it}$$

For example, in the first empirical application, Y_{it} represents the birth rate in state *i* at time *t*, X_{it} represents the welfare benefit, Z_{it} represents the sustained component of the welfare benefit and n_{it} the transitory component of the welfare benefit. The term "measurement error" is not being used in the conventional manner, because we know the precise value of the AFDC benefit. But it is still the case that the benefit level (*X*) may actually mismeasure the factor that actually affects fertility, the sustained component of AFDC benefits (*Z*). For now, the classical measurement error model is assumed, in which n_{it} and e_{it} are purely white noise, both uncorrelated with Z_{it}. It is also assumed that a stationary process generates Z_{it}.

The conventional measurement error result is for the case where $a_i = a \;\; \forall i$. In this case, if a pooled cross-sectional OLS regression of *Y* on *X* is estimated:

(3) $$Y_{it} = a + b_p X_{it} + e'_{it}$$

then:

(4) $$p\lim\left(\hat{b}_p\right) = b\left(\frac{s_Z^2}{s_Z^2 + s_v^2}\right)$$

This is the well-known result that measurement error causes coefficient estimates to be attenuated towards zero (e.g. Greene, 1993; Johnston and Dinardo, 1997). The attenuation exists because some of the variation in *X* is due to variation in *v*, which has no effect on the outcome *Y*.

In the more general case, however, when there are individual or state fixed effects:

$$\text{(5)} \qquad \text{plim}\left(\hat{b}_p\right) = b\left(\frac{s_Z^2}{s_Z^2 + s_n^2}\right) + \left(\frac{Cov\left(a_i, Z_{it}\right)}{s_Z^2 + s_n^2}\right),$$

(e.g. Johnston and Dinardo, 1997). The first term captures inconsistency due to the measurement error. The second term captures the inconsistency due to the omitted individual or state effect.

If the individual or state fixed-effect cannot be measured directly, one way to eliminate the inconsistency due to the omitted effect is to difference the data. If first differences are used, the model becomes:

$$\text{(6)} \qquad Y_{it} - Y_{it-1} = b_{fd}(X_{it} - X_{it-1}) + (e'_{it} - e'_{it-1}),$$

In this case:

$$\text{(7)} \qquad \text{plim}\left(\hat{b}_{fd}\right) = b\left(\frac{s_Z^2(1-r)}{s_Z^2(1-r) + s_n^2}\right),$$

where r is the correlation between Z_{it} and Z_{it-1} (Griliches and Hausman, 1986). This eliminates the second term from Equation (5), because differencing the data subtracts out the state fixed-effect. The measurement error term, however, shows greater attenuation towards zero for any positive correlation between Z_{it} and Z_{it-1}. If *Z* is highly correlated over time and two observations of *X* from adjoining time periods are differenced, then most of the information about *Z* will be eliminated, leaving primarily variation due to the noise component, *v*. [1]

Another way to eliminate the state-specific fixed effects is to difference observations that are more than one period apart, producing a long-differences model:

[1] Bound and Krueger (1991) give an example in which differencing the data reduces the bias from measurement error because the error is more correlated over time than the signal.

(8) $$Y_{it} - Y_{it-j} = b_{ld}(X_{it} - X_{it-j}) + (e'_{it} - e'_{it-j})$$

For this long-differences model, the result is:

(9) $$\text{plim}\left(\hat{b}_{ld}\right) = b\left(\frac{s_Z^2(1 - r_j)}{s_Z^2(1 - r_j) + s_n^2}\right),$$

where r_j is the correlation between Z_{it} and Z_{it-j} (Griliches and Hausman, 1986). As long as $r_j < r$, the long-differences model will be less inconsistent than the first-differences model. By taking differences of observations that are less correlated with each other, the variance of the signal is increased relative to the noise. If Z_{it} has a declining correlogram, then $\text{plim}\left(\hat{b}_{ld}\right)$ converges to the cross- sectional result $bs_v^2/(s_v^2 + s_z^2)$ as j becomes large.

Finally, probably the most common method of dealing with the state or individual fixed effects is to estimate a fixed-effects model. In this case, the deviations of the observations from the variable means are used in the regression:

(10) $$Y_{it} - \bar{Y}_i = b_{fe}(X_{it} - \bar{X}_i) + (e'_{it} - \bar{e}'_i) \ .$$

In practice, this model is often estimated by including indicator variables for the cross-sectional units in the regression.

The plim result for the fixed-effects model is:

(11) $$\text{plim}\left(\hat{b}\right) = b\left(\frac{s_z^2 - \frac{1}{T^2}\left[T + 2\sum_{k=1}^{t-1}(T - K - 1)\, r_j\right]}{s_z^2 - \frac{1}{T^2}\left[T + 2\sum_{k=1}^{t-1}(T - K - 1)\, r_j\right] + \left(\frac{T-1}{T}\right)s_n^2}\right),$$

(Griliches and Hausman, 1986).

Griliches and Hausman (1986) summarize a number of general results comparing the fixed-effects, first differences, and long-differences models under the assumption that Z_{it} is independent of v_{it} and stationary with a declining correlogram.[2] They show that for T>2:

(a) The inconsistency of fixed-effects and long-differences is less than that of first-differences.
(b) The inconsistency of the (T-1) long-differences estimator is less than that of the fixed-effects estimator.
(c) The relative inconsistency of differences shorter than (T-1) and fixed effects depends on more specific characteristics of the correlation structure.
(d) In the absence of measurement error, fixed-effects and differences models are all consistent.

If the estimates from the three different models are very different, and the pattern of these estimates are consistent with the results in equations (7), (9), and (11), then this could indicate bias due to measurement error. Furthermore, as long as the signal, Z_{it}, and the error, n_{it}, are uncorrelated, the estimates from all three models should be attenuated toward zero. The panel data estimates should provide a lower bound on the magnitude of the effect of interest.

These results suggest a strategy for proceeding with the empirical analysis. Estimates are obtained from fixed-effects, first-differences and long-differences models. In the absence of measurement error, these models should produce roughly similar coefficient estimates. In the presence of measurement error, the first-differences coefficient estimate should be smaller in magnitude than the fixed-effects coefficient estimate, and the coefficient estimates should increase in magnitude as longer differences are used. If both patterns are present, this suggests the presence of measurement error and indicates that all of the coefficient estimates obtained from these models are potentially attenuated towards zero.

There are other misspecifications that could generate the pattern in the fixed-effects, first-differences and long-differences models described above. For example, if there are unobserved time-varying characteristics that are correlated with both *Y* and *X*, this will bias the estimates of the coefficient on the change in *X*, even without measurement error. The bias will tend to be smallest for first-

[2] These results hold even without all of the assumptions of the classical error model. They hold even if the measurement error is correlated over time, as long as the correlation of the measurement error is smaller than the correlation of the signal. They can also hold for the case in which *Z* is not stationary, if *v* follows a stationary process. See McKinnish (1999) for details.

differences, since the unobserved characteristic will largely be differenced out, but larger for differences over longer periods that allow more time for the unobserved characteristic to change. McKinnish (1999) also shows that omitted lagged effects of X will also generate the described pattern in the panel data estimates. Furthermore, McKinnish (1999) shows that under reasonable conditions, these three forms of misspecification: measurement error, omitted time-varying characteristics, and omitted lagged effects are empirically indistinguishable from one another using traditional specification tests, unless instruments are available.

In the first empirical example, we document the pattern in the panel data estimates we describe, but, lacking an adequate external instrument, are not able to rule out alternative misspecifications as the cause. Even so, our results from the first example demonstrate an alarming sensitivity to specification that alone provides caution about the way panel data models are typically estimated and interpreted. In the second empirical example, we are able to perform additional analysis that provides convincing evidence that the pattern observed in the panel data estimates is in fact due to the differential effect of long-term and short-term changes in economic conditions.

III. Welfare and Births to Young Women

The AFDC program has been criticized by opponents who argue that, by lowing the cost of supporting a child, this program increases birth rates among poor women. This section examines the effect of welfare generosity on birth rates of young women, using a panel of state-level data on age-specific birth rates and AFDC benefit levels from 1973-92.[3] Much of the early analysis on this topic was cross-sectional in nature, using only variation in AFDC across states at one point in time to identify the effect. These studies include Janowitz (1976), Moore and Caldwell (1977), Ellwood and Bane (1985), Singh (1986), Duncan and Hoffman (1990) and Schultz (1994). Several of these studies found that there was a *negative* relationship between AFDC generosity and birth rates.

These studies and others like them were often criticized for ignoring the effect of unobserved time-constant state characteristics that might confound the relationship between AFDC and fertility. This

[3] The AFDC program was replaced by the Temporary Aid to Needy Families (TANF) program in 1997.

criticism seems well-founded given the evidence provided in Table 1, which lists the AFDC benefit and birth rate among white teens for four states in 1980. The states with low welfare benefits, Arkansas and Mississippi, have high birth rates for white teens, while the high benefit states, Connecticut and Wisconsin, have relatively low teen birth rates. The cross-sectional variation would seem to indicate that more generous benefits lower teen fertility. Of course, the reason that Arkansas and Mississippi have higher teen birth rates than Connecticut and Wisconsin probably has a lot more to do with differences between the states in social and political climate than AFDC generosity (see Moffitt, 1992). Unfortunately, these cultural differences are hard to observe, quantify, and control for in a regression.

In more recent research, investigators deal with this criticism by using panel data and adding fixed-effects. Jackson and Klerman (1994) analyzed annual age and race-specific state-level birth rates from 1968-88. After controlling for state fixed-effects, they found a significant, positive relationship between AFDC and birth rates. Matthews, Ribar and Wilhelm (1997) and Clark and Strauss (1998) performed a similar analysis and also find a small, positive effect of AFDC benefits. Argys, Averett, and Rees (forthcoming) used individual-level data from the National Longitudinal Survey of Youth (NLSY) to estimate the effect of the incremental benefit for an additional child on the fertility of AFDC recipients. Once regional fixed-effects are included, the coefficient estimate becomes small and insignificant.

Rosenzweig (1999) uses the NLSY to estimate the effect of welfare benefits on the probability a woman has a non-marital birth by the age of 22, controlling for state and cohort fixed effects. His measure of welfare generosity is the average of welfare benefits offered in the woman's state of residence between the ages of 12 and 20. This measure arguably captures much longer-term differences in welfare benefits than the contemporaneous benefits used in other studies. Using this specification, Rosenzweig finds a substantial positive effect of welfare generosity on non-marital childbearing, although a reanalysis by Hoffman and Foster (forthcoming) using the Panel Study of Income Dynamics (PSID) shows that this result is sensitive to specification.

Some other studies have focused on female headship, which combines fertility and marital decisions. Moffitt (1994) used Current Population Survey (CPS) data and includes state fixed effects and random state time trends in his model. He found the effect of AFDC benefits on female headship was insignificant for whites by significant for blacks. Hoynes (1997) performed a similar analysis using data from the Panel Study of Income Dynamics (PSID), adding individual fixed effects to the model. She found the effect of AFDC benefits to be insignificant for both whites and blacks. Hoynes stated that these results "suggest that previous studies may have overstated the effect of welfare programs on family structure"(p.91). The findings in this paper suggest that Hoynes' results might better be attributed to the lack of useful time series variation left after cross-sectional effects are removed.

Researchers use fixed-effects estimation to deflect criticisms made of earlier cross-sectional research on welfare and fertility. It is important, however to consider the costs associated with the use of fixed-effects models (see Hamermesh, forthcoming). A regression of AFDC benefits on state and year fixed-effects produces an R^2 statistic of 0.95. Once fixed effects are included in the model, 5% of the variation in AFDC benefits remains to identify the effect on fertility. Whether or not the remaining 5 variation can be used to identify the coefficient of interest depends on the nature of that variation. While the cross-sectional differences in benefit levels across states tend to persist over time, the changes across time within states tend to be much more transitory and idiosyncratic.

Figure 1 graphs the nominal AFDC benefit for a family of four over time in three states: Massachusetts, Montana and Mississippi. This benefit level is guaranteed to a family of four with no additional income, a legislated parameter of the program that is known with certainty. There are clearly permanent cross-sectional differences between the three states in their general benefit level. Looking within each state over time, the graph shows that that Massachusetts's time trend contains frequent small fluctuations. Benefits change almost annually in both positive and negative directions. Because the benefit level is not measured with error, these fluctuations reflect actual legislative action. Mississippi's time trend, on the other hand, contains no such fluctuations. The state changes benefits only twice in the period from 1970-94. Montana's time trend shows both short-term and long-term variation. There is a one-year change in 1974, but then a long-term shift in nominal benefits in 1982.

Figure 1 shows that the cross-sectional differences in benefit level eliminated by differencing are fairly permanent, but the time-series variation is both short-term and long-term. Economists have long argued that permanent changes in one's economic situation should affect behavior differently than transitory changes. For example, according to Friedman's (1957) permanent income hypothesis, consumption is a function of permanent income, but does not respond to transitory fluctuations. It therefore seems reasonable to believe that fertility will respond to changes in AFDC generosity that are expected to persist over time, but not to changes that are temporary. As childbearing is a commitment to consumption not only in the period of birth but to future periods as well, we would further expect fertility decisions to be relatively non-responsive to changes in welfare generosity that are temporary. Much of the time-series variation remaining once the cross-sectional effects are eliminated could be temporary fluctuations that have little bearing on fertility rates.[4]

In order to explore this differential effect of long-term and short-term changes in welfare generosity, a variety of panel data models are estimating using state-level data from 1973-92. The regression equation selected is a basic linear regression with a minimum of controls to act as a baseline for what has been estimated in this literature:

$$(12) \qquad \Delta BirthRate_{st} = b_1 \Delta AFDC_{st} + b_2 \Delta Earnpc_{st} + Year_t b_3 + m_{st}\,.$$

where Δ operator indicates a first difference, $BirthRate_{st}$ is the logged birth rate for a particular demographic group in state s at year t, $AFDC_{st}$ is the logged real AFDC benefit and $Earnpc_{st}$ is the logged real earnings per capita. $Year_t$ is a vector of year indicators that control for national shocks that might affect both AFDC benefits and birth rates. The long-differences models also take the form of Equation (12), except that Δ is replaced by Δ_j to indicate a j-period long difference.

The fixed-effects version of equation (12) is:

$$(13) \qquad BirthRate_{st} = b_0 + b_1 AFDC_{st} + b_2 Earnpc_{st} + Year_t b_3 + State_s b_4 + e_{st}\,.$$

[4] Moffitt (1994) and Rosenzweig (forthcoming) have also noted that using year-to-year variation in AFDC benefits is likely to understate the effect of interest.

where $State_s$ is a vector of state indicator variables.[5]

Counts of births by state, year and demographic group for 1973-92 are obtained using the Detailed Natality Files from the National Center for Health Statistics. For each state and year, these files contain either a 50% or 100% sample of birth certificates. The analysis in this paper uses birth rates for white women for three age groups: ages 15-19, 20-24 and 25-29.[6] Denominators for the birth rates are obtained from various US Census files. Births that occur in the first 9 months of a year are attributed to the previous year. Therefore, the model estimates the effect of benefits on fertility near the time of the conception, when the decision-making component presumably took place. The independent variable of interest is the AFDC benefit for a family of four with no additional income, obtained from various years of *The Green Book* publication of the U.S. House Ways and Means Committee.[7] All regressions include real earnings per capita, obtained from the Bureau of Economic Analysis' (BEA) Regional Economic Information System (REIS).[8] Earnings are divided by the population age 10 and older in order to avoid endogenous effects of changes in the birth rate.

Table 2 reports estimates of the coefficient on AFDC benefits for five different models: the pooled cross-section, fixed-effects, first-differences, 3-year long differences and 5-year long-differences. The three columns correspond to the three different age groups analyzed: ages 15-19, 20-24, and 25-29. The first row reports the coefficient estimates from the pooled cross-section. For white teens, the coefficient on AFDC benefits is negative and strongly significant, while for white women ages 20-24 and 25-29, the coefficients are positive and significant. The perverse negative coefficient for white teens is highly suggestive of the presence of unobserved state effects.

---

[5] The variance of the disturbance depends on state population. Huber-White robust standard errors are reported, rather than using population weights to correct the heteroskedasticity. This prevents larger states such New York and California from dominating the results.

[6] Results using black women can be found in McKinnish (1999). Coefficients for black women ages 20-24 and ages 25-29 exhibit the same patterns as those reported here for white women, but the results for black teens do not.

[7] Robert Moffitt provided benefits through 1989 in electronic form. Even though the benefit for a family of four is used, most of the variation in this measure reflects changes in the baseline benefit for a mother with one child, rather than changes in the incremental benefits for additional children.

[8] AFDC and earnings variables are deflating using the July CPIU, base year 83-84. Because we log AFDC and earnings, and include year effects, the coefficients on real and nominal benefits are algebraically identical.

The second row reports the results of the fixed-effects model described in Equation (13). For all three demographic groups the coefficient is positive and significant. The results indicate that a 10% increase in AFDC benefits should be associated with roughly a 1.2 to 1.3% increase in births to young white women. The results of the first-differences model described in Equation (12) are given in the third row. Recall that if the model is correctly specified and there is no measurement error, the fixed-effects and first-differences estimates are both consistent. In this case, however, the coefficient estimates from the first-differences model are an order of magnitude smaller than the fixed-effects estimates and insignificant. For white teens, the fixed-effects coefficient is .1226, while the first-differences coefficient is .0148. For white women age 20-24, the fixed-effects coefficient is .1231, while the first-differences coefficient is .0170. For white women age 25-29, the fixed-effects coefficient is .1359, while the first-differences coefficient is only .0297. The fact that the first-differences estimates are closer to zero than the fixed-effects estimates is consistent with the analytic predictions from the measurement error model discussed in Section 2.

The fourth and fifth rows contain the results for 3 and 5-year long differences. For all three demographic groups, the estimates from 3-year long differences are larger than those from first differences, and the estimates from 5-year long differences are larger than those from 3-year long differences. The increasing size of the coefficient with longer differences is again consistent with the predictions in Section 2.[9] There might be a concern that because the sample is reduced as longer-differences are used, the pattern of results might reflect the changing sample. When the models in Tables 2 were re-estimated using only the observations from 1978-92, so that all models are estimated on the same sample, the pattern of results is virtually unchanged.[10]

These results show an alarming sensitivity to model specification, particularly because the variations in specification are traditionally considered rather trivial by researchers. Furthermore, the pattern of the sensitivity is consistent with the presence of measurement error, which suggests that much

[9] The models reported in Tables 2 were re-estimated adding additional controls for economic conditions and abortion access. Adding the controls does not change the fundamental nature of the results.
[10] See McKinnish (1999) for estimates on the reduced sample.

of the current literature on this topic could be substantially understating the relationship between long-term changes in AFDC benefits and fertility rates.

IV. Lagged Internal Instruments

The results in Table 2 indicate that research on the effect of welfare benefits on fertility potentially understates the magnitude of the relationship. In order measure the extent to which the effect is underestimated, the "measurement error" must be corrected. The most common correction for measurement error is to use Instrumental Variables (IV) estimation with an instrument that is correlated with the signal but independent of the measurement error. Unfortunately, it is difficult to find such an instrument for AFDC benefits. Griliches and Hausman (1986) point out that in the absence of an appropriate external instrument, it is possible to use lagged values of the independent variable as instruments to correct the measurement error bias.

Consider first the basic differences model from equation (8):

$$Y_{it} - Y_{it-j} = b(X_{it} - X_{it-j}) + (e_{it} - e_{it-j})$$

If the measurement error, v, is uncorrelated over time, then any value of X other than X_{it} and X_{it-j}, or any function of these values, is a valid instrument for $X_{it} - X_{it-j}$.

In practice, researchers do not make use of these internal instruments very often. This section provides evidence that these instruments have not worked well in applications because many of the theoretically valid instruments are in fact weak instruments. The lagged values of X produce consistent estimates under the appropriate assumptions about the structure of the measurement error. But if the lagged values of X are only mildly correlated with $X_{it}-X_{it-j}$, the IV estimator can have large bias in small samples and can be extremely imprecise (Bound, Jaeger and Baker, 1995).

Some simple calculations can show that many of these lagged instruments are potentially weak. When estimating a first-differences model, the simplest valid instruments for $X_{it}-X_{it-1}$ are X_{it-2}, X_{it-3}, and $X_{it-2}-X_{it-3}$ (Hsiao, 1986). Suppose for simplicity that the signal, Z, is generated by an AR(1) process. Under this assumption, the strength of these instruments can be evaluated by calculating $r_{1,-2}$, the correlation between X_{it}-X_{it-1} and X_{it-2}; $r_{1,-3}$, the correlation between X_{it}-X_{it-1} and X_{it-3}; and $r_{1,-2-3}$, the correlation between X_{it}-X_{it-1} and X_{it-2}-X_{it-3}. Assuming $0 \le r \le 1$:

$$(14) \qquad r_{1,-2} = \frac{-s_Z^2 r(1-r)}{\sqrt{2s_Z^2(1-r)+2s_v^2}\sqrt{s_Z^2+s_v^2}} \ge -\frac{1}{\sqrt{2}} r\sqrt{1-r}$$

$$(15) \qquad r_{1,-3} = \frac{-s_Z^2 r^2(1-r)}{\sqrt{2s_Z^2(1-r)+2s_v^2}\sqrt{s_Z^2+s_v^2}} \ge -\frac{1}{\sqrt{2}} r^2\sqrt{1-r}$$

$$(16) \qquad r_{1,-2-3} = \frac{-s_Z^2 r(1-r)^2}{\sqrt{2s_Z^2(1-r)+2s_v^2}\sqrt{2s_Z^2(1-r)+2s_v^2}} \ge -\frac{1}{2} r(1-r)$$

where r is still the correlation between Z_{it} and Z_{it-1}. The bounds are obtained by setting $s_v^2 = 0$. The correlation in equation (14) is maximized at $r = .6667$. The correlation in equation (15) is maximized at $r = .8$ and the correlation in equation (16) is maximized at $r = .5$. Substituting these values in produces:

$$(17) \qquad 0 \ge r_{1,-2} \ge -.2722$$

$$(18) \qquad 0 \ge r_{1,-3} \ge -.2024$$

$$(19) \qquad 0 \ge r_{1,-2-3} \ge -.125$$

Bound, Jaeger and Baker (1995) suggest that researchers report the F-statistics from the first stage regression. A small F-statistic indicates that the instrument is weak. For simple regression:

$$(20) \qquad F = \frac{r^2}{(1-r^2)}(n-2),$$

where r is the sample correlation of the dependent and independent variable. Calculating the minimum number of observations required to obtain an F-statistic of 10 under the most favorable conditions, produces a minimum of 127 observations for X_{it-2}, 236 observation for X_{it-3}, and 632 observations for X_{it-2}-X_{it-3}. It is even more informative to consider a typical set of circumstances, such as one in which

$\rho = .7,\ s_Z^2 = 2$, and, $s_V^2 = 1$. Under these conditions, $\rho_{1,-2} = -.1356$, $\rho_{1,-3} = -.0949$ and $\rho_{1,-2-3} = -.0407$. In this case, the minimum number of observations necessary to obtain an F-statistic of 10 is 536 for X_{it-2}, 1,102 for X_{it-3}, and 6,029 for X_{it-2}-X_{it-3}!

The above results indicate that even though many theoretically valid internal instruments may exist, there may only be a handful of instruments that meet some minimum standard of correlation with the right-hand-side variable. The intuition of the problem is fairly simple. If the signal is highly correlated over time, differencing two adjacent observations will leave almost no signal. Therefore, when instrumenting the first-differences, one is trying to instrument an observation that is almost entirely white noise. It is also apparent that instruments that are lagged several periods behind the independent variable and instruments that are differences of lagged observations will be particularly weak. It is possible to be more strategic in our choice of instruments. For example, we can:

(a) Instrument long-differences rather than first-differences
(b) Use instruments that overlap the independent variable rather than ones that lag behind it.

For example, consider a two-year long-differences model in which X_{it-1}-X_{it-3} is used as an instrument for X_{it}-X_{it-2}. In this case:

(21) $$\rho_{2,-1-3} = \frac{s_Z^2 \rho (1-\rho^2)}{2 s_Z^2 (1-\rho^2) + 2 s_V^2} \le \frac{\rho}{2}$$

Therefore:

(22) $$0 \le \rho_{2,-1-3} \le .5$$

Under the most favorable circumstances, only 30 observations are required to obtain a first-stage F-statistic of 10. Under the more reasonable conditions that $\rho = .7,\ s_z^2 = 2$, and, $s_v^2 = 1$, $\rho_{2,-1-3} = .1767$. Therefore, only 312 observations are needed to obtain an F-statistic of 10. This "overlapped" instrument is stronger than any of the possible instruments for the first-differences model. Notice, however, that this type of instrument is particular vulnerable to violations of the assumption of uncorrelated errors.

Table 3 reports the results obtained using lagged internal instruments to estimate the relationship between AFDC and fertility. The basic differences model from equation (12), which includes controls for per capita earnings and year effects, is used. Lagged AFDC benefits are used as instruments for first differences, 2-year differences and 3-year differences. The first column describes the instrument and reports the first-stage partial F-statistic on the instrument. The "rule of thumb" is that the first-stage F-statistic for the instrument should be greater than 10. Most of the instruments in the table are well below that, and many instruments have an F-statistic less than one. Only 5 instruments in the Table can be classified as "strong" instruments, all of which have F-statistics larger than 40. None of the instruments for first-differences are strong, reinforcing that it is better to instrument long-differences. Also, all of the strong instruments overlap the right-hand side variable, rather than lag behind it, illustrating point (b).

The second column in Table 3 gives the IV results for white women age 15-19. There is clearly tremendous variance in the coefficient estimates. Some of the coefficients are greater than 2, some are close to 0 and others are less than –1. This is not surprising given how weak many of the instruments are. The third column only displays the coefficients obtained using strong instruments. Once the weak instruments are removed, the various IV coefficients are extremely similar, in the neighborhood of .055. Furthermore, the coefficients in this column are all of reasonable sign and magnitude. This procedure is repeated for white women age 20-24 in the final two columns. Once again, there is a great deal of variance in the coefficients if all instruments are considered. When the results obtained with weak instruments are removed, the remaining coefficients are similar in size and are all of reasonable sign and magnitude. The results in this table reinforce the facts that most of the lagged internal instruments are weak, and that using the weak instruments will often produce imprecise or unreasonable estimates.

Even though the IV coefficients obtained using strong instruments are stable and of reasonable sign and magnitude, it is troubling that most of the coefficients are smaller than the OLS results for the same models. For white women age 15-19, the fixed-effects estimate is .1226, while the IV estimates are in the neighborhood of .055. If the attenuation bias due to measurement error has been corrected, the IV coefficients should be larger than the fixed-effects coefficients. For white women age 20-24, the IV results are very close in magnitude to the fixed-effects estimate of .1231. The lagged instruments used in Table 3 are only valid if the measurement errors are uncorrelated. These results suggest that perhaps the measurement errors are correlated, in which case the IV estimates are inconsistent.

Griliches and Hausman (1986) suggest a strategy for testing and then dealing with violations of the uncorrelated errors assumption. Unfortunately, they do not consider the problem of weak instruments, which makes their strategy very difficult for follow. The authors suggest that models using several different lagged internal instruments be estimated and then the coefficient estimates be compared. Significant differences in the estimates should signal a violation of the model assumptions. This strategy only works if there are a number of different internal instruments available to the researcher. In other words, if all of the strong instruments are highly correlated, as they are in Table 3, then they will produce similar coefficient evidence regardless of whether or not model assumptions are violated.

If the measurement errors are correlated over time, Griliches and Hausman suggest a set of assumptions under which internal instruments are still available:

(a) If the correlation of the errors is of the form MA(k), use lags from period t-k-1.
(b) If the measurement error is stationary and the signal is not, use changes in the variation of the signal over time to identify the coefficient of interest.

Based on the analytical results, it is unlikely that an observation from period t-k-1 would be sufficiently correlated with the right-hand side variable to act as a viable instrument. Therefore, strong instruments are only available in this if assumption (b) is valid or there is special knowledge of the correlation structure of the measurement errors. In general, it would appear that lagged internal instruments are not practical unless one is truly dealing with classical measurement error. Otherwise, external instruments appear to be the only way to obtain consistent parameter estimates. In the next section, a second empirical example is pursued. This second example has the advantage that there will be defensible external instruments that can be used to correct for measurement error.

V. Local Economic Conditions and AFDC Participation

This section examines the relationship between AFDC program participation and local labor market conditions, using county-level data on earnings and AFDC expenditures from 1969-93. There have been a number of studies examining the effect of economic conditions on AFDC participation. Early work that did not control for fixed-effects included Sanders (1993) and Fitzgerald (1995), both of whom found significant effects of economic conditions on the length of welfare spells. The next generation of studies, which uses fixed-effects models to study the effect of contemporaneous changes in economic conditions on AFDC participation, includes Fitzgerald (1994), Blank (1997), Miller and Sanders (1997), and Hoynes (forthcoming). The general consensus of this literature was that there is a statistically significant but economically modest effect of economic conditions on welfare participation. Some very recent studies have used more dynamic specifications of the relationship between state or local economic conditions and welfare caseloads (Bartik and Eberts 1999; Figlio and Ziliak 1999; Wallace and Blank 1999; Mueser et al. 1999; Klerman and Haider 2000). Most of these studies have found that lagged effects of economic conditions are very important, and that the resulting long-run elasticities can be substantial in magnitude. The observed importance of lagged effects of economic conditions on welfare caseloads demonstrates that the effect of changes in economic conditions on welfare caseloads depends on the persistence of those changes. If longer-term changes in economic conditions affect welfare participation more than short-term changes, non-zero coefficients will be generated on lagged values of the economic variables (see McKinnish, 1999).

Sustained changes in economic conditions might reasonably have a larger effect on AFDC participation than temporary changes. The various costs associated with welfare receipt include the time costs of establishing and maintaining eligibility, the depreciation of human capital while on the program, and social stigma. Given these costs, a single mother might prefer to wait out a short economic downturn than go on welfare. During a longer economic contraction, the woman is likely to exhaust other resources, such as support from other family members and Unemployment Insurance, and enter the welfare program. The transition from welfare to work likewise involves costs related to the job search and work expenses such as clothing and transportation. Therefore, women are more likely to exit welfare for work during periods of long-term economic growth when they can recover the costs of these investments. In addition, to the extent that long-term economic conditions affect family structure through the creation and dissolution of marriages, responses to long-term changes in economic opportunities should be greater than responses to short-term ones.

The basic regression model estimates how changes in county-level earnings affect AFDC expenditures:

$$(23)\quad \Delta Expend_{ist} = b_0 + b_1 \Delta Earnings_{ist} + \Delta Pop_{ist} b_2 + (State_s * Year_t) b_3 + e_{ist},$$

where Δ operator again indicates a first difference, $Expend_{ist}$ is logarithm of real AFDC expenditures for county i in state s and time t; $Earnings_{ist}$ is the logarithm of aggregate real earnings; Pop_{ist} is a vector that contains the logarithm of county population and the change in the logarithm of county population from the previous year; $(State_s*Year_t)$ is the interaction of state and year dummy variables. Because state-year interactions are included in the model, it is not necessary control for AFDC benefits, which vary only at the state level. The model is also estimated using long differences and using county-level fixed effects. Annual measures of county earnings, AFDC expenditures, and county population are obtained from the Bureau of Economic Analysis' (BEA) Regional Economic Information System (REIS), for 1969-93.[11]

Table 4 reports estimates of the coefficient on county-level earnings obtained from fixed-effects and differences models of the form in equation (23), as was done in Table 2 for the previous example. The coefficient from the pooled cross-section is -.251. Once county-level fixed-effects are added, the magnitude of the coefficient is substantially smaller. The coefficient from the fixed-effects model is -.069, and the coefficient from the first differences model is -.039. In the next three rows are the long-differences models. The 3-year long-differences coefficient is -.150; the 5-year long-differences coefficient is -.203; and the 7-year long-differences coefficient is -.232. A large cross-sectional sample (all counties in the U.S.) allows the use of very long differences. The coefficient on 20-year long differences is -.307. As in the previous example, the pattern in the coefficient estimates is not driven by the changing sample. Estimating the models using just the years 1977-93, so that all models are estimated

[11] AFDC expenditures are used as the dependent variable instead of AFDC caseload. This is because data on county-level AFDC caseloads over time are not available. Because benefit levels vary only at the state level, most of the variance in expenditures at the county level will reflect changes in caseloads.

on the same sample, produces the same pattern of coefficients.[12] Once again, the pattern of the results is consistent with analytical predictions reviewed in Section 2

If the measurement error model assumptions are correct, then all of these estimates are attenuated towards zero. Under these assumptions, the coefficient from 20-year long differences, -.307, acts as a lower bound on the true coefficient. This is more than 4 times the size of the coefficient on the fixed-effects model, and more than 7 times the size of the coefficient on the first-differences models. In this application, unlike the previous one, the fixed-effects estimate is much smaller in magnitude than those obtained with long-differences. Therefore, it seems prudent for researchers to estimate both fixed-effects and long-differences models whenever feasible. This acts as an informal test for measurement error, and provides a lower bound for the true effect. Currently, it is fairly rare for policy researchers to report long-differences results, unless they are constrained to do so by their data.

While it was difficult to think of an instrument that could be used to capture long-term changes in state-level AFDC benefits, there are a number of structural changes in economic conditions that have occurred over the last two decades that could be used as an instrument in this application. For example, during the 1980s, the steel industry went through a sustained decline. Figure 2 graphs the fraction of total earnings attributed to primary metals manufacturing over time for both the entire U.S., and the eight steel-producing states used in this analysis (listed below). [13] The graph shows that the fraction of total earnings attributed to primary metals manufacturing more than halved between 1979 and 1987.

Most of the previous studies in this literature have analyzed the effect of short-term movements in general economic conditions. This negative shock to the steel industry should have a substantially larger effect on welfare participation. One reason is that this was a permanent shock. Geographic localities affected by the collapse of steel manufacturing experienced a long-term economic event, rather than just a business-cycle fluctuation. Second, at the time of the shock, the steel industry employed primarily low-

[12] see McKinnish (1999) for results from the reduced sample.

[13] For the eight states analyzed in this section, primary metals manufacturing is almost exclusively steel manufacturing. In the 1970 Public Use Micro Sample data, 79 percent of primary metals workers in the 8-state region in 1970 were steelworkers.

skilled men and paid them relatively high wages.[14] Figure 3 maps the fraction of men employed in primary metals manufacturing in 1969 by county for the US. The outlined states are the 8 steel-producing states used in this analysis: Alabama, California, Indiana, Illinois, Michigan, New York, Ohio and Pennsylvania, which were the eight largest steel producers in 1969. The concentration of this industry in a small fraction of the total counties in the U.S. is evident in the map. The steel shock should have a substantial impact on the local economies in areas that were historically large steel manufacturers. Other counties, with very little steel employment, will be relatively unaffected by the shock.

Black, McKinnish and Sanders (2000) used IV analysis to show that the steel shock generated a sizeable increase in welfare expenditures. That analysis is replicated here. Two different specifications of the instrument are considered. For the first model, the first-stage regression is:

$$\Delta Earnings_{ist} = a_0 + (S_i * Time_t)a_1 + \Delta Pop_{ist} a_2 + (State_s * Year_t)a_3 + m_{ist}, \tag{24}$$

where S_i is the fraction of total men employed in county i in 1969 that are employed in primary metals.[15] Time is a vector of indicator variables: one for the years 1970-80 and then individual indicators for the years 1981-87.

For the second model, the first-stage regression is:

$$\Delta Earnings_{ist} = a_0 + a_1(S_i * FrMetalEarn_t) + \Delta Pop_{ist} a_2 + (State_s * Year_t)a_3 + m_{ist} \tag{25}$$

where *FrMetalEarn* is the fraction of total earnings in the US attributable to the primary metals industry. In both cases, the second-stage regression model is:

$$\Delta Expend_{ist} = g_0 + g_1 \Delta Ear\hat{n}ings_{ist} + \Delta Pop_{ist}\, g_2 + g_3(State_s * Year_t) + e_{ist}. \tag{26}$$

In both models, the instrument interacts a measure of dependence on steel with a measure of changes in aggregate demand for domestic steel. In the first model, time period dummies are used to capture these changes in aggregate demand. In the second model, these changes in aggregate demand are

[14] See Black, McKinnish and Sanders (2000) for evidence.

[15] Estimates of total male employment and total male employment in primary metals in 1969 by county are obtained using the fourth county population file C from the 1970 census. Our employment counts are derived from a combination of the 1970 5% and 15% long form questionnaires, so that our counts are based on a 20% sample of the U.S population.

measured more explicitly. The first-stage regression estimate the differential trends in earnings for the counties experiencing the decline of the steel industry relative to the counties that were never dependent on steel employment. Because of the large structural decline in the steel industry during the period under study, the instrument should be highly correlated with the permanent component of earnings, but not the transitory.

Table 5 reports estimates for the various models, like those reported in Table 4 for the full U.S. The sample is restricted to the eight steel states and the years 1970-87 to avoid problems with weak instruments. Outside of these eight states and this time period, steel is not a strong instrument for local economic conditions. The coefficient estimate on county-level earnings from the pooled cross-section is large and positive, again suggesting unobserved confounding. The first-differences estimate is -.192, while fixed-effects estimate is -.584. The 3-year and 5-year long-differences estimates are -.441 and -.537. The 7-year long-differences estimate, -.582, is almost identical to the fixed-effect coefficient. In the last two rows of Table 5 are the coefficient estimates obtained from the IV models described in equations (24)-(26). In the first model, the interaction of *S* with the Time Period indicators has a first-stage partial F-statistic of 12.4 and the coefficient estimate is -.723. In the second model, the interaction of S with the fraction of US earnings from primary metals has a first-stage partial F-statistic of 34.8, and the coefficient estimate is -.809. For both models, the instrument is sufficiently "strong", the coefficient estimates are substantially larger than the fixed-effects and long-differences estimates, and they are highly significant. This suggests that OLS estimators could be severely understating the relationship between persisting changes in economic opportunities and AFDC participation. Furthermore, these results indicate that the pattern observed in the fixed-effects and differences estimates in Tables 4 and 5 is not due to other forms of misspecification, such as omitted time-varying characteristics. If that were the form of misspecification generating the results, then the IV estimates should be closer to the less-biased first-differences estimate, rather than larger than the long-differences estimates.[16]

[16]This assumes that the instrument is uncorrelated with the time-varying unobservables. The fraction of national earnings from primary metals should measure changes in demand from foreign competition and technology that are

An additional test of the measurement error model would be to find a dependent variable similar to AFDC expenditures, but one that is more responsive to transitory changes in economic conditions than AFDC expenditures. If such a dependent variable were used, then there should be a very different pattern in the long-differences and IV coefficients. Unemployment Insurance (UI) is another income maintenance program, but one that is specifically designed to act as a buffer to business cycle fluctuations. The program typically provides qualified recipients 50-70% of their previous wages for up to 26 weeks. According to the 1997 Green Book publication of the U.S. House Ways and Means Committee: "The [Unemployment Insurance] program has two main objectives: (1) to provide temporary and partial wage replacement to involuntarily unemployed workers who were recently unemployed; and (2) to help stabilize the economy during recessions."(p.327)

As the UI program is designed to sustain workers through temporary job losses, UI expenditures should be quite sensitive to transitory fluctuations in county-level earnings. Because of limits on the duration of benefits, it is unlikely that UI expenditures are more responsive to long-term changes in economic conditions than short-term ones, and very possibly less so. Making a similar argument, Black, Daniels and Sanders (1998) show that UI expenditures were relative non-responsive to the large shocks to the coal economy during the 1970's and 1980's. Here that finding is extended to the long-differences and IV analysis presented in Tables 4 and 5. The dependent variables in equation (23) is switched from annual AFDC expenditures to annual UI expenditures, also obtained from the BEA data:

(26) $$\Delta_j UIExpend_{ist} = b_0 + b_1 \Delta_j Earnings_{ist} + \Delta_j Pop_{ist} b_2 + (State_s * Year_t) b_3 + e_{ist},$$

If UI expenditures are more responsive to transitory changes in economic conditions than persisting changes in economic conditions, the pattern of results from the different models should be exactly the opposite of what was obtained using AFDC expenditures as the dependent variable. The first-differences coefficient estimate on earnings should be *larger* in magnitude than the fixed-effects

---

exogenous to the individual county. The fraction of employment, measured a decade before the decline of steel, is also likely to be exogenous coefficient; the coefficient on county-level earnings should become *smaller* in magnitude as longer differences are used. The results obtained using the full US sample are reported in Table 6. The fixed-effects coefficient is -.097; the first-differences coefficient on county-level earnings is -.232; the coefficient on 3-year long-differences is -.222; the coefficient on 5-year long-differences is -.151; the coefficient on 7-year long-differences is -.115; and the coefficient on 20-year long-differences is -.031. The estimates exhibit the expected patterns.

Table 7 reports the UI expenditures results for the subset of 8 former steel-producing states. The long-differences estimates exhibit the same pattern here as in Table 6. The fixed-effects coefficient is -.342. The first-differences coefficient is -.548. Once again, taking longer differences causes the coefficient estimate to drop in magnitude, from a high of -.551 for 3-year long-differences to a low of -.258 for 7-year long-differences. The IV results are reported in the final two rows. The first-stage F-statistics are the same as they were in Table 5, because the right-hand side of the regression has not changed. The first IV model produces a coefficient estimate of -.079, which is considerably smaller than the coefficients obtained from any of the panel data models. The second IV model surprisingly produces a positive coefficient estimate of .453. These two IV estimates are the only coefficient estimate obtained in this second empirical example that are statistically insignificant. The sample sizes in this empirical application are so large that most of the standard errors are very small. While the steel instruments produce very similar and highly significant results when the dependent variable was AFDC expenditures, they produce highly imprecise results when the dependent variable is UI expenditures.

The results from this analysis provide additional evidence that the model sensitivity seen in Tables 4 and 5 is in fact a result of the differential impact of short-term and long-term changes in conditions on AFDC participation. By considering a dependent variable in which short-term changes should have a larger effect than long-term changes, we observe exactly the reversal in the pattern of coefficients that is predicted. These results from the analysis of UI expenditures further substantiate the claim that there are differential effects of short-term and long-term variation in conditions, and that when estimating panel data models, we must be cognizant of how those differential impacts affect our findings.

VI. Conclusions

Fixed-effects and first-differences models are extremely popular because the relationship of interest is often confounded by unobserved heterogeneity in the cross-section. Unfortunately, if the independent variable is an imprecise measure of the relevant factor, coefficient estimates from these models can be severely attenuated towards zero. The time series variation that remains after removing fixed effects often largely reflects idiosyncratic changes in the independent variable that have little influence on the decision of interest. This paper uses a measurement error model in which permanent changes in the independent variable are the signal and temporary changes are the noise.

Empirical examples are presented on the effect of AFDC on fertility and the effect of economic conditions on AFDC participation. Coefficient estimates from first-differences, long-differences, and fixed-effects models are compared and found to differ in ways that are consistent with the presence of measurement error. Because no external instrument exists for the first example, the lagged internal instruments are considered, but the analysis suggests that most of these instruments are "weak." The decline of the steel industry is used as an instrument in the second example, and the resulting estimates are substantially larger in magnitude than those obtained with first-differences or fixed-effects.

These findings suggest that studies using fixed-effects or first-differences models often understate the magnitude of the effect of interest, and that caution must be taken in the interpretation of these estimates. Researchers should consider estimating and reporting results from a larger set of panel data models. Comparing fixed-effects, first-differences, and long-differences models is a simple but informative check for misspecification. Furthermore, these findings in this paper indicate that there is no easy fix in the absence of useful variation in the data. If the cross-sectional variation is tainted by omitted variable bias, and the researcher uses a panel data model to control for unobserved heterogeneity, then they should consider whether sufficient sustained variation in the independent variable remains to identify the effect of interest. If the remaining variation is very "noisy," the only fruitful strategy for obtaining a reliable estimate is likely to be the use of a viable external instrument. Therefore, researchers need to seek out data from situations in which persisting changes in the dependent variable occur.

References

Argys, Linda, Susan Averett and Daniel Rees. Forthcoming. "Welfare Generosity, Pregnancies and Abortions Among Unmarried Recipients." *Journal of Population Economics*.

Baker, Michael, Dwayne Benjamin, and Shuchita Stanger. 1999. "The Highs and Lows of the Minimum Wage Effect: A Time-Series Cross-Section Study of the Canadian Law," *Journal of Labor Economics*, 17(2):318-50.

Bartik, Timothy J. and Randall W. Eberts. 1999. "Examining the Effect of Industry Trends and Structure on Welfare Caseloads." In Sheldon Danziger (ed.), *Welfare Reform and the Economy: What Will Happen When a Recession Comes?* Kalamazoo, MI: Upjohn Institute for Employment Research, pp. 119-57.

Black, Dan A., Kermit Daniel, and Seth G. Sanders. 1998. "The Impact of Economic Conditions On Participation in Disability Programs: Evidence from the Coal Boom and Bust." Working Paper #E-203-98, Center for Business and Economic Research, University of Kentucky.

Black, Dan A., Terra McKinnish, and Seth G. Sanders. 2000. "How the Availability of High-Wage Jobs to Low-skilled Men Affects AFDC Expenditures: Evidence from Shocks to the Coal and Steel Economies." Unpublished manuscript, Carnegie Mellon University.

Blank, Rebecca. 1997. "What Causes Public Assistance Caseloads to Grow?" Working Paper #6343, National Bureau of Economic Research.

Blank, Rebecca, Christine George and Rebecca London. 1996. "State Abortion Rates: The Impact of Policies, Providers, Politics, Demography and Economic Development." *Journal of Health Economics*. 15:513-54.

Bound, John and Alan B. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*. 9:1-24.

Bound, John, David A. Jaeger and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*. 90: 443-50.

Card, David and Alan Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United State." *Journal of Political Economy*. 100:1-40.

Clark, George and Robert Strauss. 1988. "Children as Income-Producing Assets: The Case of Teen Illegitimacy and Government Transfers." *Southern Economic Journal*. 64:827-56.

Dee, Thomas S. 1999. "State Alcohol Policies, Teen Drinking and Traffic Fatalities." *Journal of Public Economics*. 72:289-315.

Dugan, Laura, Daniel Nagin and Richard Rosenfeld. 1999. "Explaining the Decline in Intimate Partner Homicide: The Effects of Changing Domesticity, Women's Status, and Domestic Violence Resources." *Homicide Studies*. 3:187-214.

Duncan, Greg J. and Saul D. Hoffman. 1990. "Welfare Benefits, Economic Opportunities, and Out-of-Wedlock Births Among Black Teenage Girls." *Demography*. 36: 519-35.

Ellwood, David and Mary Jo Bane. 1985. "The Impact of AFDC on Family Structure and Living Arrangements." *Research in Labor Economics*. 7:137-98.

Figlio, David and James Ziliak. 1999. "Welfare Reform, the Business Cycle and the Decline in AFDC Caseloads" in *Welfare Reform and the Economy: What Will Happen When a Recession Comes?* ed. Sheldon Danziger. Kalamazoo, MI: Upjohn Institute for Employment Research, 19-48.

Fitzgerald, John. 1995. "Local Labor Markets and Local Area Effects on Welfare Duration." *Journal of Applied Policy and Management*. 14:43-67.

Fitzgerald, John. 1994. "A Hazard Model for Welfare Durations with Unobserved Location Specific Effects," unpublished paper, Department of Economics. Brunswich, ME: Bowdoin College.

Freeman, Richard B. 1984. "Longitudinal Analyses of the Effects of Trade Unions." Journal of Labor Economics. 2(1):1-26.

Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton, NJ: Princeton University Press.

Green, William H. 1993. *Econometric Analysis, 2nd Ed.* New York: Macmillan Publishing Company.

Griliches, Zvi and Jerry A. Hausman. 1986. "Errors in Variables in Panel Data." *Journal of Econometrics*. 31:93-118.

Hamermesh, Daniel S. Forthcoming. "The Art of Labormetrics." in *Handbook of Econometrics Vol 5* eds James Heckman and Edward Leamer.

Hoffman, Saul and E. Michael Foster. Forthcoming. "AFDC Benefits and Non-Marital Births to Young Women." *Journal of Human Resources*.

Hoynes, Hillary. Forthcoming. "Local Labor Markets and Welfare Spells: Do Demand Conditions Matter?" *Review of Economics and Statistics*.

________. 1997. "Does Welfare Play Any Role in Female Headship Decisions?" *Journal of Public Economics*. 65:89-117.

Hsiao, Cheng.1986. *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Jackson, Catherine and Jacob Klerman. 1994. "Welfare, Abortion and Teenage Fertility." Unpublished manuscript, RAND Corporation.

Janowitz, Barbara. 1976. "The Impact of AFDC on Illegitimate Birth Rates." *Journal of Marriage and The Family*. 38:485-94.

Johnston, Jack and John DiNardo. 1997. *Econometric Methods, 4th Ed*. New York: McGraw-Hill.

Klerman, Jacob and Steven Haider. 2000. "A Stock-Flow Analysis of the Welfare Caseload: Insights from California Economic Conditions," mimeo. Santa Monica, CA: RAND.

Levine, Phillip, Amy Trainor and David Zimmerman. 1996. "The Effect of Medicaid Abortion Funding Restrictions on Abortion, Pregnancies and Births." *Journal of Health Economics*. 15:555-78.

Levitt, Steven D. 1996. "The Effect of Prison Populations Size on Crime Rates: Evidence from Prison Overcrowding Litigation." *Quarterly Journal of Economics*. 111:319-51.

Matthews, A., D. Ribar and W. Wilhelm. 1997. "The Effects of Economic Conditions and to Reproductive Health Services on State Abortion and Birth Rates." *Family Planning Perspectives*. 29:52-60 .

McKinnish, Terra. 1999. "Model Sensitivity in Research on Welfare and Fertility." unpublished PhD dissertation, Carnegie Mellon University.

Miller, Robert and Seth Sanders. 1997. "Human Capital Development and Welfare Participation." *Carnegie-Rochester Conference Series on Public Policy,* 46: 1-47.

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

Moffitt, Robert. 1992. "Incentive Effects of the U.S. Welfare System." *Journal of Economic Literature*. 30:1-61.

________. 1994. "Welfare Effects on Female Headship with Area Effects." *Journal of Human Resources*. 29:621-636.

Mueser, Peter, Julie Hotchkiss, Christopher King, Phillip Rokicki and David Stevens. 1999. "The Welfare Caseload, Economic Growth and Welfare-to-Work Policies." mimeo, University of Missouri-Columbia.

Murray, Charles. 1984. *Losing Ground.* NY: Basic Books.

Murray, Sheila E., William N. Evans, and Robert M. Schwab. 1995. "Money Matters After All: Evidence from Panel Data on the Effects of School Resources." Unpublished manuscript, University of Kentucky.

Ribar, David and Mark Wilhelm. 1999. "The Demand for Welfare Generosity." *Review of Economics and Statistics*. 81(1):96-108.

Rosenzweig, Mark. 1999. "Welfare, Marital Prospects and Non-marital Childbearing." *Journal of Political Economy*. 107(6):S3-S32.

Sanders, Seth. 1993. "Preliminary Evidence on Human Capital Production and Welfare Participation." Chapter 3 of unpublished PhD dissertation, University of Chicago.

Schultz, T. Paul. 1994. "Marital Status and Fertility in the United States." *Journal of Human Resources*. 29:637-69.

Singh, Susheela. 1986. "Adolescent Pregnancy in the United States: An Interstate Analysis." *Family Planning Perspectives*. 18:210-17.

Wallace, Geoffrey and Rebecca Blank. 1999. "What Goes Up Must Come Down? Explaining Recent Changes in Public Assistance Caseloads" in *Welfare Reform and the Economy: What Will Happen When a Recession Comes?* ed. Sheldon Danziger. Kalamazoo, MI: Upjohn Institute for Employment Research, 49-90.

Wilson, William Julius. 1996. *When Work Disappears: The World of the New Urban Poor*. New York: Alfred A. Knopf.

Table 1: Cross-Sectional Variation in AFDC Benefits and Teen Birth Rates, 1980

State	AFDC Benefits to a Family of Four	Birth Rate, White Women Age 15-19
Arkansas	188	.068
Mississippi	120	.060
Connecticut	553	.027
Wisconsin	529	.038
1980 Cross-State Correlation = -.5533		

Table 2: Relationship Between AFDC Benefit Levels and Births to Young White Women, 1973-92, Summary of Different Estimation Methods

Estimation Method	Age 15-19	Age 20-24	Age 25-29	N
Pooled Cross-Section	-.2628*** (.0263)	.1059*** (.0284)	.2352*** (.0194)	1020
State Fixed Effects	.1226*** (.0288)	.1231*** (.0224)	.1359*** (.0214)	1020
First Differences	.0148 (.0329)	.0170 (.0230)	.0297 (.0216)	969
3-year Long Differences	.0496 (.0502)	.0736 (.0511)	.0672+ (.0347)	867
5-year Long Differences	.0654 (.0554)	.0943 (.0592)	.1030* (.0452)	765

Notes: Dependent variable is the logarithm of state birth rate for specified demographic group. Table reports coefficient on logarithm of state AFDC benefit level. Per capita income and year effects are included. Huber-White standard errors reported in parentheses. Data includes all 50 states and the District of Columbia. +p-value<.1,*<.05,**<.01, ***<.001.

Table 3: Using Lagged Values of AFDC Benefits as Internal Instruments:
Strong vs Weak Instruments

	White 15-19	White 15-19 F>10	White 20-24	White 20-24 F>10
Instrument X_{it}-X_{it-1} With:				
X_{it-2} [1.00]	2.44 (2.82)		-.349 (.823)	
X_{it-3} [1.40]	1.97 (2.03)		-.272 (.645)	
X_{it-4} [1.97]	2.28 (2.54)		-.272 (.655)	
X_{it-2}-X_{it-3} [0.42]	-.428 (.922)		.120 (.440)	
X_{it-3}-X_{it-4} [0.46]	-.150 (.780)		-.273 (.995)	
X_{it-2}-X_{it-4} [0.02]	1.39 (11.97)		1.48 (12.41)	
Instrument X_{it}-X_{it-2}With:				
X_{it-3} [2.45]	-.265 (.425)		.775 (.470)	
X_{it-4} [2.53]	-.255 (.421)		.776+ (.445)	
X_{it-3}-X_{it-4} [0.04]	-.921 (4.94)		.715 (4.40)	
X_{it-1}-X_{it-3} [48.97]	.053 (.063)	.053 (.063)	.137* (.064)	.136* (.064)
X_{it-1}-X_{it-4} [44.77]	.067 (.084)	.067 (.084)	.129* (.060)	.129* (.060)
Instrument X_{it}-X_{it-3}With:				
X_{it-2} [0.32]	-1.18 (2.69)		2.49 (5.00)	
X_{it-4} [4.19]	-2.76 (.395)		.762 (.498)	
X_{it-1}-X_{it-4} [163.7]	.057 (.053)	.057 (.053)	.118* (.056)	.118* (.056)
X_{it-1}-X_{it-2} [244.2]	.056 (.062)	.056 (.062)	.116+ (.062)	.116+ (.062)
X_{it-2}-X_{it-4} [42.21]	.057 (.077)	.057 (.077)	.121+ (.061)	.121+ (.061)

Notes: First-stage partial F-statistic in square brackets. Huber-White standard errors reported in parentheses. +p-value<.1,*<.05,**<.01, ***<.001.

Table 4: The Relationship Between County-level Earnings and AFDC Expenditures, 1969-93, Summary of Different Estimation Methods

Estimation Method	Coefficient on Earnings	N
Pooled Cross-Section	-.251*** (014)	69,487
Fixed Effects	-.069*** (.010)	69,487
First Differences	-.039*** (.006)	66,119
3-Year Long Differences	-.150*** (.011)	59,998
5-Year Long Differences	-.203*** (.014)	54,008
7-Year Long Differences	-.232*** (.016)	48,034
20-Year Long Differences	-.309*** (.037)	10,762

Notes: Dependent variable is the logarithm of county AFDC expenditures. Table reports the coefficient on the logarithm of total county earnings. All regressions include controls for county population, the change in county population and year effects. Huber-White standard errors are reported in parentheses. There are 3,182 counties in the U.S. The AFDC expenditures variable is suppressed in the sample at a rate of 8.75 %. ***p-value<.001.

Table 5: The Relationship Between County-level Earnings on AFDC Expenditures, Steel States, 1970-87 – Summary of Different Estimation Methods.

Estimation Method	Coefficient on Earnings	N
Pooled Cross-Section	-.192* (.081)	11,052
Fixed Effects	-.584*** (.028)	11,052
First Differences	-.138*** (.030)	10,427
3-Year Long Differences	-.441*** (.043)	9,191
5-Year Long Differences	-.537*** (.053)	7,956
7-Year Long Differences	-.582*** (.061)	6,719
IV Model 1 [partial F=12.4]	-.723*** (.176)	10,427
IV Model 2 [partial F=34.8]	-.809*** (0.208)	10,427

Notes: Dependent variable is the logarithm of county AFDC expenditures. Table reports the coefficient on the logarithm of total county earnings. All regressions include controls for county population, the change in county population and year effects. 2SLS Model 1 instrument is the fraction of men employed in primary metals in 1969 interacted with a dummy for 1970-80 and individual dummies for each year 1981-87. 2SLS Model 2 instrument is the interaction of fraction of men employed in primary metals in 1969 and fraction of U.S. earnings from primary metals. Huber-White standard errors are reported in parentheses. There are 620 counties in the 8-state region. The suppression rate for AFDC expenditures in this sample is less than 1%. *p-value<.05,**<.01, ***<.001.

Table 6: The Relationship Between County-level Earnings and Unemployment Insurance Expenditures, 1969-93, Summary of Different Estimation Methods.

Estimation Method	Coefficient on Earnings	N
Fixed Effects	-.097*** (.009)	71,662
First Differences	-.232*** (.016)	68,074
3-Year Long Differences	-.222*** (.017)	61,816
5-Year Long Differences	-.151*** (.017)	55,852
7-Year Long Differences	-.115*** (.017)	49,721
20-Year Long Differences	-.031 (.034)	10,802

Notes: Dependent variable is the logarithm of county Unemployment Insurance expenditures. Table reports the coefficient on the logarithm of total county earnings. All regressions include controls for county population, the change in county population and year effects. Huber-White standard errors are reported in parentheses. There are 3,182 counties in the U.S. The UI expenditures variable is suppressed in the sample at a rate of 6%. ***p-value<.001.

Table 7: The Relationship Between County-level Earnings and Unemployment Insurance Expenditures, Steel States, 1970-87, Summary of Different Estimation Methods

Estimation Method	Coefficient on Earnings	N
Fixed-Effects	-.342*** (.031)	11,135
First Differences	-.548*** (.081)	10,508
3-Year Long Differences	-.551*** (.067)	9,275
5-Year Long Differences	-.346*** (.059)	8,040
7-Year Long Differences	-.258*** (.054)	6,802
IV Model 1 [partial F=12.4]	-.079 (.375)	10,508
IV Model 2 [partial F=34.8]	.453 (0.355)	10,508

Notes: Dependent variable is the logarithm of county Unemployment expenditures. Table reports the coefficient on the logarithm of total county earnings. All regressions include controls for county population, the change in county population and year effects. 2SLS Model 1 instrument is the fraction of men employed in primary metals in 1969 interacted with a dummy for 1970-80 and individual dummies for each year 1981-87. 2SLS Model 2 instrument is the interaction of fraction of men employed in primary metals in 1969 and fraction of U.S. earnings from primary metals. Huber-White standard errors are reported in parentheses. There are 620 counties in the 8-state region. The suppression rate for UI expenditures in this sample is less than 1%. *p-value<.05,**<.01, ***<.001.

Figure 1: Nominal AFDC Benefits, Family of Four, 1970-94

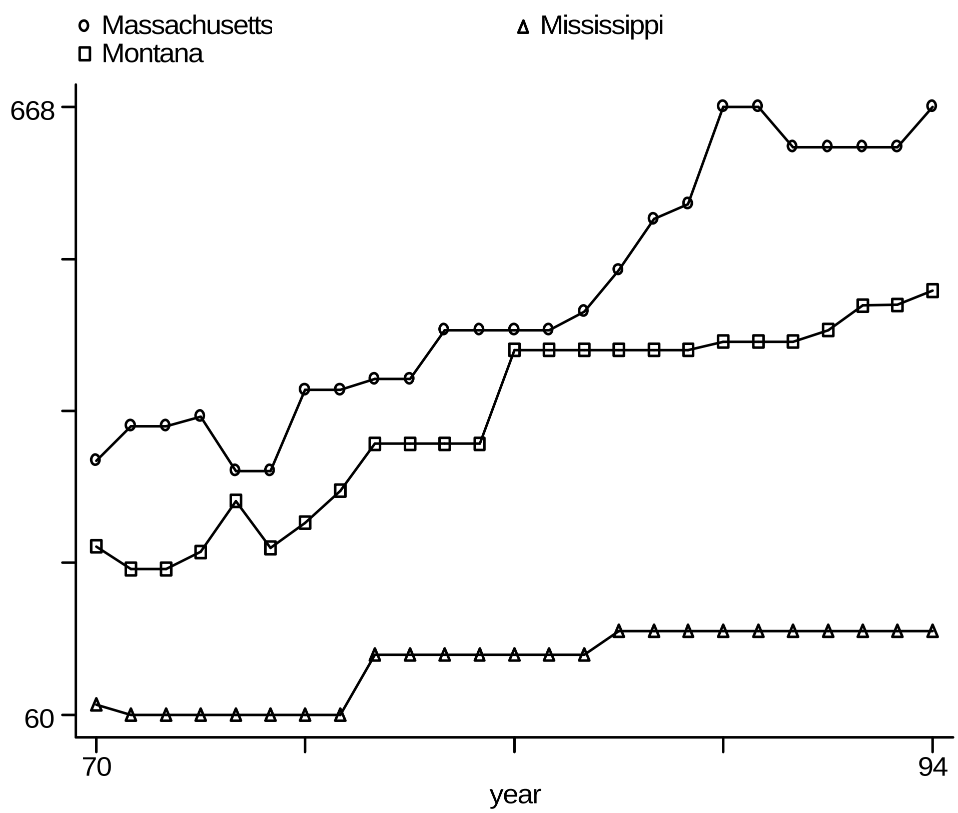

Figure 2:Fraction of Total Earnings from Primary Metal Manufacturing, 1969-94- U.S. and 8-State Steel Sample

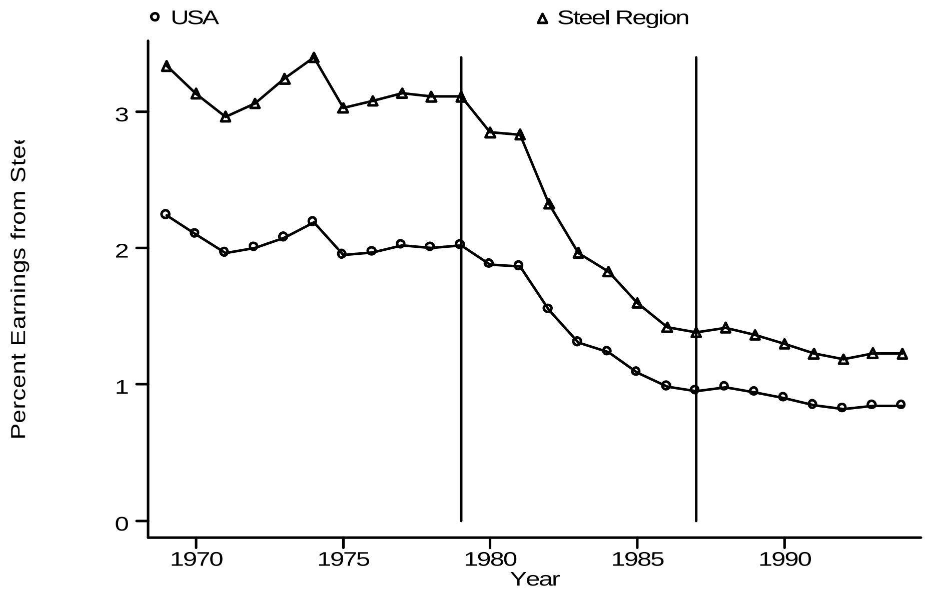

Figure 3: Fraction of Men Employed in Primary Metals in 1969