Using survey data to assess bias in the Consumer Price Index

Comparisons of self-reports with actual changes in families' financial status indicate that the CPI may measure such changes reliably

Alan B. Krueger
and
Aaron Siskind

Alan B. Krueger is Bendheim Professor of Economics and Public Affairs in the Department of Economics and Woodrow Wilson School, Princeton University, Princeton, New Jersey. Aaron Siskind is a graduate student in the Department of Economics, Princeton University.

In a clever and influential paper, William Nordhaus provides an estimate of the extent of bias in the Consumer Price Index (CPI) by comparing the net proportion of families that report an improvement in their financial situation with changes in real median income.[1] Specifically, he bases his analysis on time-series data collected by the University of Michigan's Institute for Social Research (ISR) in its Survey of Consumers. Among other things, this survey, which is used to measure consumer confidence, asks respondents whether their families' financial situation improved or worsened in the past year.

Nordhaus reasons that if real median household income rises in a particular year, more respondents should report becoming better off than worse off, and if real median income falls, more respondents should report becoming worse off than better off. In this view, constant real median income should be associated with an equal number of families reporting financial gains and losses. Nordhaus estimates the implied bias in the CPI by determining the growth rate of real median income that is associated with an equal number of families reporting an improvement, compared with a decline, in their financial situation. His point estimate suggests that the CPI is biased upwards by 1.5 percentage points. This is a novel approach to deriving an independent estimate of any bias in the CPI.

Nordhaus' method makes sense if the entire distribution of income moves with the median. But if the distribution of income changes in ways that are not related to the median, his approach could understate or overstate the bias in the CPI.

The following hypothetical example illustrates this point. Suppose the income distribution consists of five families that are ranked in order of their income in the base year. Suppose further that the family with the lowest income experiences a decline in (correctly measured) real income, the next two families experience no change in real income, and the top two families experience real-income growth. In this scenario—which might roughly mirror the U.S. income distribution over the last two decades—the median income is unchanged, so Nordhaus' reliance on the median would imply an equal number of families with real income gains and losses.[2]

However, 40 percent of families would have experienced a gain in real income and 20 percent would have experienced a decline, so, on net, 20 percent more families would report that their financial situation improved than worsened if the ISR survey question elicited accurate responses. In that case, a constant median real income would be correctly associated with an increase in the net proportion of families that reported being better off, not equal proportions better off and worse off. Thus, in this example, using the median family income to predict the net fraction of financial gainers would lead one to conclude that the CPI was biased upward, even if it in fact is unbiased. More generally, when the shape of the income distribution changes, the change in the median is not a sufficient statistic for determining the net number of

families that experienced income gains or losses.

Another potential problem concerns life-cycle income effects. The Census Bureau's estimate of household income is from the March Current Population Survey (CPS), which has a rotation-group design that should reflect the experience of repeated cross sections of households. In contrast, the ISR question asks respondents to reflect on their experience in the past year, so it is inherently longitudinal. If income rises over the life cycle for most families, the Census median income figure will understate the longitudinal growth in income, which would lead Nordhaus' method to overstate the bias in the CPI.

In this article, Nordhaus' analysis is extended. Most importantly, the Panel Study of Income Dynamics (PSID) data set is used to calculate the actual fraction of families that experienced measured increases and decreases in real income each year from 1968 to 1991. With longitudinal data, a variable can be calculated that, in principle, is more closely related to the ISR survey data on self-reported changes in financial well-being. If income is deflated properly, a regression of the net fraction of families that self-reported becoming financially better off on the actual fraction, as estimated from the PSID, should yield an intercept of zero and a slope of unity. Moreover, alternative assumptions about the possible bias in the CPI can be used to deflate real income in the microdata and then tested to see which assumption yields results that are closest to the "no bias" benchmark of a zero intercept and unit slope.

Accordingly, the next section of the article replicates Nordhaus' findings with median-income data. Also explored are the sensitivity of his results to using different percentiles of the cross-sectional income distribution, to using median income derived from the PSID, and to using an alternative measure of the CPI. This analysis finds that Nordhaus' results can be replicated and are generally robust. The section that follows presents new estimates based on the actual fraction of families whose income increased or decreased. Perhaps surprisingly, these results indicate that using the CPI to deflate family income may in fact provide an unbiased estimate of the fraction of families that report a net improvement in their financial situation. The article concludes with a discussion of the implications of the findings.

Replication of Nordhaus' analysis

Each month, the ISR's Survey of Consumers contains the following question:

> We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are _better off_ or _worse off_ financially than you were _a year ago_?

> Better Now   Same   Worse   Don't Know

In August 1997, for example, 45 percent of respondents reported that they were financially better off now, 31 percent reported that they were the same, and 24 percent reported they were worse off.

Following Nordhaus, the analysis presented here subtracts the percentage of families that report a worsening in their financial situation from the percentage that report an improvement and creates an annual series by averaging each 12 calendar months of data.[3] The resulting figures, which are reported in table A1 in the appendix, are henceforth referred to as *the net percentage of families whose financial situation improved.* Nordhaus regresses the net percentage of families whose financial situation improved on the percent change in median real household income, using the CPI-U to deflate income. His regression model is

$$(1) \qquad Y = a + bX,$$

where Y is the percentage of families that report an improvement in their financial situation, minus the percentage that report a worsening, and X is the percent change in median household income from the CPS, deflated by the CPI-U. The ratio $-a/b$ is an estimate of the percent change in measured real income that is associated with an equal number of families reporting an improvement and a worsening in their financial well-being, which Nordhaus interprets as an

| Table 1.	Regression of percentage of families reporting that they were better off minus percentage reporting that they were worse off on percent change in median real income, 1968–91 or 1968–94			
	Explanatory variable[1]			
Statistical quantity	**CPS median household income deflated by—**			**Median family income from Panel Study of Income Dynamics, deflated by CPI-U-X1**
	CPI-U[2]	**CPI-U-X1[3]**	**CPI-U-X1**	
Sample period	1968–94	1968–94	1968–91	1968–91
Intercept (*a*)	4.87	3.87	3.60	4.25
	(1.35)	(1.54)	(1.72)	(2.03)
Slope (*b*)	3.36	3.31	3.30	2.20
	(.54)	(.66)	(.70)	(.72)
a/b	1.45	1.17	1.09	1.93
	(.47)	(.56)	(.52)	(1.20)
R-squared	.61	.50	.51	.30
Average of dependent variable	5.20	5.20	5.30	5.30
Average of independent variable ..	.10	.40	.52	.48

[1] Standard errors are in parentheses.
[2] CPI-U = Consumer Price Index for All Urban Consumers.
[3] CPI-U-X1 = Experimental Consumer Price Index for All Urban Consumers.
SOURCE: Data from Survey of Consumers, Institute for Social Research, University of Michigan.

Chart 1.   Scatter diagram of Nordhaus' data
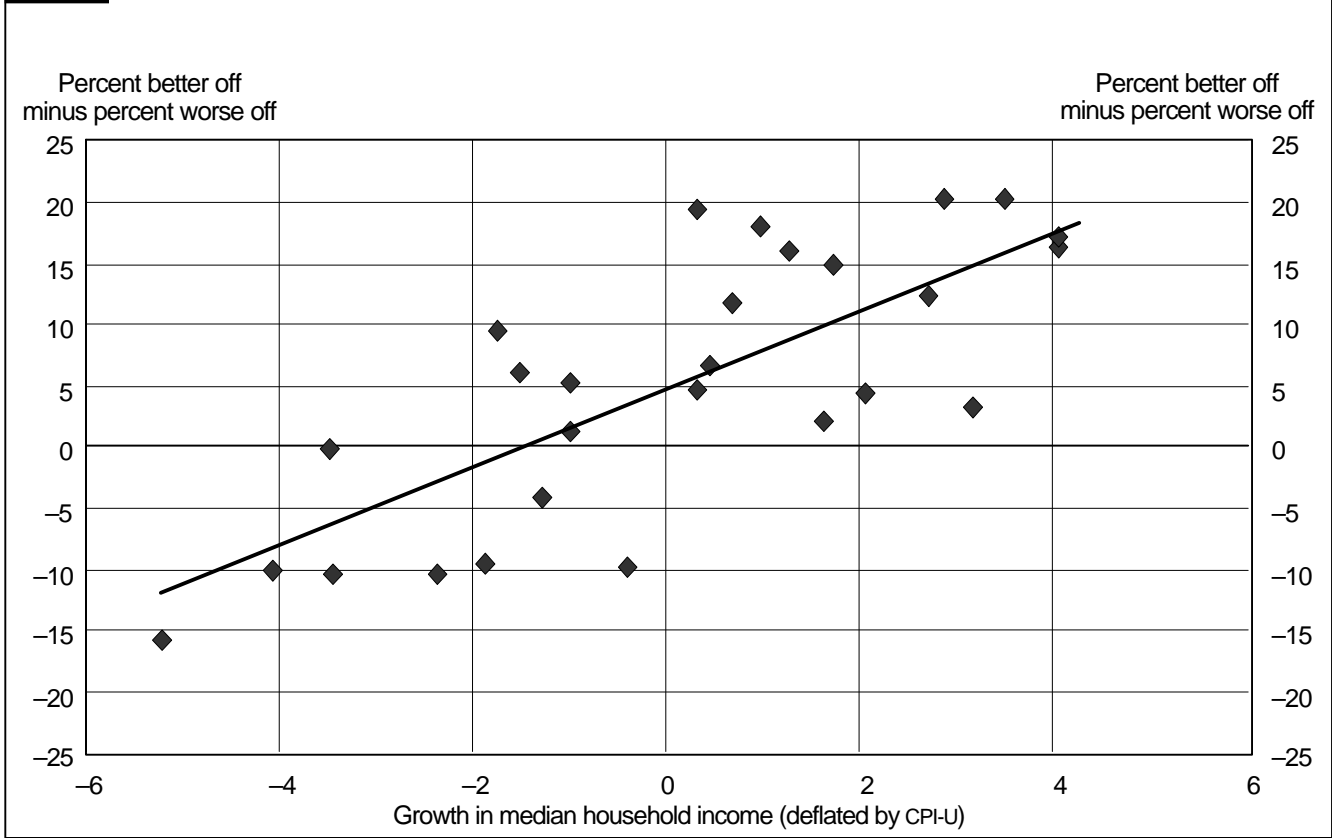
estimate of the bias in the CPI.

The column headed "CPI-U" in table 1 replicates Nordhaus' estimates: the implied bias in the CPI is about 1.5 percentage points, according to his specifications. A scatter diagram of the relationship between the two variables, *Y* and *X*, is provided in chart 1, along with the fitted ordinary least squares regression line. The *X*-intercept of the regression line corresponds to $-a/b$. Standard errors were calculated for a/b using the "delta method."

The Bureau of Labor Statistics has introduced several changes into the official CPI in past years. Although the Bureau does not retroactively change the official CPI, it has produced the CPI-U-X1, which adjusts the historical data for subsequent changes in the measurement of housing prices and therefore more closely reflects current procedures used in calculating the CPI-U. Nordhaus deflates income growth by the CPI-U, which is appropriate for estimating the bias in the historical data. Deflating by the CPI-U-X1 probably provides a better guide for the bias in the present-day CPI, however. Consequently, the first of the two columns headed "CPI-U-X1" in table 1 presents a reestimate of the same model as in the previous column, but now using the CPI-U-X1 to deflate median income growth. These results yield a smaller, but still substantial, 1.2-percentage-point-per-year bias in the CPI.

As of this writing, family income data from the PSID are available on a consistent basis only from 1967 to 1991, so the analysis that follows is restricted to that period. To examine whether the narrower period substantively affects the estimates, the second of the two columns headed "CPI-U-X1" reports a reestimate of equation (1) with CPS income data for 1968–91. This minor change hardly affects the results: the implied annual bias in the CPI is about 1.1 percentage point.

For comparison, the last column of table 1 shows results in which median real family income each year, calculated from the PSID, is the *X* variable (again using the CPI-U-X1 to deflate nominal income). The CPS and PSID median-income series are displayed in chart 2. The two median real-income measures have a correlation of 0.82 in levels, and the annual percent changes in these two measures have a correlation of 0.80. The last column of table 1 reports estimates of equation (1) using the percent change in real median family income from the PSID as the explanatory variable. The results indicate an even larger implied bias in the CPI: nearly 2 percentage points per year. Thus, the PSID and CPS median-income figures yield broadly similar results, with the PSID median-income data suggesting an even greater bias than the Census data.

Table 2 reestimates equation (1) using as the explanatory variable the inflation-adjusted percent change in selected in-

come quintile cutoffs and in the median.[4] If the shape of the log-income distribution remained constant over time, these estimates would all provide the same estimate of bias. The magnitude of the implied bias (*a/b*) varies monotonically with the income percentile that is used as the explanatory variable, with a greater bias estimated for lower percentiles of the income distribution. For example, if real-income growth for the bottom 20th percentile of the distribution is used, the implied overstatement in the CPI is 1.69 points per year; if the 80th percentile is used, the implied bias is 0.63 point. These findings are not surprising, in view of the well-documented changes in the income distribution over the last two decades, but they do suggest that changes in the income distribution could present a problem for interpreting Nordhaus' regression based on the change in median income.

Two final alternative specifications were examined to probe the robustness of Nordhaus' results. First, a multiple regression was estimated in which the net percentage of families that reported an improvement in their financial situation was the dependent variable and the percent change in *nominal* median household income and the percent change in the CPI-U-X1 were the explanatory variables. Nordhaus' bivariate specification essentially imposes the restriction that the coefficients of these two explanatory variables are equal, but opposite in sign. The analysis presented here rejects this restriction at the 0.001 level, although the percent change in nominal income does have a positive coefficient (1.90) and the percent change in the CPI does have a negative coefficient (–4.42). Second, because the model presented by equation (1) is primarily descriptive, the *X*-intercept can be estimated by regressing *X* on *Y*. In other words, the intercept from a regression of the percent change in real median income on the net percentage of families that report an improvement in their financial situation provides an estimate of the bias in the CPI (that is, a direct estimate of –*a/b*). When

such a reverse-regression model is estimated with the data from the median decile of table 2, the intercept is –0.37 (with a standard error of 0.36), indicating a smaller overstatement of the CPI than that produced by the earlier regressions. Both of these extensions tend to weaken Nordhaus' findings, although they do not overturn his central conclusion that the CPI overstates inflation.

Comparing changes in well-being

In principle, the net percentage of families reporting an improvement in their financial situation, as measured by the ISR survey (denoted *Y*), should equal the percentage of families that actually experienced an increase in real income, minus the percentage that experienced a decline in real income. (Let *Z* denote the latter variable.) If the regression

$$(2) \qquad Y = a' + b'Z,$$

is estimated, and if income is properly deflated, one would expect to find that $a' = 0$ and $b' = 1$. A joint test of these coefficient restrictions provides a more robust test of bias in the CPI than does Nordhaus' regression of *Y* on median real-income growth (*X*), because the restrictions should hold even if the shape of the income distribution changes over time. In addition, both measures reflect the life-cycle income profile.

To carry out this test, the PSID was used to calculate the percentage of families that experienced a rise or fall in income each year from 1968 to 1991. Each year, the sample consisted of all families with positive income in year *t* or year *t* –1.[5] First, income was deflated by the CPI-U-X1, and then the percentages of families that experienced a rise and a decline in income were calculated. The variable *Z* is the difference between these two percentages and is presented in table A1 in

Table 2.	Regression of percentage of families reporting that they were better off minus percentage reporting that they were worse off on percent change in selected CPS deciles of household income, deflated by CPI-U-X1, 1968–95				
Statistical quantity	**Decile**[1]				
	Second	**Fourth**	**Median**	**Sixth**	**Eighth**
Sample period	1968–95	1968–95	1968–95	1968–95	1968–95
Intercept (*a*)	4.23	4.41	3.91	3.31	2.14
	(1.70)	(1.54)	(1.50)	(1.66)	(1.80)
Slope (*b*)	2.50	3.12	3.33	3.20	3.41
	(.64)	(.64)	(.63)	(.72)	(.80)
a/b	1.69	1.41	1.18	1.03	.63
	(.87)	(.61)	(.46)	(.63)	(.61)
R-squared	.37	.48	.52	.43	.41
Average of dependent variable	5.52	5.52	5.52	5.52	5.52
Average of independent variable	.51	.36	.48	.69	.99

[1] Standard errors are in parentheses. Sample size is 28.

SOURCE: Data from Survey of Consumers, Institute for Social Research, University of Michigan.

Table 3.	Regression of net percentage of families reporting that they were better off on net percentage of families reporting higher real income[1]				
CPI adjustment factor	Constant	Slope	*R*-squared	*P*-value *for* joint *F*-test of constant = 0 and slope = 1	Mean of net proportion of families reporting increasing real income on Panel Study of Income Dynamics
2.0	8.15 (2.10)	0.91 (.26)	0.35	0.001	−3.13
1.5	6.25 (1.95)	.91 (.27)	.35	.012	−1.04
1.0	4.22 (1.97)	.90 (.27)	.34	.125	1.20
.5	2.20 (2.15)	.91 (.27)	.34	.595	3.40
.0	**−.05 (2.49)**	**.94 (.27)**	**.35**	**.954**	**5.70**
−.5	−2.48 (2.89)	.97 (.27)	.37	.384	7.99
−1.0	−5.08 (3.32)	1.02 (.27)	.39	.049	10.21
−1.5	−7.51 (3.81)	1.04 (.27)	.40	.004	12.29
−2.0	−10.12 (4.30)	1.07 (.27)	.42	.000	14.45

[1] Real income is nominal income deflated by the CPI-U-X1.

NOTE: The mean of the dependent variable is 5.30. The sample consists of annual observations from 1968 to 1991. Standard errors are in parentheses.

the appendix.[6] These data provide the right-hand-side variable for the regression model represented by equation (2), and the ISR data on self-reported changes in financial well-being used earlier form the dependent variable. Regression results are reported in the middle row of table 3 (highlighted in boldface type), and the relationship is depicted in the scatter diagram of chart 3. Notably, the linear relationship appears to have an intercept close to 0 and a slope close to 1; a formal joint *F*-test of these two coefficient restrictions has a *p*-value of 0.954, which is far from rejection of the null hypothesis.

The process of estimating this equation was repeated several times, using different hypothetical assumptions about bias in the growth of the CPI to deflate the microdata from the PSID and calculate the aggregate time-series Z-variable. In particular, 0.5, 1, 1.5, or 2 percentage points were added to or subtracted from the CPI each year to see which implied constant bias best satisfied the coefficient restrictions. (See table 3.) Surprisingly, the assumption of no bias (that is, using the unadjusted CPI-U-X1) was furthest from rejecting the restrictions that $a' = 0$ and $b' = 1$. When 1.5 points were subtracted from the annual growth in the CPI, however, the *p*-value for a joint test of $a' = 0$ and $b' = 1$ was 0.004, rejecting the no-bias

restrictions. Interestingly, over the range of hypothetical biases reported in table 3, the slope coefficient is never very different from 1.0, while the intercept varies considerably, depending on the assumed bias. Consequently, the intercept term is responsible for the rejection of large positive and large negative CPI bias assumptions.

Because the fitted regression line passes through the mean of the dependent and independent variables, and because we expect an intercept of 0 and a slope of 1, another indication that the unadjusted CPI squares quite well with the data 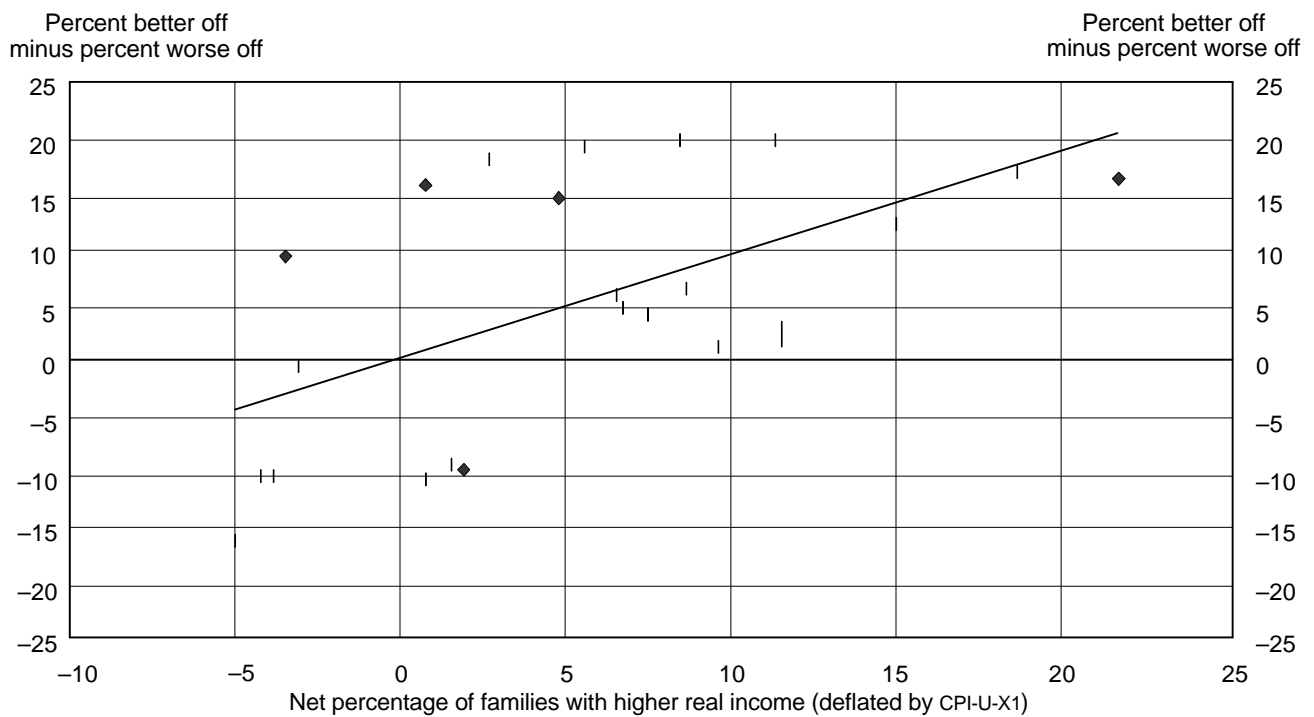is the fact that, over the 24 years studied, the average difference between the percentages of families with annual income growth and annual income declines is 5.7 percent in the PSID and 5.3 percent in the self-reported ISR data. By contrast, if the bias in the CPI were 1.5 points per year, the PSID data would indicate that 12.3 percent of families experienced rising income in the average year. Because the slope coefficient is close to 1 when this variable is the explanatory variable, the intercept must be far from 0.
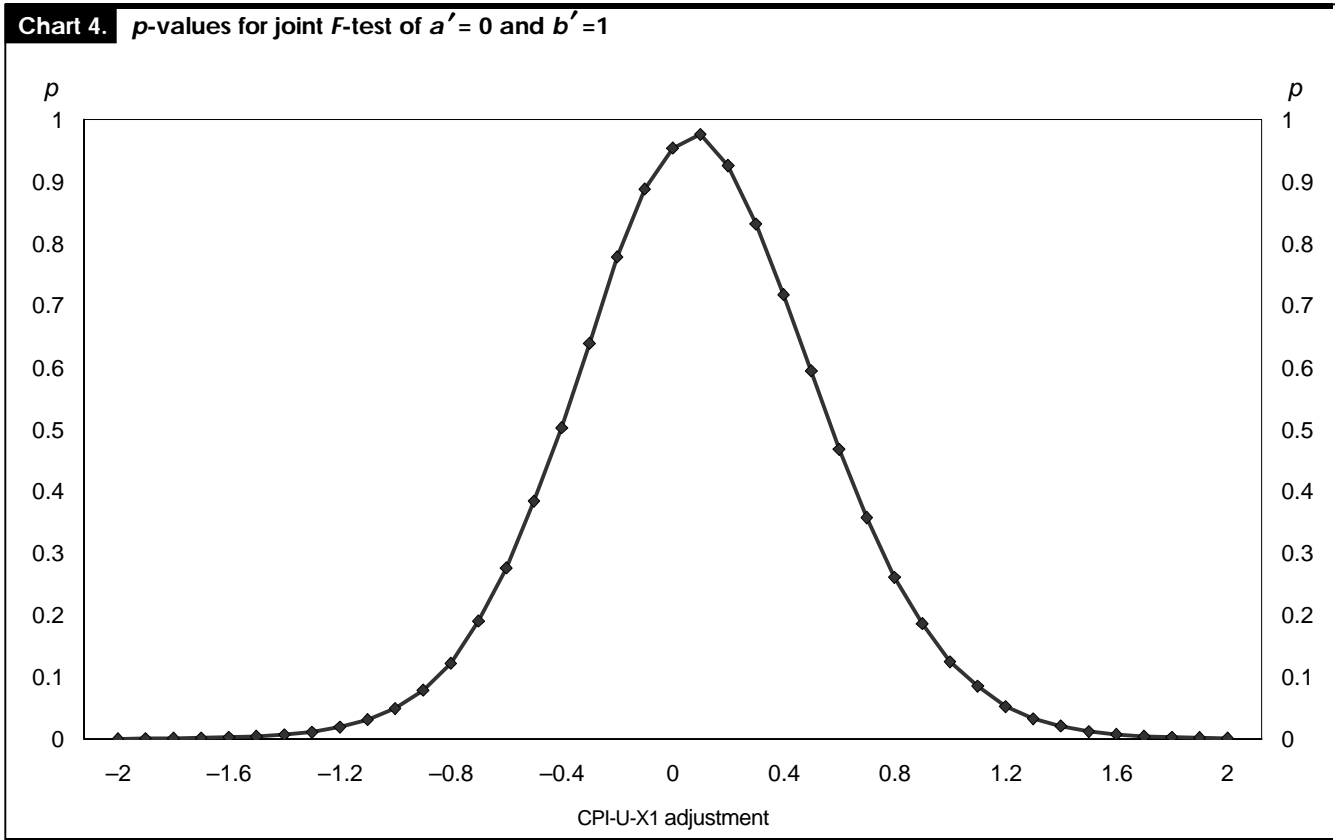
To examine more closely the value at which the no-bias restrictions best fit the data, a grid search was performed in which the assumed annual bias in the CPI was varied by ±0.1 between −2.0 and +2.0 percentage points. Specifically, the right-hand-side variable from the PSID microdata was recalculated for each of these adjusted CPI's, and equation (2) was then reestimated 40 times using each of these variables in turn as the explanatory variable. Chart 4 displays the *p*-value for the joint *F*-test of the restrictions $a' = 0$ and $b' = 1$ from each of these regressions.[7] The *F*-test is least likely to reject the restrictions if the CPI understates inflation by 0.1 percentage point per year, although the *p*-values are generally quite high if the CPI adjustment is within ±0.5 percentage point, and the regression model would have difficulty distinguishing among

Chart 2. Median real household income from Current Population survey (CPS) and median real family income from Panel Study of Income Dynamics (PSID), 1968–91

Thousands of dollars

Thousands of dollars

Chart 3. Scatter diagram of ISR and PSID measures of net percentages of families that are better off

Percent better off minus percent worse off

Percent better off minus percent worse off

Net percentage of families with higher real income (deflated by CPI-U-X1)

Chart 4. *p*-values for joint *F*-test of $a' = 0$ and $b' = 1$

biases in this range.

Three extensions of the analysis were carried out in order to check the robustness of the estimates based on the PSID. The first extension was motivated by the fact that about 30 percent of families in the ISR survey report that their financial situation is the same as in the previous year. The same analysis as in table 3 was performed, but now all families whose real income changed by less than ±10 percent were treated as having a constant financial situation. The 10-percent figure was chosen because it encompasses roughly 30 percent of the sample. If chart 4 is recalculated with these assumptions, the peak of the curve is at 0.3 percent, instead of 0.1 percent, suggesting a slightly greater understatement of price increases. The peak also becomes somewhat flatter if a 10-percent threshold for assigning a change in real income is assumed.

Second, to allow for possible money illusion (that is, the possibility that the public might misperceive nominal-income gains as leading to increased financial status, even if real income is constant), the PSID was used to create a new variable measuring the net percentage of families with *nominal* gains in income.[8] This measure was included along with the net percentage of families with gains in real income (deflated by the CPI). Interestingly, one cannot reject a unit slope for the measure of real-income gains in the ensuing regression. The net percentage of families with nominal-income gains, how-

ever, has a coefficient of –1.02 with a standard error of 0.21. These results suggest that the public does not interpret nominal-income gains as leading to increased financial status, conditional on real-income gains.

Last, the reverse regression of *Z* on *Y* was estimated; that is, the PSID measure of net percentage with real-income gains was regressed on the ISR survey measure. Because, apart from measurement errors and sampling errors, the two variables should be equal if the CPI is unbiased, one would still expect a zero intercept and unit slope from this reverse regression. However, when such a regression is estimated, the results are inconsistent with a zero intercept and unit slope. Moreover, if either a large negative bias or a large positive bias in the CPI is assumed, the zero intercept and unit slope are still strongly rejected. The reason for this rejection is that the time-series variability in the ISR measure is considerably greater than the variability in the PSID measure, even though the means of the two variables are similar when the CPI is used to deflate the PSID data. The lower dispersion in the PSID measure over time holds, regardless of whether the CPI minus 2 points or plus 2 points is used to deflate income. Why the PSID generates lower time-series variability in the net income changes is unclear, but this finding suggests that the PSID and ISR variables may be measuring different concepts. For example, swings in the self-perceived measure of financial status may reflect factors

other than just changes in real income.

UNLIKE MOST PAST RESEARCH, the results presented in this article suggest that changes in the CPI provide an accurate measure of changes in the cost of living.[9] Several caveats should be borne in mind, however, before concluding that the CPI provides an unbiased estimate of the cost of living. For one thing, the estimates presented here are imprecise, so a bias on the order of plus or minus 0.5 point per year could not be rejected at conventional significance levels. In addition, the estimates notwithstanding, there is a strong presumption that the CPI overstates the cost of living, if for no other reason than substitution bias. Much evidence also suggests that the CPI may inadequately adjust for changes in the quality of health care and other goods. However, measuring changes in the cost of living is complex and, in practice, requires strong assumptions that may not be met, such as the absence of consumer demand shifts.

For the preceding reasons, it is worth considering ways of gauging the plausibility of bias estimates in the CPI that do not rely heavily on imposed assumptions about economic behavior. Nordhaus provides an ingenious way to gauge bias in the CPI by comparing public opinion regarding financial well-being and measured changes in real median income.[10] The approach presented in the current article, which uses a more appropriate variable to measure changes in real income, appears to be an improvement over Nordhaus' pioneering effort; however, the results offer little evidence that the CPI overstates inflation.

Robert Pollak argues that, because professional conventions are required to calculate the CPI, its credibility depends critically on the public's perception that it is "not being manipulated as a policy instrument."[11] If this is the case, then using public-opinion data to gauge possible bias in the CPI may be quite valuable.

On the other hand, economists' reliance on revealed preference has a lot to be said for itself. One may question whether comparing measured income with public-opinion data such as data collected by the ISR survey provides a valid test of bias in the CPI, in both Nordhaus' approach and that presented in this article. In particular, perceived changes in financial well-being may be influenced by more factors than just changes in (correctly measured) annual real income. For example, changes in family size, expectations of future earnings or asset growth, expectations of indebtedness, college tuition costs, and home mortgage interest rates also may influence a person's perceived financial well-being. Furthermore, self-reported changes in financial status may be confounded by other, noneconomic factors that affect the national outlook, such as a military conflict or a successful mission to Mars. If these other possible influences on self-reported changes in financial well-being are correlated with changes in the proportion of families that experience real-income gains, then the simple model in equation (2) would yield a biased estimate of the extent to which the CPI under- or overstates changes in the cost of living. If, on the other hand, they amount to nothing more than additive "white noise" in the dependent variable, equation (2) will continue to yield unbiased estimates. □

Footnotes

[1] William Nordhaus, "Quality Changes in Price Indexes," *Journal of Economic Perspectives*, winter 1998, pp. 59–68.

[2] Between 1974 and 1994, the percent change in average household income for each quintile was –7 percent for the lowest quintile, –3 percent for the second quintile, 1 percent for the middle quintile, 9 percent for the fourth quintile, and 31 percent for the top quintile, using the CPI-U-X1 to deflate income. (See *Historical Income Tables for Households* (U.S. Department of Commerce, Bureau of the Census, various years), table H–3; also on the World Wide Web at <www.census.gov/hhes/income/histinc/h03.html>.)

[3] To be precise, a weighted average was calculated, where the weights were the sample size each month. (See appendix for details.)

[4] The income data are based on the CPS and reported in *Historical Income Tables for Households,* table H–1. Notice that because an additional year of data is available, the equations in table 2 were estimated with a longer sample of data than that used in table 1.

[5] The PSID income data pertain to the previous calendar year, so the data collected in the 1992 survey wave would apply to 1991. (See appendix for details of the calculations.)

[6] The observations were weighted by the PSID composite family weights in year *t*, which, in principle, should make the sample representative of the population. Note that, because it is impossible to determine whether income rose or fell for families with top-coded income in consecutive years, these observations were deleted from the sample. However, because the top code in the PSID is relatively high most years, such families make up less than 0.10 percent of the sample. Also eliminated from the sample were a small number of families whose income was bottom coded in consecutive years; the bottom code for income in the PSID is $1 per year.

[7] This technique can be viewed as a two-step version of the Fieller method for calculating confidence intervals. (See E. C. Fieller, "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society*, vol. 16, no. 2, 1954, pp. 175–85.) In that method, a 95-percent confidence interval is calculated by finding all the possible values of the null hypothesis that are not rejected by the data at the 5-percent level. In the analysis presented in the current article, imposing alternative null hypotheses about the bias in the CPI requires that the variable Z be rederived from the PSID microdata and that the joint hypothesis test on the intercept and slope then be performed.

[8] See Eldar Shafir, Peter Diamond, and Amos Tversky, "Money Illusion," *Quarterly Journal of Economics*, May 1997, pp. 341–74, for recent evidence on money illusion.

[9] For discussions of possible biases in the CPI and relevant evidence, see the entire December 1993 issue of the *Monthly Labor Review*, as well as the following articles: Brent Moulton, "Bias in the Consumer Price Index: What Is the Evidence?" *Journal of Economic Perspectives*, Fall 1996, pp. 159–77; Michael Boskin, Ellen Dulberger, Robert Gordon, Zvi Griliches, and Dale Jorgenson, *Toward a More Accurate Measure of the Cost of Living,* Final Report to the Senate Finance Committee, Dec. 4, 1996; and Matthew Shapiro and David Wilcox, "Mismeasurement in the Consumer Price Index: An Evaluation," in Ben Bernanke and Julio Rotemberg, eds., *NBER Macroeconomics Annual, 1996* (Cambridge, MA, MIT Press, 1996), pp. 93–142. For evidence on new goods, see especially Timothy F. Bresnahan and Robert J. Gordon, eds., *The Economics of New Goods* (Chicago, University of Chicago Press, 1997).

[10] Nordhaus, "Quality Changes."

[11] Robert Pollak, "The Consumer Price Index: A Research Agenda and Two Proposals," *Journal of Economic Perspectives,* winter 1998, pp. 69–78; quote from p. 75.

APPENDIX: About the data

This article uses data from the University of Michigan's Institute for Social Research (ISR), the Census Bureau, and the Panel Study of Income Dynamics (PSID).

The net percentage of families whose financial situation improved, which is the dependent variable in the regressions presented in the article, is constructed from data from the ISR's Survey of Consumers. This ongoing survey asks a sample of roughly 500 families per month a series of questions about their views concerning various economic issues, such as their current and expected future financial situation and current and future business conditions. The specific survey question used in the article is "Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?" Each quarter, the ISR reports an index that equals the percentage of families responding that they were better off financially than they were a year ago, minus the percentage of families reporting themselves worse off, plus 100. The analysis in the article computes the yearly net percentage as the average of the four calendar quarters of the index, weighted by the number of observations per quarter, minus 100. (See table A1.)

In the Census Bureau data, median household income is obtained from table H-5 of the *Historical Income Tables for House-holds*; quintile income is obtained from table H-1.[1] The article also uses the CPI-U and CPI-U-X1, which is taken from table B-60 of the 1997 *Economic Report of the President*. In replicating Nordhaus' results, for example, nominal median household income from table H-5 is deflated by the CPI-U, and then the percentage change from year to year is computed.

Lastly, data from the PSID are used to compute the net percentage of families with higher real income from year to year, which is the Z-variable in equation 2. Because there is no way to directly track families from one year to the next, the analysis examines families with the same head in consecutive years.

The variables used from the PSID family data are the interview number, total family money income,[2] and family weight (the core weights for the sample, which excludes the new Latino sample for 1990–92, code v18943 in 1990). Also used are the 1968 identification number, individual number, interview number, status, and variables associated with the individual's relation to the head of the

Year	Net percentage of families better off, from ISR survey of consumers	Percentage change in median real income, from PSID	Estimated percentage better off minus percentage worse off, from PSID			Number of families
			CPI	CPI minus 1 percentage point	CPI plus 1 percentage point	
1968	16.42	5.92	21.67	25.32	16.72	4,130
1969	12.38	2.68	15.04	18.46	11.52	4,170
1970	5.98	−.19	6.57	11.18	2.44	4,388
1971	1.31	−.21	9.54	13.78	5.03	4,568
1972	17.11	4.83	18.63	22.37	14.62	4,784
1973	4.33	2.23	7.51	11.55	3.19	4,988
1974	−10.04	−4.50	−4.14	−.56	−8.27	5,213
1975	−10.26	−2.27	−3.86	.34	−9.34	5,358
1976	2.14	1.84	11.51	15.81	7.47	5,513
1977	6.63	3.28	8.58	14.64	3.00	5,671
1978	3.39	1.93	11.46	16.46	7.87	5,812
1979	−9.43	1.05	1.52	6.45	−3.62	5,965
1980	−15.74	−5.42	−4.96	−.64	−9.77	6,157
1981	−10.24	.57	.81	4.93	−3.63	6,253
1982	−9.75	−1.76	1.92	5.71	−2.27	6,321
1983	4.74	1.55	6.89	10.76	2.54	6,417
1984	20.24	3.00	11.43	16.34	7.06	6,449
1985	15.00	1.31	5.02	9.37	.24	6,557
1986	20.25	1.34	8.61	13.37	3.60	6,571
1987	18.06	.61	2.89	8.51	−1.88	6,613
1988	19.49	−.76	5.61	10.51	1.54	6,673
1989	16.00	−.26	.90	6.67	−3.78	6,702
1990	9.50	−5.11	−3.22	1.47	−8.06	6,827
1991	−.25	−.21	−3.01	2.23	−7.50	6,867
1992	−4.01	—	—	—	—	—
1993	5.26	—	—	—	—	—
1994	11.76	—	—	—	—	—
1995	14.25	—	—	—	—	—
1996	15.50	—	—	—	—	—
Mean (1968–91)	5.30	.48	5.70	10.21	1.20	....
Standard deviation (1968–91)	11.46	2.83	7.20	7.06	7.38	....

Table A1. **Key variables used in analysis, 1968–96**

NOTE: The CPI used to deflate income is the CPI-U-X1. Number of families pertains to preceding three columns. Dash indicates data not available.

family, all from the PSID individual data.

The 1968 identification number and individual number were concatenated to create unique individual identification numbers so that individuals could be tracked over time. Family and individual data were then merged year by year on the basis of the interview number, and only the family heads' observations were retained, using the status variable and the relation-to-head variables; an individual is a head of the family if relation to head = 1 for 1968, status = 1, and relation to head = 1 for 1969–82, and if status = 1 and relation to head = 10 for 1983–92. The result of these operations was a data set that includes total family money income, the family head's individual identification number, and the family weights for each year. With this data set, real total family money income is compared from one year to the next for each family, using the head's individual identification number to match families. These data are then used to compute the net percentage of families with higher real income in year t compared with year $t-1$, utilizing the CPI-U-x1 (plus or minus up to 2 percent) to deflate nominal income each year. Families are included whenever they are in the sample in 2 adjacent years.

Because of income censoring, some family heads are eliminated from the analysis. Total family money income is top coded at $99,999 from 1968–79 (using 1967–78 income), $999,999 in 1980 (using 1979 income), and $9,999,999 thereafter. Observations for which total family money income is above $99,999 in consecutive years from 1968 to 1979 and those for which income is top coded in 1979 and above $99,999 in 1980 in 1979 dollars are eliminated,[3] as no determination can be made as to whether those families' real income rose or fell over the years in question. (For example, if a family's income is top coded in 1976 and 1977, but falls below $99,999 in 1978, that observation is eliminated from the 1976–77 comparison, but is counted as having lower real income in the 1977–78 comparison.) Total family money income is bottom coded at $1 in all years, so observations for which total family money income is reported at that level in consecutive years also are eliminated, because no determination can be made as to whether those families' real income rose or fell.

Also excluded are observations in which the head of the family changed from year to year. If, for example, the head of a certain family is different in 1976 than in 1977, but remains the same in 1978, that family is eliminated from the 1976–77 comparison, but is included in the 1977–78 comparison (because only the head of the family is being tracked).

Footnotes to the appendix

[1] *Historical Income Tables for Households* (U.S. Department of Commerce, Bureau of the Census, various years), tables H–5 and H–1.

[2] Total family money income (variable number v16144 in the 1988 PSID data set) is "the total of all members' earnings, transfers, and asset income from the prior calendar year." For 1988, this is the sum of variable numbers v14911 through v14918, v14922 through v14925, v14928, v14930, v14932, v14933, v14943, v14946, v14947, v14949, v14952, v14954, v14956, v14957, v14959, v14961, v14963, v14964, v14967, v14968, v14970, v14973, v14975 through v14977, v14979, v14981, v14983, v15061, v15066, v15071, v15076, v15081, and v15089 through v15099.

[3] Observations are comparisons of a family in year t with that same family in year $t-1$. Families need not be in the sample all years. Note that 1980 nominal income is deflated by the CPI-U-x1.