

Jack E. Triplett is the chief economist at the Bureau of Economic Analysis. Robert McCahill provided research assistance.

Measuring Consumption: The Post-1973 Slowdown and the Research Issues

Jack E. Triplett

What happened to U.S. economic growth after 1973? Before 1973, productivity, real compensation, and real per capita consumption all showed strong annual growth—in the 2.5 percent to 3.0 percent range (see Table 1). After 1973, growth diminished in all three measures. The post-1973 retardation in U.S. productivity growth—from 2.9 percent per year before 1973 to only 1.0 percent per year since 1973—has been a research topic for nearly 20 years. More recently, attention has turned to the reduction in the growth rate of real earnings that began about the same time.¹ The growth rate of U.S. real per capita consumption also abruptly slowed after 1973. In the decade and a half before 1973, real per capita consumption grew 3.0 percent per year. In the subsequent two decades, it grew only a little more than half as quickly—1.7 percent per year (see Table 1).

All three reductions in economic growth rates—productivity, real earnings, real per capita consumption—tell a superficially consistent story, up to a point. If the rate of productivity improvement fell, so must the growth of real earnings, absent changes in the functional distribution of income. And if the growth of real earnings slowed, so ultimately must the growth in real per capita consumption, even with the reduction in the savings rate that has also been observed.

However, data problems—defects in measurement—play a very prominent role in the research agendas on productivity and real earnings.² Here, I take a critical look at the data on real per capita consumption.

Real consumption is measured from the bottom up (i.e., components of consumption are estimated and deflated separately). If measurement errors exist, they must therefore be specific errors that affect particular components—either an error in the estimation of the current-dollar consumption expenditures for a particular component or an error in the individual price index used for deflating that component. Thus, if one believed that the overall growth rate of real consumption were too low because of measurement errors, those errors ought to show up differently in the growth rates of consumption sectors that are difficult to measure (e.g., high-technology electronic products or services). Conversely, if the measured overall level of real consumption were too low because difficult-to-measure sectors were understated, then the sectors that have fewer measurement problems (e.g., food) ought to show higher-than-expected growth (i.e., a lower-than-expected post-1973 slowdown). Accordingly, I examine components of real per capita consumption, not just the aggregate data.

I first review the slowdown in consumption growth rates and then turn to the measures of current-dollar consumption expenditures—the numerator of the real consumption ratio. Subsequently, I discuss the price statistics used for deflation, with particular attention to the report of the CPI Commission (1996).

THE POST-1973 SLOWDOWN IN PER CAPITA CONSUMPTION GROWTH

As already noted, the growth of real per capita personal consumption expenditures (PCE) slowed more than 40 percent

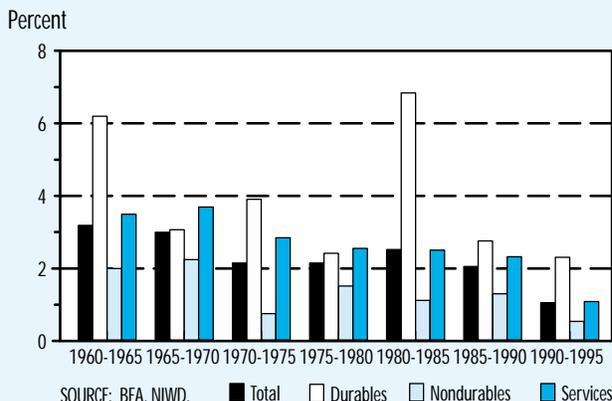
This article represents my personal views and is neither an official position of the Bureau of Economic Analysis (BEA) nor of the U.S. Department of Commerce. I am greatly indebted to the following individuals for discussions and comments on the article's content: B.K. Astrosic, Barry Bosworth, Charles Schultze, Gerald Donahoe, Zvi Griliches, Robert Parker, and especially Brent Moulton. Valuable advice and assistance on the data were provided by Clint McCully and Greg Key of BEA and Raphael Branch and Stephanie Shipp of the Bureau of Labor Statistics. The article benefitted from a seminar presentation at the Bank of England, for which I thank William Allen.

¹ Levy and Murnane (1992, p. 1,333) note, "Nineteen-hundred-seventy-three marked the end of rapid real earnings growth and the beginning of slower growth bordering on stagnation."

² Griliches (1994) discusses measurement issues for productivity. Consumer Price Index (CPI) Commission (1996) discusses possible errors in the CPI, which provides deflators for calculating real earnings growth and real consumption growth.

Figure 1

Growth in Per Capita Personal Consumption (PC)



capita consumption grew half as fast after 1973 as before. The categories of “other durables” (which contains computers and electronic equipment which have experienced rapid quality change) and “other” (i.e., nonhousing) services are usually thought to be relatively hard to measure; growth rates in these two components declined by 20 percent and 35 percent, respectively. Thus, the consumption slowdown is not concentrated in the hard-to-measure areas of consumption. In fact, it is substantially larger in the easier-to-measure components (and in motor vehicles, which show the greatest post-1973 slowdown).

These trend consumption growth rates are not timing accidents of the business cycle. Peak-to-peak business cycle growth rates are tabulated in Table 2: In *none* of the peak-to-peak periods since 1973 did per capita real consumption growth reach the peak-to-peak rates seen before 1973.⁴ The same statement is true for most of the consumption categories in Table 2: Peak-to-peak growth rates after 1973 do not reach their pre-1973 rates.

Rates of increase over decades tell the same story: Each of the two decades after 1973 falls short of the 1963-73 growth rate in total PCE, and also for every consumption category, including nonhousing services.

Data for shorter periods confirm the pervasiveness of the consumption slowdown. Figure 1 shows five-year periods ending in decades and half-decades. In no five-year period after 1973—even the peak growth years of 1980-85—has per capita real consumption ever reached the growth rates that prevailed in the 1960s. This intertemporal statistical regularity holds for total PCE and for every major PCE component—with the exception of durables (1980-85) and marginally for 1970-75. It is commonplace that the United States has become a services economy; however, even for the services category, growth rates for five-year periods after 1973 never again attained the rate of growth of services in the 1960s.

Figure 2 illustrates the same data, organized by five-year periods centered on 1973. This organization corresponds a bit better to business cycle movements than

Table 1

Average Annual Percentage Growth Rates

| | 1959-73 | 1973-95 |
|-------------------------------|---------|---------|
| Productivity output per hour* | 2.9 | 1.0 |
| Real hourly compensation* | 2.4 | 0.3 |
| Real per capita consumption | 3.0 | 1.7 |

*Nonfarm business sector.

SOURCES: Productivity and real hourly compensation figures from U.S. Bureau of Labor Statistics, Office of Productivity and Technology (unpublished matrix showing average annual growth rates between pairs of years). Real per capita consumption figures from Table 2 of this article.

after 1973, from 3.0 percent per year to 1.7 percent per year.³ The striking aspect of the consumption slowdown is how universal it has been: Growth rates for every major consumption category—durables, nondurables, and services—declined after 1973 (Table 2). The more detailed components of consumption show a similar picture: Per capita consumption growth rates declined after 1973, and they declined in both the relatively easy-to-measure components, and also in the hard-to-measure components. Food and housing are relatively easy to measure; in both these components, per

³ Before January 1996, the Bureau of Economic Analysis (BEA) used a fixed-weight Laspeyres index to calculate real gross domestic product (GDP), including real PCE, in which 1987 price weights were used for all years. Currently, a chain Fisher Ideal index number system is used. The post-1973 consumption slowdown is also evident in the previous data, in which the growth rates were: 1959-73, 2.8 percent; 1973-94, 1.6 percent. (PCE for 1995 was not published using the old system).

⁴ Table 2 probably understates these differences. Because comparable pre-1959 NIPA consumption data are not at present available, Table 1 omits the strong upward surge in per capita consumption of the 1950s.

Table 2

Average Annual Growth Rates: Real Per Capita Personal Consumption Expenditures (PCE)

| Period | Total PCE | PCE Durables | | | PCE Nondurables | | | PCE Services | | |
|----------------------|-----------|--------------|----------------|-----------|--------------------|------|--------------|--------------|--------------------|------|
| | | All Durables | Motor Vehicles | Durables* | All Non-durables † | Food | Non-durables | All Services | Housing Services ‡ | |
| 1959-73 | 3.03 | 5.35 | 5.51 | 5.21 | 1.89 | 1.13 | 2.75 | 3.43 | 3.52 | 3.39 |
| 1973-95 | 1.72 | 2.77 | 1.24 | 3.98 | 0.89 | 0.58 | 1.21 | 2.08 | 1.69 | 2.23 |
| 1973-84 | 1.72 | 2.51 | 1.97 | 2.92 | 0.83 | 0.54 | 1.14 | 2.27 | 2.10 | 2.33 |
| 1984-95 | 1.73 | 3.04 | 0.52 | 5.05 | 0.95 | 0.61 | 1.29 | 1.90 | 1.29 | 2.13 |
| Decades | | | | | | | | | | |
| 1963-73 | 3.55 | 6.30 | 5.89 | 6.65 | 2.47 | 1.83 | 3.18 | 3.71 | 3.60 | 3.77 |
| 1973-83 | 1.47 | 1.47 | 0.67 | 2.10 | 0.65 | 0.51 | 0.81 | 2.17 | 2.01 | 2.24 |
| 1983-93 | 1.97 | 3.81 | 2.14 | 5.17 | 0.96 | 0.54 | 1.40 | 2.16 | 1.51 | 2.41 |
| Peak-to-Peak Periods | | | | | | | | | | |
| 1960-69 | 3.28 | 5.64 | 5.28 | 5.94 | 2.19 | 1.56 | 2.89 | 3.65 | 3.59 | 3.67 |
| 1969-73 | 3.06 | 6.08 | 6.42 | 5.76 | 1.84 | 0.76 | 3.04 | 3.24 | 3.46 | 3.15 |
| 1973-81 | 1.29 | 0.36 | -1.50 | 1.84 | 0.61 | 0.39 | 0.86 | 2.14 | 2.40 | 2.02 |
| 1981-90 | 2.49 | 5.27 | 5.23 | 5.40 | 1.34 | 1.00 | 1.70 | 2.60 | 1.66 | 2.98 |

* Less motor vehicles. † Less food. ‡ Less housing.

SOURCES: Total PCE durables, nondurables, and services are from the national income and product accounts (NIPA), Table 8.3: Selected Per Capita Product and Income Series in Current and Chained Dollars. Survey of Current Business, August 1996. Motor vehicles, food, and housing are from unpublished BEA detail, provided by Greg Key, National Income and Wealth Division. Other columns: Special tabulations provided by Sherman Hammack, National Income and Wealth Division, converted to per capita basis by Robert McCahill.

does Figure 1, but the picture is similar. The strongest half-decade after 1973 (1983-88) does not reach the growth rates of the strongest pre-1973 half-decade (1963-68), although it is roughly comparable (except for motor vehicles) to the second slowest half-decade (1968-73). But other post-1973 half-decades show far slower growth than half-decades before 1973.

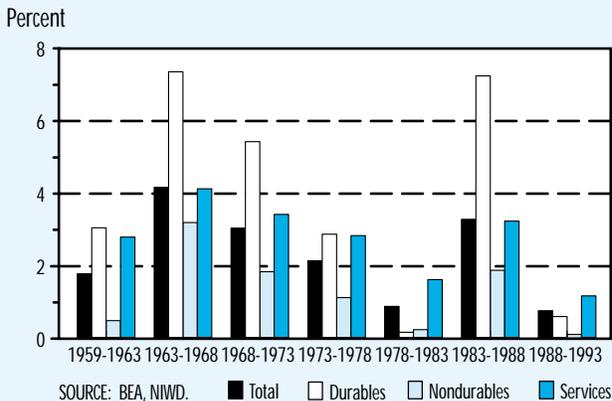
In summary, the consumption growth rate decline before and after the major watershed year of 1973 has been pervasive within subperiods and across major components. The data show a more-or-less abrupt slowdown at 1973, relative stagnation through the 1970s, especially toward the end of the decade, and recovery in the 1980s. Still, growth is not as strong as in the 1960s.

Notice an important part of the consumption growth story: Despite the slowdown after 1973, per capita consumption in 1995 was substantially higher than it had been two decades before. A 1.7 percent annual growth rate is a 46 percent increase in the per capita living standard in 22 years. Contrary to some assertions in the press, the United States has not experienced stagnant living standards. Any notion that per capita living standards have not increased during the past two decades is conclusively refuted by the data on real consumption.

The data tell their own story. But the data have been criticized, particularly the deflators. Before turning to the deflation issues, I address in the next section a relatively neglected topic—issues in

Figure 2

Growth in Per Capita Personal Consumption Expenditures (PCE)



SOURCE: BEA, NIWD. ■ Total □ Durables □ Nondurables ■ Services

measuring the numerator of the real consumption ratio.

HOW IS CONSUMPTION MEASURED?

I first consider how “consumption” is measured in the national income and product accounts (NIPA), because national accounting conventions will affect the interpretation that one can put on the data.

PCE: What’s Included?

The term used in national accounts is not *consumption*, but *personal consumption expenditures*, or PCE. Consumption and consumption expenditures are not precise synonyms. Five rules, which correspond to five types of expenditures, determine the composition of PCE.⁵

- (1) New goods and services purchased by individuals, mainly from the U.S. business sector.
- (2) New goods and services purchased by nonprofit organizations (that serve individuals), mainly from the U.S. business sector.
- (3) Used goods purchased by individuals and nonprofit organizations from other sectors. These entries require offsets elsewhere in the accounts, to avoid double counting.⁶

(4) Goods and services purchased by individuals (and also by nonprofits, if data are obtained) abroad (i.e., foreign travel). Expenditures by nonresidents are deducted to obtain net foreign travel. Net foreign travel also requires offsets elsewhere (to imports and exports) to avoid double counting.

(5) Specific imputations that are exceptions to the usual GDP rule of recording only market transactions.

The rationale in U.S. Department of Commerce (1990) for expenditure rules 3 and 4 is that inclusion of these items makes PCE “more useful for the analysis of consumer behavior.” PCE imputations (expenditure rule 5) are explained as necessary to keep GDP invariant to a list of five conditions. As examples, food grown and consumed on farms is imputed to PCE so that GDP is invariant to whether the food is marketed, and rent is imputed for owner-occupied housing in PCE to keep GDP invariant to whether the housing stock is rented or owner-occupied. The five expenditure rules are not justified in U.S. Department of Commerce (1990) by reference to the concept of consumption, but they are generally consistent with it.

It can be difficult to see the implications of all these accounting rules, and working through them can be somewhat tedious. To illustrate the accounting system’s implications for the interpretation of consumption, I consider whether food consumed at this conference’s luncheon is consumption in the national accounts. Actually, there are two questions: Is the lunch “consumption” within the PCE definition? And is it *food* consumption?

On the first expenditure rule, new goods purchased by individuals from the U.S. business sector are part of PCE. Thus, if we each buy our own lunch and we are not on business, government, or other expense accounts, the lunch is consumption and it is food consumption (specifically, purchased meals and beverages).

If the lunch were provided by a nonprofit institution (e.g., Murray Weidenbaum’s

⁵ Much of the following is drawn from U.S. Department of Commerce (1990). For economy of space, the description is simplified and numerous details are omitted.

⁶ Additionally, PCE includes a dealer margin on purchases of used goods.

Center for the Study of American Business at Washington University), it is part of PCE because purchases by nonprofit educational and research institutions are included in PCE (expenditure rule 2). However, this lunch is not food consumption in PCE; it is instead put in educational expenditures.

Suppose this conference were held in Canada. Food consumed during foreign travel, other than business travel, is not counted as food consumption, but rather as net foreign travel (expenditure rule 4). Thus, the lunch would be consumption in the U.S. NIPA, but it is not food consumption. Note that food consumed in nonbusiness domestic travel appears in the purchased-meals category of PCE. In part, this distinction is data-source driven: Food consumed on domestic trips will be recorded in the sales of U.S. restaurants—the major source for PCE purchased meals—but data on foreign travel are obtained from travel surveys that may not have reliable detail by types of purchases.

If the lunch were provided by a governmental unit, such as the Federal Reserve Board of Governors, it is neither food consumption nor consumption. It is counted in the calculation of GDP because it is a government purchase, and government purchases are final products in the accounts.

Finally, if lunch is provided by a business unit that is not one's employer, it is not consumption. Rather, it is an intermediate input purchased by business. Because intermediate inputs are not final products, such a lunch does not even enter into the calculation of GDP. (If lunch is provided by an employer as part of compensation, then it is included in NIPA wages and salaries and for consistency also appears in PCE. If the employer-provided lunch is on a business expense account, however, it is an intermediate expenditure and is not in PCE.) Because this conference lunch is paid for by the Federal Reserve Bank of St. Louis, a profit-making institution, the lunch is not *food* consumption, and is not *consumption* at all.

The lunch example shows that the conventions for measuring PCE in national accounts do not always give us a number that corresponds to what we intuitively

think of under the rubric “consumption.” Many of these accounting conventions serve to keep the different parts of the accounts consistent and to avoid double counting. Nevertheless, economists use PCE as a measure of consumption. The measurement may not fit the concept of consumption as economists use it.

PCE: Data Sources for Current-Dollar Estimates

Given the definitions for PCE, the first step in computing per capita real consumption is to estimate aggregate, current-dollar PCE. PCE is compiled component by component, with no single estimating method applying to every component. Table 9 of U.S. Department of Commerce (1990) contains separate source data descriptions for each of the 90 lines that were then published in the NIPA consumption tables. The footnotes to Table 9 list 58 major data sources for PCE, ranging from major government data programs (such as the Census of Retail Trade and the Census of Services Industries) to private-sector sources (such as the Edison Electric Institute, the National Automobile Dealers Association, and the National Football League).

As with other parts of the NIPA, the PCE is estimated in three stages, each of which revises earlier estimates—quarterly, annual, and benchmark estimates. The most important data source for quarterly and annual PCE estimates is the monthly Retail Trade Survey conducted by the Census Bureau. Data available annually and data having too long a lag to be incorporated into the quarterly compilation of GDP are introduced in an annual revision.

Quarterly and annual changes in PCE components are periodically (usually, every five years) “benchmarked” to levels derived from the BEA input-output (I-O) tables and the economic censuses. Put another way, the PCE benchmark level, established by the I-O table and the economic censuses, is extrapolated forward by available monthly, quarterly, and annual data until it is again benchmarked at the end of the next five-year period. Thus, any deficiencies that may

exist in current economic surveys may be corrected in principle by the periodic benchmarking process.

For most PCE components, the benchmark-year estimate is determined through what BEA calls the “commodity flow” method. Table 10 of U.S. Department of Commerce (1990) describes eight steps in the commodity flow method:

1. *Identify commodities purchased by persons, using commodity lists from the economic censuses and the Standard Industrial Classification (SIC) manual.* For example, major appliances—including refrigerators, dishwashers, and washing machines—are designated as consumer products in the PCE. Therefore, all subsequent steps apply to sales of these products. Some of the commodity lists used to identify PCE commodities, however, are egregiously out of date.⁷ Failure to bring new products into price indexes quickly enough is a recurring theme of the price index literature, as noted by the CPI Commission (1996). Such a lag also poses a potential problem with the current-dollar data (“on-line” services are an example).

2. *Estimate total domestic shipments of each of these commodities, primarily from detailed product shipment information in the economic censuses (i.e., the censuses of manufacturers, service industries, and housing).* This step requires that total U.S. production (e.g., of major appliances, food commodities, or telephone services) be estimated accurately.

3.&4. *Add imports and subtract exports.* These steps convert U.S. production to an estimate of domestic supply.

5. *Add trade margins, transportation costs, and taxes for each commodity.* These adjustments are necessary to value the consumption commodities at prices purchasers pay. Many steps are required that must be done separately for each commodity. The main information sources are the economic censuses.

6. *Subtract change in trade inventories.* To estimate inventories by commodity involves

a number of steps that need not be spelled out here, except to note that inventory data are collected by industry and must be reworked considerably to get inventory-by-commodity detail.

7. *Allocate total domestic purchases by commodity, as calculated in the first six steps, among business, government, and household purchasers.* This is a crucial step. In the 1996 comprehensive revision, new information on the allocation of restaurant meals between businesses and households raised the level of PCE (and therefore GDP) and increased the PCE’s rate of change. For most commodities, the PCE proportion of domestic supply is obtained residually, by subtracting government and business purchases from total domestic supply. “Estimates of business purchases are derived in part from Census Bureau data on purchased materials and services, but because such data are not available for all business, most business purchases must be estimated using other data and, where necessary, judgment in place of data” (Department of Commerce, 1990, p. 33). Data on purchased inputs—especially business services and purchases by state and local governments—are not the strong points of the economic censuses. Department of Commerce (1990) gives as an example the difficulty of allocating carpet production among household, business, and government purchases.

8. *Adjust data from the first seven steps for intersector sales of used commodities, where applicable.* The best known example is used vehicle sales from the business and government sectors to the consumer sector.

Thus, benchmark estimates for home-consumed food combine food shipments and margins data from the censuses of agriculture, manufacturers, wholesale trade, and retail trade and from foreign trade statistics. Next, an estimate for food used for business purposes, primarily in restaurants and by institutional food providers, is subtracted. Benchmark estimates for purchased meals and beverages

⁷ Autogiro is a product entry in the 1987 SIC manual. The autogiro, a bit like a cross between a biplane and helicopter, has not been produced since 1934. A number of musical instruments listed in the 1987 manual (including auto-phones and octophones) are also obscure, even to experienced musicians.

are obtained from the censuses of retail trade and services industries, with additional information from government sources (tax receipts). Again, BEA must estimate the proportion of total restaurant sales attributable to business purposes, with the residual being restaurant meal purchases in the PCE.

For a small number of commodities, the BEA uses direct estimates of consumer expenditures. For example, Consumer Expenditure Survey data provide estimates of consumer expenditures on personal computers, auto and truck rentals, nursery schools and day care, and a few other commodities. Direct estimation is typically used for components where data on business purchases are considered unreliable (see Step 7).

An Alternative Data Source: The Consumer Expenditure Survey

One way to evaluate PCE current-dollar consumption is to examine alternative consumption data sources. The major alternative source for consumer expenditures is the Bureau of the Census/Bureau of Labor Statistics Consumer Expenditure Survey (CE).

In the CE, expenditures are collected from two samples of 5,000 “consumer units”—a *diary* for frequently purchased items (e.g., food) and a *quarterly recall instrument* for larger items (e.g., autos) and for regularly-billed consumption items (e.g., utilities and telephones). The two samples are blown up to national totals, using probability-of-selection weights and adjustments to ensure that ratios of persons in the consumer units by age, sex, and race conform to national U.S. population totals.⁸ The BLS integrates the two surveys and publishes the results. For 1984 to the present, the CE total applies to the entire (urban and rural) civilian noninstitutional population.

For comparison with the PCE, several attributes of the CE need emphasis. First, like the PCE, the CE is a measure of consumer expenditures, not necessarily of consumption. A lunch served at a conference, for example, is excluded from the CE’s defini-

tion of consumer expenditures. The PCE excludes this as well. Also excluded from the CE are expense account meals and—unlike the PCE—any expenditures (including conference lunches) of nonprofit institutions. The list of exclusions from CE is similar to the exclusions in the PCE, except for the treatment of nonprofit organizations and the CE’s exclusion of expenditures for gifts.

Components of the CE are subject to sampling and nonsampling errors. According to BLS, nonsampling errors can be attributed to many sources, such as definitional difficulties, inability or unwillingness of respondents to provide correct information, errors in collection, and response errors.⁹

PCE and CE: Relative Strengths and Weaknesses

The two data sources on consumption—PCE and CE—have offsetting strengths and weaknesses. The CE collects information on what households purchase. For many components, the PCE obtains consumer expenditures residually: by subtracting from total domestic supply those purchases of consumption commodities that are made by sectors other than the consumer sector. Normally, one expects that a direct measure of an economic variable is more accurate than an indirect and roundabout estimating procedure.

Both sources of information on consumption are subject to nonsampling, or reporting, biases. Reporting biases are known to be serious in some CE components. Consuming units drop out of the quarterly survey before completion, perhaps representing serious response bias, because attrition is probably not random. In addition, the CE’s sample size (5,000 consumer units), is certainly too small for almost any use for which one wants consumption data.¹⁰ Moreover, the CE lacks any natural universe statistic to which the sample estimates of consumption can be benchmarked. Thus, the CE’s small sample size and lack of a benchmarking statistic means that its estimates for smaller components, particularly (e.g., household textiles), are not as reliable as one would

⁸ See U.S. Department of Labor’s BLS Handbook of Methods (1992, p. 174). Note that these adjustments, or “blow up” factors, essentially ensure that the count of persons in the consumer units matches the count of persons in the population—not that consumption expenditures in the CE sample match some universe tabulation of consumer expenditures.

⁹ U.S. Department of Labor (1992, p. 175).

¹⁰ The recently announced increase from 5,000 to 7,500 CE consumer units is a positive, but grossly insufficient, step. The CE is the federal government’s only general-purpose survey of consumer expenditure. It is widely employed for all kinds of analytic purposes—tax and other policy analyses, for example. For comparison, the Canadian consumer expenditure survey will soon have a sample size of 36,000.

want for serious research on consumption. Also, the weights for the *individual 207* basic components of the CPI are not determined accurately from a CE of only 5,000 consuming units, although it may also be true that the variance imparted into the *overall* CPI may be small.

As noted earlier, retail trade surveys are the backbone of the PCE. Retail trade surveys are much larger than the CE—22,000 establishments in the Annual Retail Trade Survey. Such surveys are also subject to reporting biases and nonsampling errors that are probably not documented as thoroughly as have been those of the CE. One particularly troublesome survey error is “birth bias”—no adequate mechanism exists for bringing new retail establishments immediately into the sampling frame. Business survey reports are generally based on business records and so may be less subject to memory errors. The availability of economic censuses every five years permits benchmarking of monthly retail trade surveys that should greatly attenuate the effects of reporting and other biases in these surveys on the PCE.¹¹ The U.S. Department of Commerce methodology paper on personal consumption expenditures (1990) emphasizes the discipline provided by the I-O table: Everything must go somewhere. The I-O provides a series of cross-checks that impose consistency on the data.

On the other hand, the I-O discipline is probably better at the higher levels than it is for the details. Total production of curtains, drapes, sheets, and other household textiles may be allocated in a consistent manner among different final users, but is the consumption share correct? The finer the level of detail, the more likely that the long chain of computations necessary to reach the PCE's indirect estimate of consumer spending will have cumulative errors that affect the totals. In an unrelated paper (Triplett, 1996a), I found it very difficult to determine the consumption of semiconductors by the U.S. computer equipment industry. At this level of detail, the I-O table also rests on bits of data and bits of assumption.

Thus both consumption data series have weaknesses. However, both have strengths.

The individual components of PCE and CE have been studied too little to permit conclusions about which is better and what can be learned from comparing the two. In the next section I present some pertinent information.

PCE AND CE LEVELS AND GROWTH RATES

Level Comparisons: PCE and CE

For matched components of the PCE and CE, Branch (1994) computes correspondence ratios for 1989-92. The correspondence ratio for component *i* is simply the level of its national estimate derived from the CE divided by the corresponding estimate from the PCE. These correspondence ratios have recently been updated to 1995: Table 3 contains an extract of the updated rates supplied by Branch.

Regrettably, classifications of consumption goods and services in the PCE and the CE (and therefore the CPI) differ. This noncompatibility has always plagued users of consumption data. It also makes it far more difficult for personnel within the two agencies to understand and explain the differences. Branch (1994, p. 48) made adjustments to a number of CE components to make them comparable to methodology in the PCE. However, noncomparability may remain. Certain components, including owner-occupied housing, health care, and some smaller items, were excluded from Branch's comparisons because of the degree of their noncomparabilities. Accordingly, aggregate PCE and CE levels in Table 3 should be interpreted as totals for comparable PCE and CE components only.

Branch emphasizes that certain consumption components with high CE/PCE correspondence ratios (including autos, rent, and gasoline, for which the ratios are essentially unity) also have high reporting rates in the CE and low coefficients of variation. Essentially, this is the good news: Where internal statistical analysis of CE

¹¹ Data for small businesses are not collected in the economic censuses, but are estimated from administrative sources.

Table 3

Comparison of CE and PCE: 1992-95

| Expenditure Category | Ratio of CE to PCE | | | |
|---|--------------------|------|------|------|
| | 1992 | 1993 | 1994 | 1995 |
| Food, total | 0.72 | 0.71 | 0.70 | 0.68 |
| Food at home | 0.74 | 0.74 | 0.73 | 0.73 |
| Food away from home* | 0.68 | 0.65 | 0.64 | 0.62 |
| Alcoholic beverages | 0.40 | 0.34 | 0.35 | 0.34 |
| Rent, utilities, and public services † | 0.98 | 0.97 | 0.99 | 0.95 |
| Rented dwellings, total | 1.04 | 1.01 | 1.03 | 0.97 |
| Utilities, fuels, and public services | 0.92 | 0.91 | 0.94 | 0.92 |
| Telephone | 0.89 | 0.89 | 0.88 | 0.85 |
| Household operations ‡ | 0.84 | 0.80 | 0.83 | 0.77 |
| Household furnishings and equipment | 0.66 | 0.65 | 0.65 | 0.66 |
| Apparel and services | 0.62 | 0.57 | 0.55 | 0.56 |
| Transportation | 0.72 | 0.71 | 0.74 | 0.68 |
| Vehicle purchases § | 1.08 | 1.05 | 1.14 | 1.04 |
| Gasoline and motor oil | 0.93 | 0.92 | 0.94 | 0.92 |
| Other vehicle expenses ê | 0.36 | 0.35 | 0.36 | 0.32 |
| Maintenance and repairs, total | 0.31 | 0.30 | 0.29 | 0.24 |
| Vehicle rental and other charges | 0.84 | 0.83 | 0.91 | 0.94 |
| Public transportation | 0.58 | 0.60 | 0.69 | 0.61 |
| Entertainment | 0.58 | 0.58 | 0.53 | 0.51 |
| Fees and admissions | 0.53 | 0.52 | 0.51 | 0.47 |
| Televisions, radios, sound equipment | 0.65 | 0.71 | 0.61 | 0.57 |
| Pets, toys and playground equipment | 0.62 | 0.63 | 0.59 | 0.63 |
| Other entertainment supplies, equipment | 0.50 | 0.42 | 0.37 | 0.37 |
| Personal care products and services | 0.66 | 0.63 | 0.64 | 0.63 |
| Reading | 0.53 | 0.52 | 0.50 | 0.46 |
| Tobacco products and smoking supplies | 0.55 | 0.57 | 0.55 | 0.59 |
| Miscellaneous # | 0.27 | 0.23 | 0.26 | 0.24 |

NOTE: Sums may not equal totals due to rounding. Expenditure estimates for home ownership, insurance, capital improvements, health care, finance charges, education, and cash contributions are excluded from the comparisons.

SOURCE: U.S. Department of Labor, Bureau of Labor Statistics, Consumer Expenditure Survey, 1994-95, bi-annual bulletin, forthcoming.

* Excludes school lunches and meals as pay.

† Includes rent for tenant-occupied dwelling units and lodging away from home and at school. Rent in the CE is contract rent, which includes utilities for some renters. The CE covers direct costs of utilities and fuels by homeowners and renters. In PCE, data are for space rent, which excludes charges for utilities. PCE data cover total expenditures for utilities and fuels even if paid by landlords.

‡ Excludes amounts for baby-sitting, day-care centers, and care of invalids or the elderly.

§ PCE estimates are derived using estimates of dealer margin (a concept that cannot be matched to CE) and wholesale value of net transactions between persons and government, foreigners, and nondealer businesses. CE data on vehicle purchases and trade-ins were combined to approximate total value of new vehicle purchases. CE data on used vehicle purchases, trade-ins, sales, and losses were combined to approximate the value of net transactions of used vehicles.

ê Includes vehicle rentals, maintenance and repairs, and other vehicle charges. The estimates exclude aircraft rentals, vehicle licenses, vehicle inspection, and vehicle registration.

CE estimates exclude expenditures for other properties.

components suggests we can have confidence in the CE data, the component also agrees with the PCE estimate derived from retail trade information.¹²

At the other end of the correspondence ratio scale are “vice” products, alcoholic beverages and tobacco (CE/PCE correspondence ratios of 0.39 and 0.54, respectively). It is widely accepted that respondents to consumption surveys typically under-report these expenditures, so one expects low CE/PCE ratios. Gambling losses, for which the PCE has three categories amounting to some \$40 billion, also appear under-reported in the CE (personal conversation with Clint McCully). For vice products, the PCE is probably more accurate, because it records retail sales, rather than the quantities people say they buy. Branch (1994, p. 48) notes: “PCE data are based mostly on administrative and establishment data, which we expect to be more complete, whereas the CE data are collected via a household survey, which is subject to under-reporting.” On the other hand, one should not overlook the inherent shortcoming of the PCE’s residual-estimation methodology. PCE estimates of alcohol consumption depend on estimates of business purchases of alcoholic beverages. Because consumption is determined residually, by subtracting business (and government) purchases from domestic supply, the PCE for alcoholic beverages will be too high if business purchases are underestimated.

In short, for both high- and low-correspondence ratios, internal statistical analysis of the CE suggests confidence in the PCE totals. For most consumption components, however, CE/PCE correspondence ratios lie somewhere in the middle.

The correspondence ratio for household furnishings and equipment (including furniture, major and small appliances, and household textiles) was only 0.66 in 1995. Some major appliances included in new houses—mainly refrigerators and laundry equipment—are removed from NIPA housing investment and placed in PCE. They are excluded from CE expenditures on major appliances, so in this case one expects PCE to exceed CE. However, household furnishing

and equipment estimates in PCE depend on the accuracy of reported business and government purchases (e.g., curtains for offices, towels for hotels, and furniture for both).

The ratio for radio, television, and sound equipment is only 0.57. For books and magazines (reading supplies), the ratio is 0.46. It is not clear why these components should be subject to under-reporting on this scale in the CE. Both business and government purchase radio, television, and sound equipment. Are business and government purchases of these products under-reported? If so, PCE estimates will be too high. For components having middling correspondence ratios, one needs to examine possible biases in both CE and PCE.¹³

For food components, additional information exists from private-sector tabulations of grocery store sales. An extract of comparisons made by Branch (1994) is presented in Table 4.

The CE/PCE correspondence ratio for food at home is 0.74, so the CE reports roughly 25 percent lower total at-home food expenditures in 1992 than does the PCE. In turn, however, estimates of grocery store food sales from two trade sources (Table 4), are lower yet—10 percent to 20 percent lower than the aggregate CE estimate. Part of the difference is undoubtedly accounted for by sales of food by nonsupermarket retail outlets, which are included in both CE and PCE. Yet, the sizes of the differences in Table 4 are surprising. For most of the detailed components of food consumption, PCE food expenditures are higher than all three alternative sources, with the possible exceptions of fish and seafood and fresh fruits and vegetables, where the data are inconclusive. The increasing sale of fresh fruits and vegetables in informal “farm markets” and so forth should show up in the CE, but not in estimates based on grocery store sales. On the other hand, the *Supermarket Business* level of fresh fruit and vegetable expenditures presumably includes sales to business. This would make its total higher than the CE and PCE.

Branch (1994) also compares CE expenditures for applicable consumption

¹² Obviously, correspondence ratios will also be high where CE data are the basis for PCE, such as vehicle rentals.

¹³ Branch (1994) also noted ratios that cannot be reconciled completely. The CE’s small sample size means that it may not be accurate at detailed levels.

Table 4

Food Expenditures, 1992

| Food Category | PCE | Ratios, Levels of Alternative Sources to PCE | | | |
|-------------------------------------|--------------------|--|------|----------------------|--------------------|
| | Total (\$Billions) | PCE | CE | Supermarket Business | Progressive Grocer |
| Total | 350.4 | 1.00 | 0.74 | 0.67 | 0.62 |
| Cereals and cereal products | 22.0 | 1.00 | 0.64 | 0.63 | 0.55 |
| Bakery products | 39.7 | 1.00 | 0.68 | 0.63 | 0.55 |
| Beef, pork, other meat, and poultry | 72.1 | 1.00 | 0.81 | 0.75 | 0.73 |
| Fish and seafood | 6.3 | 1.00 | 1.23 | 1.03 | 0.44 |
| Eggs | 2.6 | 1.00 | 1.08 | 0.52 | 0.72 |
| Fresh milk and cream | 11.1 | 1.00 | 1.21 | 0.67 | 0.81 |
| Other dairy products | 21.5 | 1.00 | 0.78 | 0.56 | 0.65 |
| Fresh fruits and vegetables | 28.2 | 1.00 | 0.90 | 1.29 | 1.01 |
| Processed fruits and vegetables | 26.6 | 1.00 | 0.62 | 0.56 | 0.43 |
| Sugar and other sweets | 27.6 | 1.00 | 0.37 | 0.27 | 0.17 |
| Fats and oils | 9.2 | 1.00 | 0.78 | 0.69 | 0.57 |
| Nonalcoholic beverages | 41.8 | 1.00 | 0.51 | 0.52 | 0.45 |
| Miscellaneous prepared foods | 46.4 | 1.00 | 0.83 | 0.64 | 0.72 |

SOURCE: Branch (1994).

components with Energy Department surveys of energy purchases and with information in the American Housing Survey. For most of these components, different consumer expenditure surveys give fairly consistent results, many of them with correspondence ratios between 0.9 and 1.1. Because these rent and utilities components are also among the components for which PCE and CE are relatively close (Table 3), comparisons across these alternative consumer expenditure surveys confirm previous judgments about components with high CE/PCE correspondence ratios. However, they cast little light on the many consumption components for which PCE and CE estimates differ.

In sum, if one believes the recent rate of growth of PCE is too low (because PCE is somehow missing substantial amounts of consumer expenditure), directly collected consumer expenditure data do not show it. For most categories of consumption, PCE national levels are above national totals from the CE and also above estimates of other available alternative sources. If directly collected consumer-expenditures

data (for, say, food) were thought to be accurate, then recent PCE consumption measures are too high.

Growth Rate Comparisons

For the purposes of this article, growth-rate comparisons between PCE and CE are even more relevant than are level differences. They pose a number of statistical problems.

The ongoing, quarterly CE began in 1980. Data for 1982-84 were used to construct weights for the CPI in part because data for the first two years of the quarterly CE survey were considered less reliable than subsequent collections. However, consistent CE data are available only after 1984. For my purposes, I have used an unpublished tabulation of 1982 CE data (available within BEA) that contains adjustments to the 1982 data for comparability with later CE surveys.¹⁴

Before the advent of the quarterly CE, expenditure surveys were conducted in 1960-61 and 1972-73. The 1960-61 survey

¹⁴ I am indebted to Greg Key, BEA Consumption Branch, for providing this information.

Table 5

PCE – CE Average Growth Rates: 1973-94

| | Full period | First Subperiod | | Second Subperiod | |
|-------------------------------|-------------|-----------------|---------|------------------|---------|
| | 1973-94 | 1973-82 | 1973-84 | 1982-94 | 1984-94 |
| Total | 0.4 | 0.4 | -0.1 | 0.4 | 1.0 |
| Durable goods | -0.3 | -0.9 | -1.1 | 0.2 | 0.6 |
| Motor vehicles | -0.9 | -0.8 | -1.2 | -1.0 | -0.6 |
| Durables, less motor vehicles | 0.2 | -1.1 | -1.1 | 1.2 | 1.7 |
| Nondurables | 1.1 | 1.1 | 0.6 | 1.0 | 1.5 |
| Food | 1.0 | 1.8 | 1.2 | 0.3 | 0.7 |
| Nondurables less food | 1.1 | 0.2 | -0.1 | 1.9 | 2.5 |
| Services | 0.4 | 0.3 | -0.2 | 0.4 | 1.0 |
| Housing | 0.2 | 0.8 | 0.1 | -0.2 | 0.3 |
| Services less housing | 0.6 | -0.1 | -0.4 | 1.1 | 1.7 |

SOURCES: 1972-1973 CE detail: *Handbook of Labor Statistics* 1978, "Bulletin 2000," Table 130, pp. 466-68. Robert McCahill adjusted the CE detail, using data obtained from Branch (BLS), for comparability to later years. 1984-94 CE detail: BLS on the Internet: <gopher://hopi2.bls.gov:70/00/Special Requests/ce/standard/y84-94.pm%09+Text/plain>29 Aug 1996. PCE detail: National Income and Wealth Division. NIPA Underlying Detail History 1959-1995 (three floppy disks). Release date August 20, 1996.

is generally regarded as noncomparable with later surveys and inferior methodologically. BLS has adjusted the 1972-73 CE survey for consistency with the post-1984 survey. I use these adjusted data here.¹⁵ Fortunately, the timing of the 1972-73 CE survey corresponds well to the break in consumption growth discussed earlier. This means that we have more-or-less consistent data on consumption growth from CE surveys covering the entire post-1973 consumption slowdown, except for 1974-1981.

Table 5 displays differences in growth rates between PCE and CE. Because the 1972-73 CE is a two-year average, a two-year average of the PCE data was also computed for comparability and used in the growth rates reported in Table 5.¹⁶ I write "1973," but this should be understood as the 1972-73 average.

Because of the nature of the PCE and the CE survey, I have broken the 1973-94 period alternatively at 1982 and 1984: The year 1982 coincides with a GDP benchmark, so it is better for the PCE part

of the comparisons. But, as noted above, 1984 is somewhat better for comparison with later years' CE.

Over 1973-94, PCE grew more rapidly than CE by 0.4 of a percentage point per year. In 21 years PCE consumption has therefore grown 43 percentage points more than CE consumption (11 percent of the 412 percent growth in aggregate PCE consumption during this interval).¹⁷

One might ask whether measured differences in PCE-CE growth rates are an artifact of the CE's inconsistency between 1973 and the end of the period. It is, however, precisely the period for which CEs are consistent—namely, 1984-94—where PCE-CE growth differentials are the largest. If one splits the 1973-94 interval at 1984, all the 20-year PCE-CE growth-rate difference occurred in the final 10 years (1.0 point per year), and the difference between PCE and CE growth rates between 1973 and 1984 is inconsequential (as shown in Table 5, line 1). On the other hand, if 1982 is used to split the 1973-94 period, the excess PCE growth is the same in both

¹⁵ Adjusted data supplied by Raphael Branch.

¹⁶ Therefore, the PCE growth rates in Table 5 differ slightly from those in Table 1, where 1973 (rather than the average of 1972-73) is used.

¹⁷ Note that the percentages cited in this sentence are aggregate, current-dollar growth rates. They are therefore not inconsistent with the real per capita growth rates of Tables 1 and 2.

subperiods (0.4 of a percentage point). Although the 1982-84 period may deserve more attention, the post-1984 CE is consistent over time. It is a better survey than were earlier CEs, so I emphasize the 1984-94 comparisons.

Among the components, the excess of PCE over CE growth rates (or deficient growth in CE, depending on how one looks at it) for the full period is strongest in nondurables (1.1 percentage points per year, equally distributed between food and nonfood) and in services less housing (0.6 of a percentage point per year). In view of my earlier discussion of the PCE measurement of food consumption, note that the PCE recorded a larger growth in food consumption than did the CE during this 21-year period (1.0 percentage point per year). The opposite is true for durable goods: The growth rate difference is negative (-0.3 points overall). This means that the CE grows more rapidly than the PCE. Most of the excess CE growth in durables is in motor vehicles (0.9 points).

When growth-rate differentials are split into subperiods, the patterns are perplexing. The excess of PCE over CE growth rates is considerably higher in the 1984-94 interval than before, in most of the components in Table 5. In durables less motor vehicles, nondurables less food, and services less housing, the sign of the difference changes from a negative to a strong positive. Even in motor vehicles, the change after 1984 is in the positive direction because a strong negative growth-rate difference in the earlier period is replaced by a smaller negative difference after 1984. The sole exception is food, which shows a smaller growth rate discrepancy in the later period. Still, PCE growth after 1984 exceeds CE growth by 0.7 points per year.

Notice that motor vehicles grew more rapidly in the CE in both subperiods. Other durables, however, grew considerably faster in the CE in the *earlier* part of the 1973-94 period (-1.1 percentage points in Table 5), but faster (by 1.7 percentage points) in the PCE in the *latter* part of the 1973-94 period.¹⁸ Although nondurables as a group grew more rapidly in the PCE in both subperiods,

excess PCE growth was primarily in food in the first subperiod (1.2 points) and heavily in nonfood nondurables in the second (2.5 points, compared with a small negative difference in 1973-84). PCE rental housing grew only slightly faster than CE in both subperiods; however, PCE nonhousing services lagged CE growth (-0.4 of a point) for 1973-84 but grew substantially more than the CE measure (1.7 points) in the second.

It is difficult to know what to make of this. Are recent PCE growth rates too slow (thus accounting for the measured per capita consumption slowdown)? Directly collected consumption data do not show it. If anything, CE data suggest that the PCE may have grown too rapidly. Moreover, cases such as vehicles, where the CE grows more strongly, suggest deterioration in correspondence ratios.

We need to do a great deal more work on reconciling CE and PCE growth rates because—without a more adequate reconciliation—we cannot understand whether we have reliable data on consumption from either source.

Conclusions on the Comparison

My discussion on the comparison of levels and growth rates of PCE and CE has been motivated by the data on the post-1973 slowdown. Is there any evidence that the numerator of the real consumption ratio is subject to some downward bias? Is there any evidence that understatement of current dollar consumption in the PCE might have contributed to the post-1973 slowdown in measured per capita consumption?

The evidence reviewed is far from conclusive. One should not over interpret what that evidence says. However, other than perhaps auto expenditures, there is no evidence that the PCE is understating current dollar consumption levels or growth. In fact, for non-auto components, CE levels and growth rates are consistent with the PCE's overstatement of current-dollar consumption growth.

We have little data to evaluate PCE growth rates *before* 1973. However, data

¹⁸This difference is probably not associated with personal computers because CE expenditures on personal computers are used in PCE.

sources for PCE were better *after* 1973: Expanded coverage of services (beginning with the 1977 economic censuses and greatly expanded censuses in the 1980s), and improved data for benchmarking are examples; a comprehensive list of improvements would be long. Better source data clearly improved the accuracy of PCE. Did better data also contribute to the measured post-1973 slowdown? Did data improvements lower the PCE growth rate? We do not know for sure. The post-1973 slowdown could still have been a statistical mirage if pre-1973 consumption growth rates were biased upward.¹⁹

DEFLATION ISSUES

If there is bias in real PCE growth, professional opinion points to the deflators. Most of the deflating indexes for PCE are components of the CPI (73 percent, based on the 1994 composition of PCE).

The CPI Commission (1996) estimated that the aggregate CPI contains an upward bias of 1.1 percentage points per year and suggested that the current bias is greater than it once was. If consumption deflators are upward biased, then real consumption growth is downward biased. If the entire 1.1 CPI bias carried over into the PCE, and no bias existed before 1973, then more than half the post-1973 consumption slowdown might be a statistical illusion.²⁰

This section reviews three sources of measurement error discussed by the CPI Commission. These three sources accounted for about half (0.5 point) of the total 1.1 point bias estimate. I show in this section that only part of the Commission's estimate for these three sources carries through to the PCE, and that these sources do not contribute to the real consumption slowdown. Quality change and new products are discussed in the following section.

Commodity Substitution Bias

The CPI is subject to commodity substitution bias because it uses the fixed-weight Laspeyres formula to aggregate its 207 basic components (i.e., its commodity detail). The CPI Commission called this *upper*

level substitution bias and estimated the bias at 0.15 - 0.20 of a percentage point per year, which is consistent with empirical research.²¹ The Commission recommended use of a superlative index number formula to eliminate substitution bias in the CPI.

However, the PCE is not subject to substitution bias, because the PCE is already computed by a superlative index number. The PCE is not deflated by the aggregate CPI. Deflation in the PCE is done component-by-component, and deflated components are aggregated by means of the Fisher Ideal index number formula. Diewert (1976) shows that the Fisher Ideal index is a superlative index number. He also shows that, because a superlative index number is a second-order approximation to the unknown true index number, superlative index numbers are free from substitution bias. Fisher measures of PCE extend over the entire period 1959 to the present. Accordingly, no substitution bias occurs in real per capita consumption, and substitution bias cannot be a factor in the post-1973 consumption growth slowdown.

Outlet Substitution Bias and Basic Component Bias

Outlet substitution bias exists when new, lower-cost retail outlets provide opportunities that were not available when the CPI sample of stores was selected. Basic component bias exists when the estimator used for one of the 207 CPI basic components does not yield the true expected value for that component (the CPI Commission called this *lower level substitution bias*).²² These two biases are intertwined in a way that has been difficult to disentangle.

The CPI Commission put outlet substitution bias at 0.1 percentage point.²³ Outlet substitution bias in CPI components carries over directly into real PCE. Is it larger after 1973, as some have asserted? Retailing seems to have experienced a "revolution" each decade since at least the 1920s, and no evidence exists that changes in retailing were proportionately greater in the recent than in the more distant past. Because the CPI Commission's estimate is

¹⁹ Again, the CE adds to the puzzle. Although the 1960-61 survey is now regarded as methodologically inferior to later surveys, the level of expenditures estimated from it provides the closest agreement to PCE of any consumption survey. See Slesnick (1992).

²⁰ That is, $1.10 \times .73 = 0.80$ points (because CPI components are used as deflators for 73 percent of PCE). This compares with the slowdown, which is $3.0 - 1.7 = 1.3$ points.

²¹ Empirical estimates of aggregate CPI substitution bias appear in Braithwait (1980), Manser and McDonald (1988), and Aizcorbe and Jackman (1993).

²² Attaching this name to the research results on which the estimate is based implies that the commodity substitution paradigm describes the problem of estimating basic components. It may not. Space precludes discussion of this question here.

²³ The estimate is based on a single study (Reinsdorf, 1993), extrapolated to other components of the CPI to which it might apply. Although this approach undoubtedly puts too great weight on a single study, one expects the sign of outlet substitution bias to be positive, and the CPI Commission's estimate is small. For these reasons, it is not necessary to evaluate the evidence for purposes of this paper. In its response to the CPI Commission, the Bureau of Labor Statistics (1997) questions the Commission's extrapolation of Reinsdorf's study.

Table 6

CPI Item Replacements: 1995

| | Proportion of Annual CPI Quotes* | Percent of All Replacements | Probability of the Event During the Year† |
|---|----------------------------------|-----------------------------|---|
| All item replacements | .0390 | 100.0 | .468 |
| Comparables (direct comparisons) | .0254 | 65.1 | .305 |
| Overlaps | .0005 | 1.3 | .006 |
| Deletions, traditional | .0057 | 14.6 | .068 |
| Deletions, class-mean | .0032 | 8.2 | .038 |
| Direct quality adjustments | .0041 | 10.5 | .049 |
| Continuously priced (no replacements during the monthly pricing interval) | .9610 | 0.0 | .532‡ |

* All CPI item replacements in the components tabulated, as a proportion of all CPI price quotes collected annually in the same category (NB: CPI items that are priced monthly have 12 price quotes per year, so one item replacement per year per item would yield a proportion of .0833). Source: Moulton and Moses (1997, Table 4).

† Assumes all items studied by Moulton and Moses (1997) are monthly items.

‡ The probability that no replacements occur during the year in a monthly item ($= 1 - .468$).

small, historical change in the size of the outlet substitution bias (if any) could not make an appreciable contribution to the post-1973 consumption slowdown.

The CPI Commission estimated basic component bias at 0.5 of a percentage point historically, and 0.25 of a percentage point currently (because of methodological changes BLS made in 1995 and 1996). The 0.5 estimate comes from comparing a geometric-mean estimator for the items within a basic component with the estimator used since 1978 for the CPI. Basic component bias in the CPI also carries over in principle to real PCE growth in those components.

The current basic component bias problem in the CPI begins in 1978. The bias arises because a faulty statistical estimator was devised to implement the then-new probability sample. The BLS made an interim adjustment in 1995, estimated at a little over 0.1 percentage point. This adjustment was carried back to 1978 in the PCE.²⁴ A second BLS change, amounting to another tenth of a point, was made in 1996. Neither this second change nor the remaining quarter point accepted by the CPI Commission have been incorporated into the deflators for real PCE. If one accepts both these

numbers, the PCE would be biased downward in the period since 1978 by perhaps a quarter of a point per year (0.35×0.73).

Is the post-1978 basic component bias in the CPI a statistical factor that contributes to the slowdown in measured consumption growth? The CPI's history before 1978 complicates assessment of this question.

The probability sampling procedures put into the CPI in 1978 were intended to correct another CPI bias that was associated with BLS's former purposeful (and possibly unrepresentative) outlet and commodity sampling procedures. No estimate exists of this earlier bias, nor is the sign of the bias known. The context of the discussion by the Price Statistics Review Committee (1961)—also known as the Stigler Committee report—suggests that the committee believed it was an upward bias. If so, two forms of basic component bias occur in the CPI—one before 1978 and one after 1978—both upward. The relative bias, *pre-1978* and *post-1978* is unknown, but it could not have contributed much to the post-1973 slowdown.

In summary, substitution bias, though present in the CPI, is absent from the PCE and cannot affect the post-1973 real consumption slowdown. Outlet substitution bias affects both CPI and PCE but might be as

²⁴ The effect on real PCE varied by year because of earlier BEA adjustments to housing and other factors. It raised real PCE in some years but lowered it in others. The net effect was small. See Survey of Current Business (January/February 1996).

large before 1973 as after. Basic component bias also affects both CPI and PCE but, again, might be nearly as large before 1973 as after. One part of the historical basic component bias discussed by the CPI Commission has already been adjusted out of the PCE, and at least some of the remainder replaces an earlier upward CPI bias of unknown size that existed before 1978. Consequently, the net effect of these three measurement errors on the measured real consumption slowdown is probably small.

QUALITY CHANGE AND NEW PRODUCTS

The CPI Commission grouped quality change and new products together in its discussion. It is often unclear whether something new is a new product, a new variety of an existing product, or a quality change. The Commission estimated that quality change and new products create an upward bias in the CPI of about 0.6 percentage point annually (CPI Commission, 1996, Table 2).

Combining the discussion of quality change and new products has some advantages. Yet, new products and quality change have different implications for CPI procedures, so combining them in the report obscured the strategy for improving the CPI. Indeed, none of the CPI Commission's fifteen recommendations addresses improving methods for quality change and for new products in the CPI. This is particularly surprising because over half of the Commission's estimated CPI bias came from these sources.

Moreover, combining the discussion of new products and quality change also obscured the distinction between quality changes that are observed within the CPI sample and quality changes that take place outside the sample. In the CPI, probability samples of stores and of products (services) are selected; quality change only intrudes on the CPI computation when a product in the sample is "pushed out" because it is no longer available in a particular CPI retail outlet. The appearance of improved prod-

ucts/services outside the CPI sample has no impact on the CPI, unless the improved product/service displaces an item inside the sample.

There are two questions: What is the bias, if any, of CPI procedures for handling quality change, when quality changes appear on CPI items? Do CPI pricing procedures systematically overlook or under-represent some of what is "new" in the economy? The CPI Commission emphasized the latter question, and that is one of its major contributions to the price index literature. But its analysis of quality changes inside the CPI is the major weakness of its report.

Indeed, some economists contend that studying internal CPI quality adjustment procedures is neither necessary nor relevant. "For most categories in the CPI, the extent of current quality adjustments is irrelevant to an assessment of the treatment of quality change in the commission's report, simply because most of our estimates of quality change bias are valid independent of how the BLS arrives at its estimates of price change or the extent to which its adjustments for quality change are large or small" (Gordon, 1997). For the reader who accepts that view, the rest of the present paper is beside the point.

I believe, however, that understanding and analyzing the implications of CPI quality adjustment procedures is important and relevant. Such information complements independent studies for evaluating CPI quality bias. It is also useful for evaluating independent price index studies, which are not in every case the only truth.

This section on quality change first considers the treatment of quality changes in the CPI. The implications of the methods used in the CPI for handling quality change are not well understood by economists; the CPI Commission did not discuss them adequately, and some of these methods overadjust for quality change, so that *improving* quality can generate *downward* bias in the CPI—which is not what economists usually expect. The second subsection discusses quality changes outside the CPI sample, and the third considers new prod-

ucts. The final subsection wraps up by considering whether changes in the amount of CPI quality change bias can account for the slowdown.

Methods For Handling Quality Change in The CPI

The following paragraphs, expanded from Triplett (1990), explain the treatment of quality changes in the CPI (indeed, in all price indexes) and the implications for price index bias. Empirical estimates use new data from Moulton and Moses (1997).

The quality problem in constructing price indexes arises because of “item replacements” (often called “item substitutions” in BLS materials): a 1997 model car is replaced by the 1998 model, a 25” television set is replaced by one with a 27” screen, a candy bar is replaced by one that is one-quarter ounce smaller or larger, a wheelbarrow with metal handles is replaced by one with wooden handles. Other item replacements occur that are not so clearly “quality” changes: a woman’s spring dress is replaced by a fall style (perhaps made with better material in addition to the style change), a volleyball disappears from a sporting goods store (perhaps because of the end of volleyball season or perhaps because volleyball has become less popular), and must be replaced in the CPI with some other item of sporting equipment. In most of these item replacements, some “old” product is not hanging around to be supplanted gradually by a new one.

About 4 percent (0.0390) of the price quotations collected for the CPI in 1995 involved a replacement (Moulton and Moses, 1997). Although some CPI items are priced bimonthly, most CPI components have monthly pricing cycles, so that each monthly item has twelve price quotations in a year. If all components that Moulton and Moses studied were monthly items, the expected number of replacements per CPI item during the year would be about 0.47—that is, $0.0390 \times 12 = 0.468$. On average, each CPI item encounters about one-half a replacement each year.²⁵ Because multiple replacements occur on some items, Moulton

and Moses (1997) report that about 30 percent of CPI items experienced at least one replacement during 1995. This high replacement rate indicates the pervasiveness of the quality-change problem for price index construction.

When an item that is currently included in the CPI sample disappears from a retail outlet or changes in specification, two prices are observed—the price of the old item, a , in period one, P_{a1} , and the price of the replacement item, b , in period two, P_{b2} . The true price index (I_T) for this item is obtained by adjusting the *price relative*, P_{b2}/P_{a1} , by the true but unknown measure of quality change, A_T , or:

$$I_T = P_{b2}/P_{a1} - A_T.$$

Three alternative mechanisms for dealing with quality change are used in the CPI (a fourth mechanism, overlap pricing, also exists but is employed infrequently). For the components Moulton and Moses studied, about 65 percent of CPI item replacements were handled by the direct comparison method (Equations 1a-1b, below), about 11 percent by the direct quality adjustment method (Equations 2a-2b), and about 23 percent by the two forms of the deletion method (Equations 3a-3b). See Table 6. Each quality change method implies an explicit or an implicit quality adjustment, as explained in the following.

Method One: Direct Comparison. The replacement item is judged essentially equivalent to the one it replaces (that is, the quality difference is “small”). Then, the quality adjustment, A_1 , is zero, and the price index, I_1 , for this item is the full amount of the price relative:

$$(1a) \quad A_1 = 0$$

$$(1b) \quad I_1 = P_{b2}/P_{a1}$$

Moulton and Moses (1997, Table 7) estimate that, in 1995, $I_1 = 1.0251$. Replacements that were deemed comparable were associated with an (arithmetic) average price increase of 2.51 percent in the month

²⁵Suppose there are four items, each priced monthly, so there are 48 annual price quotes. Now suppose there are two item substitutions during the year, either one substitution each in two of the four items, or perhaps two substitutions in one item, the other three items remaining unchanged through the whole year. The proportion of substitutions to total quotes is a little over 4 percent (2/48). But the expectation of item substitution during the year is one-half.

Table 7

Analysis of Monthly Price and Quality Changes, CPI Item Replacements, 1995

| | Direct Comparison (<i>i</i> =1) | Direct Quality Adjustment (<i>i</i> =2) | Deletion Method | | All Quality Adjustments ⁷ |
|---|-------------------------------------|---|-------------------------------|------------------------------|--------------------------------------|
| | | | Traditional (<i>i</i> =3) | Class Mean (<i>i</i> =4) | |
| Observed mean price relative (P_{b2i}/P_{a1i}) ¹ | | | | | |
| Arithmetic, untrimmed | 2.51 | 4.25 | 30.73 | 23.79 | 17.35 |
| Arithmetic, trimmed ² | 2.51 | 4.10 | 5.79 | 11.96 | 6.77 |
| Logarithmic, untrimmed | 2.32 | 3.80 | 4.44 | 8.61 | 5.63 |
| Mean quality adjustment (A_i) ³ | | | | | |
| Arithmetic, untrimmed | 0 | 1.59 | 30.39 | 18.62 | 14.26 |
| Arithmetic, trimmed ² | 0 | 1.44 | 5.45 | 6.79 | 3.68 |
| Logarithmic, untrimmed ⁴ | 0 | 1.18 | 4.19 | 4.49 | 2.78 |
| Mean quality-adjusted (pure) price (I_i) | | | | | |
| Arithmetic, untrimmed ⁵ | 2.51 | 2.66 | 0.34 | 5.17 | 3.09 |
| Logarithmic, untrimmed ⁶ | 2.32 | 2.62 | 0.25 | 4.12 | 2.85 |

SOURCES and NOTES: 1) Computed by adding the corresponding I_j and A_j rates.
 2) Extreme trim ("Method B") — Moulton and Moses (1997).
 3) Moulton and Moses (1997), table 9; but see footnote 4
 4) Special tabulation by Karin Moses of BLS. Differs from corresponding estimates in Moulton and Moses (1997), owing to correction of a computing algorithm error.
 5) Moulton and Moses (1997), Table 7
 6) Special tabulation by Karin Moses of BLS.
 7) Excluding direct comparisons
i = subscript indicating quality change method

in which the replacement occurred (see Table 7). This increase was several times the 1995 average in months where no item replacement occurred (0.12 percent, Moulton and Moses 1997, Table 7). Note that 96 percent of monthly CPI price quotes do not involve an item replacement.

In the direct comparison case, any quality change between variety *a* and variety *b* will be missed. Quality is probably improving, on balance, that is, $A_T > 0$. If so, direct comparison imparts an upward bias into the price index (and a downward bias if quality is deteriorating).²⁶ Note, however, that the average price change for the direct comparison cases (2.51 percent) is not higher than the quality-adjusted price changes for CPI cases where a direct quality adjustment is made (2.66 percent—Table 7). This suggests that the upward bias from ignoring quality in the direct

comparison cases is small. Direct comparison is the sanctioned method for cases where the quality difference between varieties *a* and *b* is small, so it is reasonable that the quality errors are also small (though they might be pervasive).

Additionally, quality changes might have been missed by the pricing agents in the field (on field procedures, see Duff, 1997), so that some of the "continuously priced" CPI items might also have experienced quality change. Undetected quality changes also bias the CPI upward if, on average, undetected quality changes are improvements.

Method Two: Explicit Cost-Based or Hedonic Quality Adjustment. The new item is considered different from the old one, and information exists to make a quality adjustment, λ , where λ is some function of the production cost differential

²⁶ This is not a complete statement, because it does not consider the BLS pricing specification, which holds constant some of the characteristics of the product. Thus, substitutions are conditional (Moulton and Moses's term) on the pricing specification and improving quality in the market might be consistent with deteriorating quality in the CPI sample. An example involving refrigerators appears in Triplett (1971, pp. 190-94). Though the example is not unrealistic, I presume that it is not the dominant case in the CPI.

between items a and b , or is derived from a hedonic function. Then (expressing all variables as rates of change):

$$(2a) \quad A_2 = \lambda, \text{ and}$$

$$(2b) \quad I_2 = (P_{b2}/P_{a1}) - \lambda.$$

Cost-based quality change adjustments are used in the CPI mainly for vehicles and gasoline, and hedonic adjustments for clothing and rent.

Moulton and Moses (1997) present estimates for 1995 for the terms in Equations 2a-2b. A CPI component is calculated as the arithmetic mean of price relatives. Taking arithmetic means across all the examples of direct quality adjustments in the CPI gives (see Table 7):

$$\begin{aligned} P_{b2}/P_{a1} &= 0.0425 \\ \lambda &= 0.0159 \\ I_2 &= 0.0266 \end{aligned}$$

These estimates have the following interpretation: Had these quality changes been ignored—that is, had the old item and the modified item been treated as direct comparisons—prices would have risen by 4.25 percent in the month in which the quality change was observed. After the quality adjustment of 1.59 percent, price increases of 2.66 percent entered the CPI for these products. Table 7 also presents some alternative calculations that are explained in the following subsection.

Conditional on the accuracy of the direct quality adjustments, substantial true price increases occurred on these products, even after quality adjustment. Note that both “raw” and “quality-adjusted” arithmetic mean price increases for items that had received direct quality adjustments are substantially greater than the average increases for unchanged items (0.12 percent, in 1995). These numbers indicate that price increases in 1995 were likely to coincide with product changes.

The price index will be biased if the true quality change differs from the estimated quality adjustment, that is, if $A_T \neq \lambda$. Though the direction of the bias is unknown, it has long been suspected that the use of

manufacturers’ cost data for quality adjustment in the CPI tends to overadjust for the value of quality change, particularly in the case of automobiles.

The CPI Commission judged that all within-sample CPI quality change methods, including manufacturers’ cost, impart upward bias (CPI Commission, 1996, page 38). The Commission asserted that cost-based quality adjustments for automobiles did not include manufacturers’ changes that increased durability (CPI Commission, 1996, page 53) or those that reduced defects (“an important source of quality improvement that is not taken into account in the CPI . . . [is] the marked decrease in the incidence of defects . . . as measured by the J.D. Power survey” [CPI Commission, 1996, page 55]). Although automobiles have undoubtedly become more rust-resistant and more reliable, the CPI Commission’s and Gordon’s (1997) assertion that these quality improvements are missed in the CPI are not well informed. Bureau of Labor Statistics (1997) listed changes, such as increased use of corrosion-resistant metal, for which cost-based quality adjustments for automotive durability have been made in the CPI. Reduced defects must also have come about from changes made by the car makers. In my experience in the BLS, the auto manufacturers never overlooked quality changes when they submitted costs to the BLS. Rather, manufacturers tried to attribute too much price change to quality improvements—I recall one auto manufacturer’s contention that removing the 90 and 100 numerals from a speedometer ought to qualify as improved quality in an automobile. It is conceivable that some quality changes are worth more to the consumer (the correct theoretical basis for quality adjustment in the CPI) than they cost to produce. However, the Commission’s idea that quality adjustments are systematically overlooked by the manufacturers when they make reports to the BLS is inconsistent with my experience with these data and also inconsistent with alternative evidence. All published hedonic studies of automobiles that cover the era of CPI cost-based quality adjustments have produced hedonic price

indexes that rise more rapidly (not less rapidly) than the CPI new automobile index.

Method Three: Deletion. The replacement item is judged noncomparable, but no direct quality adjustment is available. In this case, the item is deleted from the CPI in the month that it changes or exits, and its price change is imputed from price changes of other items in the same component of the index.

The deletion method is given several names in BLS materials: “link method,” or “link without overlapping prices” method. This is inexact terminology because all quality change methods involve links of some form. My use of the word “deletion” emphasizes the fundamental property of the method—the price of the item that changes or is replaced is dropped from the index in the month in which the replacement occurs.

The deletion method’s price imputation creates an *implicit* quality adjustment. The implicit quality adjustment and the imputed price index for the CPI item that changed are given by (again expressing the variables as rates of change):²⁷

$$(3a) \quad A_3 = P_{b2}/P_{a1} - \sum_j w_j (P_{j2}/P_{j1}), j \neq a, b, \text{ and}$$

$$(3b) \quad I_3 = \sum_j w_j (P_{j2}/P_{j1}), j \neq a, b \\ = P_{b2}/P_{a1} - A_3.$$

The second line of Equation 3b is just a rearrangement to emphasize that the deletion method can be expressed as a quality adjustment to a price relative, an adjustment that is formally equivalent to Equation 2b, above.

Evaluation of the Deletion Method. The implications of the deletion method are poorly understood, so it requires an extended discussion. Since the research reported in Triplett (1971), I have been convinced that the deletion method over-adjusts for quality change, or—what is the same thing—it misses price change because it inappropriately counts price change as quality change. This judgment is accepted by BLS staff but rejected by the CPI Commission.

It may be useful to view Equations 3a and 3b from the perspective of the true,

quality-adjusted price index, I_T :

$$(4) \quad I_T = P_{b2}/P_{a1} - A_T$$

Equation 4 says that the true price index for the item that changed is the price relative for the new and the old, adjusted by the true quality change. Comparing Equations 3b and 4, we can describe the bias from application of the deletion method by:

$$(5a) \quad \text{Bias} = I_3 - I_T$$

$$(5b) \quad = A_T - A_3.$$

Consider, first, Equation 5b. The sign of the bias created by the deletion method depends on whether the true quality change, A_T , is greater or less than the implicit quality change adjustment, A_3 , created by the deletion method. That is, the sign of the bias depends on whether $A_T > A_3$ or $A_T < A_3$: The bias under the deletion method occurs when the method over-adjusts or under-adjusts for quality change. The price index bias does not depend on whether quality is improving or deteriorating: *It does not depend on whether $A_T > 0$ or $A_T < 0$.*

Because the bias depends on $A_3 - A_T$, and not on A_T , evidence, anecdotes, or introspection about the prevalence, direction, or magnitude of A_T shed no light, by themselves, on the bias in the CPI from the deletion method. Improving quality ($A_T > 0$) creates a downward bias in the price index when $A_3 > A_T > 0$; it creates an upward bias when $A_T > A_3 > 0$. Deteriorating quality ($A_T < 0$) biases the index upward, when the implicit quality adjustment for deteriorating quality is too small, and downward when it is too large. The sign of the bias is entirely an empirical matter that requires (1) measuring A_3 , the size of the implicit adjustment in the CPI, and (2) comparing it with some estimate of A_T .

Now consider Equation 5a: This shows that the deletion method’s bias depends on whether the imputed price change, I_3 , is greater or less than the true price change, I_T . That is, *the bias depends on whether the true, quality-adjusted price*

²⁷Equation 3a is the same as Equation 3 in Triplett (1990). Recently, what the BLS calls a “class mean” method has been introduced. This qualifies the price quotations that go into the bracket on the right-hand side of Equations 3a and 3b—that is, the class mean method restricts the observations, j , that are used in the imputation—but the basic method is still described by Equations 3a and 3b. Note that in the U.S. Producer Price Index, $A_3 = P_{b2}/P_{a1}$ and $I_3 = 0$, an imputation procedure that has greater bias than the CPI procedure. The PPI method is also used in the CPIs of many countries.

change for the item that changed is greater than or less than the measured price changes in items j that were used to impute I_3 (Equation 3b). CPI bias occurs when too much or too little price change is imputed from items that did not change in specification to the item that did. The deletion method is biased if price changes (upward or downward) are more likely when CPI item replacements occur.

The annual model changeover of new cars is a well-known example of a CPI item replacement. Although vehicles receive cost-based quality adjustments in the CPI, the deletion method has also been employed for cars at times, and using autos as an example of the problems posed by the deletion method has heuristic advantages.

Suppose that each car model had a life of two years, and was relatively unchanged in the "off" year. Or suppose that new car models were introduced throughout the year, rather than in an introduction season in the fall. Equation 3b shows that the price change of a new car model whose quality changed would be imputed from the prices of cars whose quality *did not* change in that month.

Now suppose that it is more likely for car manufacturers to make changes in prices (either up or down) in a month when a new model is announced. The supposition is clearly realistic. In this case, the deletion method is biased toward no price change. In the extreme case where no price change occurred except when a new model was introduced, the price index would never change (because each of the P_{j2}/P_{j1} terms in Equation 3b shows no change). In this case, all of the price change takes place when the models change, and the deletion method removes all price change from the CPI.

Thus, from the perspective of Equation 5a, the CPI bias depends on the price imputations. One might ask: What do we expect of those imputations? Is there evidence that the quality-adjusted prices estimated with the deletion method (i.e., I_3) behave systematically differently from price changes that are not imputed in this way? One

can, of course, also use information about the quality adjustment, the value of A_3 , to assess the bias.

CPI Data on Implicit Quality Adjustments and Price Imputations. Moulton and Moses (1997) provide data from which estimates for all three terms of Equation set 3a-3b can be computed. The data for 1995 are shown in Table 7.

The analysis of deletions is complicated by several factors. First, two forms of the deletion method are used in the CPI. The "traditional" method is the one described above: For example, price imputations are based on all automobile models that did not change. In the newer, "class mean," version of the deletion method, the items j that are used for the imputation are restricted to ones for which quality also changed in the same month—for example, automobile models that received an explicit quality adjustment.

The class mean method implies that the true price change for changed models with no explicit quality adjustment equals the true price change for models for which explicit quality adjustments were made, or changed models that were judged comparable. The traditional deletion method implies that changed and unchanged models have the same true price change, which is less plausible. Separate calculations for the two forms of deletion appear in Table 7.

For administrative reasons, deletion is also used in the CPI for cases that do not correspond to what we normally think of as quality change. This situation introduces a second complication because the administrative data base from which the Moulton and Moses calculations were made does not permit distinguishing true quality changes from other deletions. For example, CPI classes are sometimes defined broadly enough to encompass a group of related products such as sports equipment. When a volleyball is not available in a CPI retail outlet, pricing may switch to some other kind of sports equipment. If so, the deletion method will be used because obviously any price difference between, say, a volleyball and a tennis racket should not affect

the CPI. Several other examples are presented in Moulton and Moses (1997).

Additionally, CPI basic components are computed as arithmetic means of price relatives. Some non-quality deletions in the data base may involve large price differences, and the arithmetic mean is not symmetric for large increases and large decreases. Some quality changes also involve large price changes.

On the logic that items that are replaced by items with significantly higher or lower prices are more likely to have been deleted for non-quality reasons, Moulton and Moses (1997) trim large price changes from the data. They also compute logarithmic means, which also diminishes the influence on the mean of large changes.

Three of the six alternative calculations presented in Moulton and Moses (1997) are summarized in Table 7. Consider first the untrimmed arithmetic means. Because CPI components are arithmetic means, these calculations are the relevant ones for determining the effect of deletions on the CPI, as it is currently calculated and published. The untrimmed arithmetic means answer the question: What would the CPI have been if all the item replacements that were handled by deletion were instead ignored and compared directly?

On average, when item replacements that were handled by deletion were encountered in 1995, very large "raw" price changes were recorded, 30.73 and 13.79 (Table 7). As noted, the magnitude of this P_{b2}/P_{a1} ratio may reflect deletions for reasons other than quality changes.

For traditional deletions, implicit quality adjustments, A_3 , were around 30 percent, so the average, quality-adjusted price change, l_3 , for these item replacements was only 0.34 percent. For the class-mean form of the deletion method, the implicit quality adjustment was lower (18.62 percent), and the price increase after quality adjustment was substantially higher (5.17 percent).

The untrimmed arithmetic average probably overstates the amount of quality adjustment in CPI item replacements. Some of the 24 percent to 30 percent raw price differentials, and some of the 8 percent to 30

percent implicit adjustments, pertain to item replacements that were not quality changes, as we usually think of them. On the other hand, Moulton and Moses (1997) omitted all size adjustments, such as those for the candy bar in the example mentioned earlier. An earlier version of the Moulton and Moses calculations suggests that size adjustments would add another 0.80 to the untrimmed arithmetic mean, for a total of 2.56 points. If size adjustments had not been made in the CPI, the CPI would have risen more than it did. The untrimmed estimates understate CPI quality adjustments for this reason.

As shown in Table 7, the trimmed CPI observations have lower implicit quality adjustments, a result that is almost true by construction. Trimming excludes the largest raw price differentials from the analysis, and implicit quality adjustments in all deletions are a large proportion of raw price differentials. Trimmed arithmetic quality adjustments, A_3 , amounted to 5.45 percent for the traditional deletion method and 6.79 percent for the class-mean cases.

Quality-adjusted price changes were not computed for the trimmed sample; I assume that the true price changes were the same in the trimmed and untrimmed samples, an assumption which implies that the raw price change was substantially lower for observations in the trimmed sample (for example, only 5.79 percent for traditional deletions in the trimmed sample, compared with 30.73 percent for the same method in the untrimmed sample).

Moulton and Moses (1997) also compute logarithmic means from the same data. A logarithmic mean is not affected by asymmetry of large changes, so it provides an alternative to trimming. As expected, all the logarithmic means in Table 7 are lower than the corresponding untrimmed arithmetic means.

However, the CPI is computed with arithmetic, not logarithmic, means. Logarithmic means accordingly answer a different question: How would a logarithmic mean CPI differ if all deletions were treated as direct comparisons? The logarithmic quality-adjusted price changes, l_i , in Table 7 provide information.

The essential relationships are the same for logarithmic as for arithmetic means: The traditional deletion method creates implicit quality adjustments that account for a very high proportion of the raw price changes from item substitutions (4.19 of 4.44 percent raw price change). Moreover, when the deletion method is employed, the resulting logarithmic quality-adjusted price change (at only 0.25 percent) is substantially lower than when other quality change methods are used. This is parallel to the results for arithmetic means. Quality-adjusted price changes from traditional deletion are suspiciously low. Correspondingly, application of the class-mean method yields an implicit quality adjustment that is a smaller proportion of the raw price change and an imputed price change that is considerably higher (4.12 percent).

Table 8 (p. 32) rearranges data from Moulton and Moses (1997) to provide comparisons for major CPI components. Little discussion is necessary because the relationships are the same as in the aggregate data. Raw price differences for item replacements handled by deletion are very large, perhaps surprisingly large. But whether one examines trimmed or untrimmed data, calculated with the actual arithmetic mean CPI or a hypothetical logarithmic mean CPI, the large implicit quality adjustment from traditional deletion removes a very high proportion of the price difference from the index, leaving a very small quality-adjusted price change. For transportation item replacements, for example, and using the arithmetic mean calculation, traditional deletion removes all of the raw price difference ($A_3 = P_{b2} / P_{a1}$), leaving zero adjusted price change. For new cars handled by traditional deletion, $A_3 > P_{b2} / P_{a1}$, leading to negative quality-adjusted price change. In the logarithmic calculation, the implicit quality adjustment removes nearly all of the raw price change, leaving quality-adjusted price changes of 0.04 and 0.10 for these two cases.

The class-mean version of deletion generally results in a smaller implicit quality adjustment, though there are exceptions. However, in every category in

Table 8, the class-mean method leads to a larger quality-adjusted price change than is the case for traditional deletions, and this statement holds for both arithmetic calculations and logarithmic ones. For transportation item replacements handled by the class-mean method, for example, quality-adjusted prices rose by about 4 percent according to the arithmetic calculation (zero for traditional deletions). When the traditional deletion method was used, quality-adjusted prices for apparel, commodities, and new cars fell; when the class-mean method was employed, they rose (by 12¹/₂ percent and 4¹/₂ percent, respectively). In all these cases, the logarithmic calculation gives results of the same magnitude.

What does one make of this? I summarize by turning back to the question at the beginning of this subsection: "Do quality-adjusted price changes imputed by the deletion method behave systematically differently from price changes that are not imputed in this way?"

First, the deletion method is biased only if price changes are more likely when new models or varieties are introduced. Moulton and Moses show that this condition is met—price changes are more likely with CPI item replacements. When replacements do not occur (96 percent of the time), much price stability is observed, at least in low inflationary environments like 1995. Though some readers of Moulton and Moses (1997) have expressed surprise at this observation, it is not really so surprising: The matched model method would work fine so long as price changes for the unchanged items paralleled the true price changes for models that were changed, in which case we would not need to worry much about quality change. But prices of changed and unchanged models don't move in parallel; quality change is a problem for measuring prices because prices are more likely to change when models change.

Second, deletion creates downward bias when price increases accompany item replacements. The Moulton and Moses data show that quality-adjusted prices rose

Table 8

1995 Price Changes and Quality Change, Item Substitutions in the CPI Handled by Deletion

| | Arithmetic, untrimmed | | | Arithmetic, trimmed | | | Logarithmic, untrimmed | | |
|-----------------------------|-----------------------|--------|-------|---------------------|-------|---------|------------------------|-------|-------|
| | P_{b2}/P_{a1} | A_i | I_i | P_{b2}/P_{a1} | A_i | I_i^a | P_{b2}/P_{a1} | A_i | I_i |
| Food | | | | | | | | | |
| Traditional | 29.07 | 28.65 | 0.42 | 4.13 | 3.71 | 0.42 | 1.70 | 1.35 | 0.35 |
| Class-mean | * | * | * | * | * | * | * | * | * |
| Housing (non-rent) | | | | | | | | | |
| Traditional | 37.31 | 36.99 | 0.32 | 6.47 | 6.15 | 0.32 | 8.98 | 8.67 | 0.31 |
| Class-mean | 16.20 | 12.31 | 3.89 | 11.17 | 7.28 | 3.89 | 4.90 | 2.21 | 2.69 |
| Apparel & Upkeep | | | | | | | | | |
| Traditional | 17.08 | 17.55 | -0.47 | 15.56 | 16.03 | -0.47 | 9.47 | 10.77 | -1.30 |
| Class-mean | 48.85 | 36.46 | 12.39 | 25.82 | 13.42 | 12.39 | 22.68 | 13.86 | 8.82 |
| Apparel Commodities | | | | | | | | | |
| Traditional | 19.85 | 20.19 | -0.61 | 18.28 | 18.89 | -0.61 | 11.56 | 13.06 | -1.50 |
| Class-mean | 48.85 | 36.45 | 12.39 | 25.82 | 13.42 | 12.39 | 22.68 | 13.86 | 8.82 |
| Transportation | | | | | | | | | |
| Traditional | 15.36 | 15.36 | 0.00 | 9.17 | 9.17 | 0.00 | 5.86 | 5.82 | 0.04 |
| Class-mean | 9.30 | 5.34 | 3.96 | 8.93 | 4.97 | 3.96 | 6.00 | 2.27 | 3.73 |
| New Cars | | | | | | | | | |
| Traditional | 4.32 | 4.39 | -0.07 | 4.32 | 4.39 | -0.07 | 3.25 | 3.15 | 0.10 |
| Class-mean | 8.95 | 4.47 | 4.48 | 8.54 | 4.06 | 4.48 | 7.39 | 3.13 | 4.26 |
| Medical | | | | | | | | | |
| Traditional | 62.01 | 61.82 | 0.19 | 4.43 | 4.24 | 0.19 | 10.73 | 10.64 | 0.09 |
| Class-mean | * | * | * | * | * | * | * | * | * |
| Entertainment | | | | | | | | | |
| Traditional | 21.27 | 19.94 | 1.33 | 6.96 | 5.63 | 1.33 | 4.67 | 3.63 | 1.04 |
| Class-mean | 127.28 | 124.76 | 2.52 | 15.11 | 12.59 | 2.52 | 12.12 | 10.49 | 1.63 |
| Other | | | | | | | | | |
| Traditional | 27.85 | 27.25 | 0.60 | 6.67 | 6.07 | 0.60 | -0.98 | -1.57 | 0.59 |
| Class-mean | 7.49 | 4.70 | 2.79 | -3.73 | -6.52 | 2.79 | -1.77 | -4.10 | 2.33 |

^a Assumed the same as I_i in the arithmetic, trimmed column.

* No observations in 1995.

SOURCES: Same as Table 7. Logarithmic A_i column is revised from Moulton and Moses (1997), as indicated in Table 7, and logarithmic I_i is a special tabulation for this paper by Karin Moses of BLS.

on average for the items that changed (Tables 7 and 8), no matter what method of quality adjustment is used. We know that prices don't always increase when product quality increases, because we know from much price index research that prices often fall when quality changes occur, particularly for high-tech products

and services. But much of the CPI is not high tech. What applies to electronic products, pharmaceuticals, communications services, and commodities that experience rapid rates of technological change does not, the data suggest, necessarily apply to shampoo, hair brushes, vacuum cleaners, wheelbarrows and a whole range of "low-

tech” products that are not challenging as research projects, but still make up much of the CPI.

Third, among CPI item replacements, those that are quality adjusted implicitly by the deletion method have lower quality-adjusted prices than item replacements that are handled in some other way. In Table 7, traditional deletions rose by only 0.34 percent after adjustment, measured by arithmetic means; price changes from other methods resulted in increases of from 2½ percent to 5 percent. The comparable logarithmic means yield increases of 0.25 percent, compared with a range of 2.3 percent to 4.1 percent. In short, price increases imputed from the deletion method look suspiciously low. Moreover, this outcome is a regularity, as shown in Armknecht and Weyback (1989); it is not unique to 1995.

Triplett (1971) showed that the deletion method can produce downward price index bias when quality is improving and prices are rising. Partly in response to the analysis in the literature, and partly because of internal analysis by CPI staff, the potential downward bias of the deletion method—the fact that it can miss part of the price increases that take place, incorrectly recording them implicitly as quality changes—has been of concern to BLS staff. Armknecht (1996) remarks:

At one time in the CPI the rule of thumb for assessing the quality content when substitutions occurred was ‘when in doubt, link it out.’ This practice resulted in some true price changes being removed as quality change.

The evidence across CPI components suggests that downward bias from deletions can be a serious problem in the CPI, but it is perhaps not compelling evidence. Gordon’s (1997) discussion of the CPI Commission’s report emphasizes research price indexes that are “independent” from the CPI: “The difference between these quality-adjusted independent price indexes and the corresponding CPI indexes . . . forms the basis of our [the Commission’s]

estimates of bias.” Is there independent evidence of downward CPI bias from deletion?

Two hedonic studies, both conducted by BLS staff, have shown downward bias in CPI components where deletions were involved. Apparel, studied by Liegey (1993), has long been known to have a downward bias because of deletions. Randolph (1988) estimated aging bias in the CPI rent index, which occurred because of imperceptible monthly deterioration in the rental units priced for the CPI. In itself, the aging bias was caused by direct comparisons of rents in the face of small monthly declines in quality as the rental unit deteriorated ($A_T < 0 = A_1$). The paper did not directly address the question of what happened in the CPI at the inevitable renovation. When the unit was renovated, returning the unit to its initial quality (or even upgrading it with, for example, better appliances and so forth), the renovated unit was treated as noncomparable, so the rent difference associated with renovation was deleted from the index. The price change for the deleted unit was imputed by I_3 (Equation 3, above). It is common knowledge that rental increases are more likely to occur when units are vacated and when they are renovated, a situation that implies that deletion misses price increases ($I_3 < I_T$). Although Randolph estimated the downward bias in the CPI that was created by monthly quality declines, he did not study deletions in the CPI rent index.

The CPI Commission, in a departure from its stated methodology, did not accept Randolph’s study, apparently because the Commission misinterpreted it. “Randolph (1988) estimates . . . aging bias . . . , a concept that represents the effect of depreciation net of any maintenance and renovation expenditures” (CPI Commission, 1996, page 42). Randolph’s estimates were not net of renovation because price changes associated with renovation were already deleted from the CPI. The CPI Commission substituted a series of back-of-the-envelope calculations leading to the conclusion that CPI rent was biased *upward*, not downward. Moulton and Moses (1997) show that the Commission’s calculations contain errors.

Randolph was not the only researcher to find downward bias in the CPI rent index, so it is puzzling why the Commission did not follow its usual practice in this case.

The CPI Commission maintained that the deletion method was upward biased. The following passages illustrate (all from CPI Commission, 1996): "To the extent that the deletion method is used, the CPI consists disproportionately of commodities of constant quality which may be further along in the product cycle" (p. 36). "This list of BLS methods reveals at least four potential sources of upward bias [including] the use of the deletion method that bases price change on models that are unchanged in quality and may be further along in the product cycle." And, of rental housing, "Alternative units are rotated in, with the overlap handled by deletion. If there is a general tendency for more recently constructed units to have more and better appliances, central air conditioning, and other amenities that were not present in previous decades, there is the possibility of an upward bias in the CPI". (In the latter case, the Commission may have been thinking of direct comparison, or method one.)

The Commission did not explicitly address or reference published literature that suggested downward bias from the deletion method, so it is not possible to characterize its position completely. However, if I can translate its language into the variables in Equation set 3, the Commission was saying that $l_3 > l_T$, so it must also have been saying that the prices P_j (the models that did not change) are always rising relative to the true value of l_T (see the passages quoted above), and the implicit quality adjustment (Equation 3a) is always too small. I believe it means to cite as evidence in this context (the language in the report is not always clear) research by individual members of the commission on personal computers and other electronic products, as well as pharmaceuticals.

When the new varieties or models are the vehicles for price reductions, and when the prices of the old do not fall fast enough to keep up, the Commission's view of the

direction of bias from deletion is correct: The deletion method misses price decreases and thereby biases the CPI upward. This expectation is shown by Equation 3b. Moreover, the examples from electronics, drugs, and some other high-tech products conform to the model the Commission apparently had in mind, and I fully concur that this research points undisputably to upward bias in these components. I would go further: Those high tech products and services have posed measurement problems in the past and are likely to pose problems at least as great in the future unless more resources are spent to gather data to estimate hedonic functions (a recommendation that was not, regrettably, among those of the Commission).

The issue is not whether the Commission's model is correct for the cases to which it applies and for the research on high-tech products and services the Commission cites. The issue is whether this is the *only* model, and whether this model of price behavior applies to the whole CPI.

After all, the model in which prices of existing products rise (or fail to fall) when confronted with competition from lower-priced new varieties must pertain to very special cases. In drugs, this phenomenon is caused by price discrimination, in computers it is caused by the speed of technical change and the characteristics of buyers who do not want to be on the technological frontier. It is perhaps exacerbated by list-transactions price problems in the research data base. Is there any evidence that prices of unchanged models of cars rise in the month when improved models are introduced? When one candy bar maker increases or decreases the size of the bar by a quarter ounce, thereby increasing or decreasing the price per ounce, do we expect that the others will always increase their prices, and by more? If not, the Commission's model of deletion goes in the wrong direction and does not describe the expected bias from application of the deletion method.

Ultimately, the prevalence in the empirical data of upward bias or downward bias from use of the deletion method can only be resolved by studies on individual

products and services. For this paper, an open mind on the direction of bias from the deletion method has three applications:

First, research has sometimes shown negative bias in CPI components; one cannot explain these studies if all CPI quality adjustment methods, including deletion, are upward biased, as the CPI Commission contends. Second, if traditional deletion is not downward biased (as the data reviewed above suggest), then the BLS development of the class-mean method must have been a blunder, a particularly pernicious one in view of the much greater price increase that it creates (refer to Tables 7 and 8). I believe that the class-mean method deserves further study, but I accept the importance and relevance of downward bias caused by traditional deletions; it was the reason for development of the class-mean method. Third, the empirical finding that the traditional deletion method has, on balance, been downward biased underlies the review of changes in CPI methods and the size of CPI biases over time that follows; I use that review to determine whether changes in measurement methods could have contributed to the consumption slowdown. If deletion cannot contribute to downward bias, then my review is built on a faulty premise.

Quality Changes Outside the CPI Sample

The standard CPI “matched model” pricing method involves (a) repeated pricing of the identical model or product variety until it is no longer available, combined with (b) application of the deletion method to shift over to a new model or variety when the old one exits the sample.

Studies on a number of technologically dynamic products have shown that the matched model method misses price decreases under certain not atypical circumstances: (1) New product varieties are introduced at a price-quality ratio below those of previously-existing varieties. (2) Prices of previously-existing varieties adjust slowly, if at all, to the new price/quality regime set by the newly-introduced varieties (in some cases, prices

of the old varieties actually rise). (3) conditions (1) and (2) imply that quantities of the new varieties rise rapidly and those of the old fall until the old disappear, perhaps without ever reaching a new market equilibrium between new and old price-quality ratios.

Product studies that have confirmed the matched-model method’s defects under these circumstances include studies of computers (Dulberger, 1989), of semiconductors (Dulberger, 1993), and of certain pharmaceuticals (Berndt, Griliches, and Rossett, 1993). In each of these cases, the researcher carefully replicated the CPI or PPI calculation procedure and demonstrated that the source of the price measurement error lies in late introduction of new varieties (which causes the CPI and the PPI to miss rapid price declines after introduction), combined with the use of the deletion method at the point where the BLS shifts pricing from the old product to the newer one. A number of the studies compared PPI, not CPI, indexes. However, the results probably extend to the CPI with respect to most electronic goods, including televisions (preliminary TV price indexes are reported in Gordon, 1997). Prices are falling for these products, new varieties are the vehicles for introducing new technologies, and similar market conditions apply. The PPI results probably apply also to “high-tech” services, such as telecommunications (Hausman, 1997), that use electronic equipment or components as inputs.

Three reasons exist for upward bias from these products with falling prices. First, even if the new model of an electronic product replaces an old one in the CPI at its point of introduction (it is thus inside the sample) one expects upward bias, because the expected CPI bias from the deletion method is upward when prices are falling. Consider the price imputation of Equation 3b: Because the deletion method misses the downward price change accompanying introduction of the model, this method will bias the CPI upward. Thus, the deletion method’s bias for quality changes inside the CPI sample is completely symmetric: When quality is improving, rising prices create downward bias and

falling prices create upward bias. The deletion method's bias is more nearly a function of the sign of the true price change, l_T , than of the quality change.

Some weak evidence for the deletion method's symmetry comes from Table 8. Many consumer electronic goods are in the "entertainment" category. The quality-adjusted price change for traditional deletions in this commodity category (1.33 percent, arithmetic untrimmed; 1.04 percent, logarithmic) is the largest for any category in Table 8. I speculate that it would be larger still if electronic goods were split off from other goods in this category.

A second source of upward bias occurs because the new model may not appear in the CPI sample for some time after its introduction. Twenty percent of the CPI's probability sample is replenished each year, so new products inevitably lag in entering the sample, and in some documented cases new products have apparently fallen between the lines of the CPI product classification system. The CPI thus misses price decline if the prices of new products that are not in the sample decline relative to the prices of their counterparts in the sample. Though one might think that the prices of obsolescent goods should decline relative to advanced ones, existing price index literature for high-tech products suggests that the older products often become more expensive.

A third source of bias is the welfare gain at the point of introduction of the new model, which the CPI also misses. This source of bias is discussed in the next section.

New Products

The CPI Commission considered welfare gains from the introduction of new products in a cost-of-living index. Suppose a new product (or a new variety of an existing product) is introduced at time 2, at price P_2 . If P_{r1i} is the reservation price at which demand for the new product is zero in period 1 for consumer i , then the ratio P_2/P_{r1i} appears in consumer i 's cost-of-living index. This is obviously a declining price. Aggregating over all consumers gives the familiar consumer surplus triangle.

Hausman (1997), in a widely cited and path-breaking study, estimated the welfare gain from the introduction of apple-cinnamon Cheerios, and concluded that introduction of this one new cereal would have reduced a cost-of-living index for cereals by about 1.5 percent to 1.7 percent.

No adjustments for consumer surplus are currently made in the CPI, nor does consumer surplus contribute to real per capita consumption, as it is now measured. One expects upward price index bias from neglecting new products,²⁸ and therefore downward bias in PCE.

There is too little empirical basis to allow estimating the magnitude of new-product effects on the overall CPI or on real PCE. Existing estimates appear exaggerated. First, the discussant for Hausman's (1997) study showed that Hausman's welfare gain estimate was upward biased (a finding that implies that his cereals price index was downward biased). Second, the CPI Commission's "guesstimates" of consumer surplus (for example, 20 percent for increased variety of fresh fruits and vegetables, 5 percent for microbreweries and an increased variety of imported wine, 10 percent for new products among draperies, furniture, and soap) also appear upward biased, judging from a variety of back-of-the-envelope calculations carried out in response to the Commission's report. On the other hand, Gordon (1997) points out that the Commission did not try to include consumer surplus calculations across all consumption categories, so overstatements in the categories for which it did consider consumer surplus may be offset by ignored surplus in categories where surplus was not estimated.

This is an important topic for future research. Stimulating that research may be one of the lasting contributions of the CPI Commission, similar to the contribution of the Stigler Committee report, which stimulated the empirical estimates of commodity substitution bias that were available to the Boskin Commission.

With respect to the post-1973 consumption slowdown, consumer surplus from new products would affect the slowdown

²⁸One should, of course, net out product disappearances, which might disproportionately affect some population groups, particularly the poor and the elderly.

only if their unmeasured effects were greater after 1973 than before. Though one often hears assertions that the pace of introduction of new products has accelerated, there are no data to confirm such an assertion, which needs historical perspective. It seems unlikely that the welfare gain from another variety of Cheerios is greater than the welfare gain from the introduction of Cheerios itself, which occurred well before 1973. Gordon (1994) remarks that it is implausible that the proliferation of electronic products in the 1980s and 1990s could make greater changes to living standards than were made by the automobile, major appliances, telephones, and so forth in the decades before 1973. It is also remarkable that a major article on quality errors in price indexes published 30 years ago (Nicholson, 1967) contains a list of new products—including new drugs, improved television sets, coffee bars, and increased availability of fine wines, and of fresh fruits and vegetables in the winter—that parallels the CPI Commission's (1996) list.

Although unmeasured consumer surplus biases PCE downward, no one knows whether the rate of introduction of new products increased after 1973, and whether consumer surplus accumulated at a faster rate. The effect on the slowdown could be either positive or negative. Because we do not know, I assign the effect a provisional value of zero.

The Wrap-up: What Can Be Said About Deflation Bias and the Real Consumption Slowdown?

To summarize the discussion so far:

- Substitution bias cannot be a factor in the post-1973 consumption slowdown because real consumption is measured with a superlative index number both before and after 1973.
- Outlet substitution bias is too small (the CPI Commission estimated it at 0.1) to account for any substantial part of the post-1973 consumption slowdown.
- Basic component bias existed in a different form before 1978, so it can make little net contribution to the post-1973 consumption slowdown, even though it might affect growth rates both before and after 1973.
- New products (consumer surplus) bias might contribute to the measured deceleration in the growth rate of real per capita consumption after 1973, provided that the rate of new product introductions was more rapid after 1973 and that they contributed greater consumer surplus than the new products of earlier days. But little information exists.
- Assessing the contribution of quality change bias to the slowdown is more complicated.

For quality bias to account for part of the post-1973 consumption slowdown, the CPI must have been biased upward more severely because of quality change after 1973 than before. Is there any evidence for the conjecture of increasing quality change bias? I consider whether quality adjustment methods in the CPI changed in a way that is consistent with greater upward bias after 1973.

1959-1973 Period. Evidence from the price index literature, BLS documentation, and (more compellingly) anecdotes all suggest that quality change in the CPI received more attention after around 1960 than in the 1950s. A major new formal method for making quality change adjustments—manufacturers' data on the cost of quality changes on new cars—was initiated in 1959-1960, and the quality problem also assumed more importance at the operational level in BLS after 1960.

Studies in 1965 and 1966 (summarized in Triplett, 1971) reported that about 7 percent and about 10 percent of CPI price quotations involved item replacements, compared with the 4 percent rate reported by Moulton and Moses (1997) 30 years later (see Table 6). Roughly three-fifths of

the item replacements were handled by direct comparison in the 1960s (about the same as in 1995). Deletions accounted for a third or more of the cases (higher than in 1995), and all were traditional deletions because the class-mean method had not been invented. Direct quality adjustments in the 1960's were very rare (omitting size adjustments for food packaging, around 2-3 percent of item replacements). There were probably changes in procedures during the 1959-73 period. For example, anecdotes in the late 1960s reported that auto manufacturers' cost data was initially subject to less stringent review than it was later on. However, I take the 1959-73 period as the base for considering the period following 1973.

We do not know what the CPI quality bias was before 1973. In the 1960s, economists frequently guesstimated CPI bias at around 3 percent per year (three times the CPI Commission's estimate for the 1990s), with quality change accounting for much of the bias, but the 3 percent number rested on very little evidence. Because it is sometimes easier to estimate changes than levels, that is the approach I take in the following. Whatever the quality bias in the CPI was before 1973, do changes in CPI procedures after 1973 move that unknown bias in an upward or downward direction?

1973-1983 Compared with Pre-1973. Though the CPI revision of 1978 was undoubtedly the most comprehensive and far-reaching ever, little in it was directly concerned with improving CPI quality adjustments. However, the introduction of in-store probability sampling of items indirectly (and unexpectedly) reduced the incidence of item replacement in the CPI from between 7 percent and 10 percent to 4 percent (see Moulton and Moses, 1997, Table 4). This also reduced the incidence of in-sample quality change in the CPI, which (other things equal) would have reduced the quality bias, whatever it was, relative to pre-1973. Additionally, a new CPI sample rotation methodology was installed (a new probability sample of items and outlets was drawn every three

years, later stretched out to five years); sample rotation on a probability basis reduced the amount of missed out-of-sample quality changes because sample replenishment brought new varieties of products into the CPI faster than had been the case before 1978.

With respect to within-sample quality adjustments, somewhere along the line there was a gradual shift away from direct comparisons toward deletions, and, to a lesser degree, direct quality adjustments. Recall from the earlier discussion that the expected bias from direct comparison is upward and from traditional deletion is downward, when quality is improving and prices are rising, and prices were certainly rising in the 1970s. These changes should have tipped the quality error in the CPI in the downward direction. Contributing additionally to downward quality bias in the CPI was the BLS decision to treat regulatory changes in automobiles (smog devices and to a lesser extent, perhaps, mandated safety equipment) as quality improvements in automotive transportation, rather than—the correct treatment—as a tax on transportation levied to support cleaner air.²⁹ This change had major effects that began with expanded regulation about 1973. A partial and small offset to this shift toward downward bias in the new car indexes was more vigorous BLS examination during the 1970s of the cost data for automobiles, which attenuated, perhaps, some of the downward bias from overadjustment.

Right at the end of this period, the BLS changed the way it computed the cost of owner-occupied housing. Although this was a conceptual issue, not primarily one of quality change, it removed a substantial upward bias that was present in the 1960s and, especially, the late 1970s.

Gordon (1990) found that the difference between his "alternative" price indexes and BLS indexes (PPI, as well as CPI) was smaller in the last years covered in his book than it was in the earlier period. His finding is consistent with smaller quality error in the CPI after 1973, though that may not be the only explanation. In another part of his study, Gordon recomputed real PCE

²⁹The CPI Commission (1996) also pointed to this downward bias.

durables using his alternative price indexes (Gordon, 1990, table 12.11, page 553). Growth rates both before and after 1973 are higher when the alternative indexes are used as the deflators, but the post-1973 consumption slowdown is about the same.³⁰

I judge that CPI quality bias moved in the downward direction in the 1973-83 period, compared with the bias before 1973. This conclusion implies a shift toward upward bias in the PCE after 1973, the wrong direction to explain the post-1973 consumption slowdown. The main argument going the other way is the speculation that the relatively modest improvements in methods did not keep up with the difficulty of the problem. An ever larger share of consumption went into components that are difficult to measure, such as electronics (where studies show upward CPI quality bias) and services.

1983-1987 Compared with 1973-83. I know of no major changes to CPI quality adjustment procedures in this interval. The 1978 CPI revision was innovative, but late; the 1987 one came in on schedule, but without significant improvements. For lack of anything better, I presume that whatever quality error existed before 1983 also extended to 1987.

Post-1987 Period. In Triplett (1988), I suggested that the CPI was downward biased by quality change because research had shown that three major CPI components—clothing, new cars, and housing—were downward biased because of overadjustment for quality change. I still believe that was correct, in 1988. These three downward-biased components accounted for nearly half of the CPI's weight. In contrast, the CPI Commission noted that electronic products, where research has shown strong upward bias, account for only around 2 percent of the weight.

But around 1987-88, the BLS began to correct three sources of downward bias in CPI quality adjustment procedures. First, the BLS eliminated the downward bias in rental housing documented by Randolph

(1988), and also by other researchers. Second, the BLS also made changes to reduce the downward bias in the clothing indexes that emerged from Liegey's (1993) study, but which was discussed extensively even earlier (see Triplett, 1988). Both these downward CPI quality biases existed in the pre-1973 period. Third, the pendulum began to swing away from the traditional deletion method for handling quality change, which had a serious downward bias. BLS introduced the "class-mean" procedure, which, the evidence in Tables 7 and 8 shows, generally leads to more quality-adjusted price increase. Additionally, the CPI now uses more direct comparison than formerly—a change which one would expect to contribute upward bias. Moulton and Moses (1997) maintain that these are better direct comparisons than they used to be, so the expected increase in upward bias might be small. The BLS also introduced quality adjustments into the used car index, which was seriously biased upward.

The three major changes that eliminated downward biases affected a large proportion of the CPI. Thus, correcting these downward biases increased the measured rate of inflation, compared with the less accurate pre-1987 index—they moved the unknown quality bias upward. As already noted, the changes to rent and to clothing also increased the post-1987 measured rate of inflation, relative to pre-1973. Conversely, the used car changes lowered the inflation rate, post-1987 relative to earlier. The net effect of the class-mean change on the pre-and post-1973 measures is harder to determine.

Removing elements of downward quality bias in the CPI has, on balance, probably moved the overall quality bias in an upward direction after 1987. This is consistent with the CPI Commission's assertion that the current quality bias in the CPI is greater in a positive direction than it once was, though for somewhat different reasons than the Commission gave. Additionally, out-of-sample quality change might have increased after 1987, though there is no direct evidence: "The growing importance of such hard-to-measure categories as consumer electronics and medical

³⁰Gordon's data extend to 1983. Using his table 12.11, page 553, growth rates (not per capita) in consumer durables are: 1959-73, 7.4 percent; 1973-83, 3.6 percent. In comparison, per capita growth rates (this paper, table 2) are: 1959-73, 5.3 percent; 1973-83, 1.5 percent. Per capita growth rates are of course always lower.

services may have increased the significance of quality change bias in the past decade” (CPI Commission, 1996, p. 32).

Overall, Pre-and Post-1973. Considering changes in CPI procedures, it seems unlikely that increasing upward quality error in the CPI accounts for the post-1973 consumption slowdown. The slowdown was too abrupt, and too large, to have been a statistical illusion created by changes in the way quality change is handled in components of the CPI.

Additionally, the only abrupt change around 1973 was the BLS treatment of smog devices in automobiles which, because it creates a downward bias in the CPI, goes in the wrong direction to explain the slowdown. Other changes in the CPI between 1973 and around 1987 also appear to have moved CPI bias downward. A possible change in the offsetting direction is increased out-of-sample quality change caused by increasing complexity of new products and by shifts in the consumer share of more complex products and of hard to measure services. But even this change is attenuated to a degree because the CPI after 1978 was better in its coverage of new products than it was before. Taking these CPI changes together, they would have increased the PCE rate of growth after 1973, not slowed it.

Correction of downward CPI biases after 1987 means that overall quality bias in the current CPI has moved in the positive direction, which would reduce PCE growth, post-1987. But this reduction would not be sufficient to account for the post-1973 consumption slowdown, because it is a post-1973, not a post-1987, phenomenon.

CONCLUSIONS

I began by asking whether there was evidence that the post-1973 slowdown in the rate of growth of U.S. real per capita consumption was a statistical illusion created in whole or in part by measurement error. The paper has taken the form of a long search for clues.

The absence of evidence for the statistical illusion story does not mean that U.S.

consumption data are without flaws. Indeed, economists should have serious concern about the adequacy of consumption data. Potential problems with current-dollar consumption data were discussed. There can be no question that errors exist in the procedures for measuring the price indexes that are used for deflation.

But part of the evidence on the current-dollar data (examining alternative data sources on consumer expenditures) goes in the wrong direction for the statistical illusion conjecture because it suggests that the growth of real per capita consumption may have been overstated. Part of the evidence on the deflators also goes in the wrong direction, offsetting, at least in part, the perhaps more widely understood upward bias problems in the CPI (such as substitution bias, or the measurement of health care costs).

Do the data contain error? Clearly, they do. Are the data errors preponderantly in the same direction so we can say with a high degree of certainty that consumption growth is understated? The evidence is not sufficiently extensive to support a fully informed judgment. It is probably true that the net effect of all deflation measurement errors creates downward bias in real consumption, and the non-measured part of consumption growth (that is, consumer surplus) also causes understatement.

Even if the errors that exist did go in the same direction on balance, so that measurement errors cause understatement of the growth of real consumption, the statistical illusion hypothesis would still require a positive answer to a third question: Do errors intensify after 1973 so that they could account for the abrupt slowdown in the upward course of U.S. living standard growth? This is the most doubtful proposition of all.

REFERENCES

Aizcorbe, Ana M., and Patrick C. Jackman. “The Commodity Substitution Effect in CPI Data, 1982-91,” *Monthly Labor Review* (December 1993), pp. 25-33.

Armknrecht, Paul A. “Improving the Efficiency of the U.S. CPI,” *International Monetary Fund Working Paper*, September 1996.

- _____ and Donald Weyback. "Adjustments for Quality Change in the U.S. Consumer Price Index," *Journal of Official Statistics* (1989:2), pp. 107-23.
- Berndt, Ernst R., Zvi Griliches, and Joshua G. Rosett, "Auditing the Producer Price Index: Micro Evidence from Prescription Pharmaceutical Preparations," *Journal of Business and Economic Statistics* (July 1993), pp. 251-64.
- Braithwait, Steven D. "The Substitution Bias of the Laspeyres Price Index: An Analysis Using Estimated Cost-of-Living Indexes," *American Economic Review* (March 1980), pp. 64-77.
- Branch, E. Raphael. "The Consumer Expenditure Survey: A Comparative Analysis," *Monthly Labor Review* (December 1994), pp. 47-55.
- CPI Commission—Advisory Commission to Study the Consumer Price Index. *Toward a More Accurate Measure of the Cost of Living: Final Report to the Senate Finance Committee*, 1996.
- Diewert, W. Erwin. "Exact and Superlative Index Numbers," *Journal of Econometrics* (May 1976), pp. 115-45.
- Duff, Christina. "Is the CPI Accurate?" *Wall Street Journal*, January 16, 1997, p. A1.
- Dulberger, Ellen R. "The Application of a Hedonic Model to a Quality-Adjusted Price Index for Computer Processors," *Technology and Capital Formation*, Dale W. Jorgenson and Ralph Landau, eds., MIT Press, 1989, pp. 37-75.
- _____. "Sources of Price Decline in Computer Processors: Selected Electronic Components," *Price Measurements and Their Uses*, Murray F. Foss, Marilyn E. Manser, and Allen Young, eds., Conference on Research in Income and Wealth, Studies in Income and Wealth, Vol. 57, University of Chicago Press for the National Bureau of Economic Research, 1993, pp. 103-24.
- Executive Office of the President, Office of Management and Budget. *Standard Industrial Classification Manual—1987*, GPO, 1987.
- Gordon, Robert J. *The Measurement of Durable Goods Prices*, University of Chicago Press for the National Bureau of Economic Research, 1990, pp. 490-515.
- _____. "Comment," *Brookings Papers on Economic Activity* (No. 2, 1994), pp. 324-30.
- _____. "Comment," *Brookings Papers on Economic Activity* (No. 1, 1997), pp. 350-66.
- Griliches, Zvi. "Productivity, R&D, and the Data Constraint," *American Economic Review* (March 1994), pp. 1-23.
- Hausman, Jerry A. "Valuation of New Goods Under Perfect and Imperfect Competition," *The Economics of New Goods*, Timothy Bresnahan and Robert Gordon, eds., Conference on Research in Income and Wealth, Studies in Income and Wealth, Vol. 58, University of Chicago Press for the National Bureau of Economic Research, 1997, pp. 209-37.
- Levy, Frank, and Richard J. Murnane. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature* (September 1992), pp. 1333-81.
- Liegey, Paul R. Jr. "Adjusting Apparel Indexes in the Consumer Price Index for Quality Differences," *Price Measurements and Their Uses*, Murray F. Foss, Marilyn E. Manser, and Allen Young, eds., Conference on Research in Income and Wealth, Studies in Income and Wealth, Vol. 57, University of Chicago Press for the National Bureau of Economic Research, 1993, pp. 209-26.
- Manser, Marilyn E., and Richard J. McDonald. "An Analysis of Substitution Bias in Measuring Inflation, 1959-85," *Econometrica* (July 1988), pp. 909-30.
- Moulton, Brent R., and Karin E. Moses. "Addressing the Quality Change Issue in the Consumer Price Index," *Brookings Papers on Economic Activity* (No. 1, 1997), pp. 305-49.
- Nicholson, J.L. "The Measurement of Quality Changes," *The Economic Journal* (September 1967), pp. 512-30.
- _____. "Measurement Issues in the Consumer Price Index," Bureau of Labor Statistics paper prepared for the Chairman of the Joint Economic Committee, June 1977.
- Price Statistics Review Committee. *The Price Statistics of the Federal Government*, U.S. Congress, Joint Economic Committee, Government Price Statistics Hearings, Part 1, 87th Congress, 1st Session, January 24, 1961; also published as: National Bureau of Economic Research, General Series, Number 73.
- Randolph, William C. "Housing Depreciation and Aging Bias in the Consumer Price Index," *Journal of Business and Economic Statistics* (July 1988), pp. 359-71.
- Reinsdorf, Marshall. "The Effect of Outlet Price Differentials on the U.S. Consumer Price Index," *Price Measurements and Their Uses*, Murray F. Foss, Marilyn E. Manser, and Allen Young, eds., Conference on Research in Income and Wealth, Studies in Income and Wealth, Vol. 57, University of Chicago Press for the National Bureau of Economic Research, 1993, pp. 227-60.
- Slesnick, Daniel T. "Aggregate Consumption and Saving in the Postwar United States," *The Review of Economics and Statistics* (November 1992), pp. 585-97.
- "Improved Estimates of the National Income and Product Accounts for 1959-95: Results of the Comprehensive Revision," *Survey of Current Business* (January/February 1996), pp. 1-31.
- Triplet, Jack E. "Quality Bias in Price Indexes and New Methods of Quality Measurement," *Price Indexes and Quality Change: Studies in New Methods of Measurement*, Zvi Griliches, ed., Harvard University Press, 1971, pp. 190-4.
- _____. "Price Index Research and Its Influence on Data: A Historical Review," paper presented at National Bureau of Economic Research, Conference on Research in Income and Wealth, Fiftieth Anniversary Conference, Washington, DC, May 1988.

_____. "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy," *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth*, Ernst R. Berndt and Jack E. Triplett, eds., Studies in Income and Wealth, Vol. 54, University of Chicago Press for the National Bureau of Economic Research, 1990, pp. 207-33.

_____. "High Tech Industry Productivity and Hedonic Price Indexes," *Industry Productivity: International Comparison and Measurement Issues*, OECD Proceedings, Organization for Economic Cooperation and Development, Paris, 1996.

Wynne, Mark A., and Fiona D. Sigalla. "The Consumer Price Index," *Economic Review* Federal Reserve Bank of Dallas (1994:2), pp. 1-22.

U.S. Department of Commerce, Bureau of Economic Analysis. *Personal Consumption Expenditures*, Methodology Paper Series MP-6, U.S. Government Printing Office, June 1990.

U.S. Department of Labor. *BLS Handbook of Methods*, Bureau of Labor Statistics, Bulletin 2414, 1992.