The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work

By DAVID W. WILCOX*

This paper investigates the sources and methods used to construct the aggregate data on consumer spending in the United States, searching especially for imperfections that may have implications for the outcome of empirical work. The paper identifies two such imperfections: sampling error and compositional error. It then presents several examples intended to illustrate that these imperfections may be empirically important and that appropriate remedies for them often can be devised. The paper concludes by suggesting some guidelines for empirical practice. (*JEL* E21, C82)

When researchers test and reject an implication of a theoretical model, they usually assume that the model is in error and that subsequent investigation should be directed toward the development of alternative models that might better account for the observed characteristics of the data. They usually spend little effort investigating the characteristics of the data themselves or the suitability of the data for use in the application at hand. This paper reverses these priorities and investigates the source data and estimation methods used to construct the retail-sales and aggregate consumption data in the United States, searching especially for imperfections that might have implications for the outcome of empirical work.

Section I provides a brief introduction to the sources and methods underlying the construction of the retail-sales and personal-consumption-expenditures (PCE) data in the United States and identifies two such imperfections. The first is that retail sales is measured with error. Part of this error is sampling variation and simply reflects that the monthly retail-sales estimates are derived from a sample rather than a complete enumeration of all stores. The remainder of the measurement error is nonsampling error and stems from such problems as reporting errors, coding errors, definitional difficulties, and imperfections in imputation procedures. Although relatively little can be said about the statistical properties of the nonsampling error, quite a bit is known about the sampling error. Unfortunately, measurement error in retail sales feeds into the estimates of PCE, because the retail-sales data comprise an important building block for the consumption estimates.

A second imperfection arises because the product composition of retail sales is not known at the monthly frequency. Thus, to take one example, the Census Bureau does not receive monthly reports on sales of apparel. Instead, it receives monthly reports on sales at various types of stores, such as apparel stores and department stores, which sell apparel. Strictly speaking, the only source of direct information about the product composition of sales by type of store is the Census of Retail Trade, which is con-

*Federal Reserve Board, Washington DC 20551.
An earlier version of this paper carried the title "What Do We Know About Consumption?" The views expressed are those of the author and not necessarily those of the Board of Governors of the Federal Reserve System or other members of its staff. I thank members of the staff at the Federal Reserve Board—especially William Cleveland—for helpful discussions, employees of the Bureau of Economic Analysis and of the Census Bureau—especially William Bell—for much advice, seminar participants at the University of Virginia and especially two anonymous referees for helpful comments on an earlier draft, and Maura Doyle, Jo-Ann Holden, and Stephen Helwig for excellent research assistance. Remaining errors are my own responsibility.

ducted once every five years. Monthly estimates of consumer spending by type of product are constructed under the assumption that the composition of sales within each category of stores is fixed from month to month.

The heart of the paper is composed of several examples intended to illustrate that these imperfections may be empirically important and, equally worth emphasizing, that appropriate remedies for them often can be devised. Section II focuses on sampling error and shows that it influences the data in two important ways. First and most obviously, sampling error adds noise to the data. In the context of many standard theoretical models of consumption, such noise, if not properly controlled for, would be interpreted as signaling greater uncertainty in the economic environment than is actually present. Noise also inappropriately reduces the coherence between series. The paper presents an example in which the objective is to estimate the relative importance of common and idiosyncratic shocks in city-level retail sales and shows that proper adjustment for the sampling error in those data materially changes the empirical conclusions.

Sampling error also distorts the autocorrelations of the data. Of course, this would be true at the first lag even if the sampling error were not serially correlated. The sampling error in fact turns out to be highly serially correlated, partly for reasons related to the design of the Retail Trade Survey. As a result, even autocorrelations at other than the first lag are influenced by the presence of the sampling error. The paper presents a second example in which the objective is to infer the depreciation rate of a composite consumption good from the autocorrelation of spending at the first lag. In general, sampling error drives a wedge between the economically meaningful autocorrelation and the observed autocorrelation and so causes the estimated depreciation rate to differ from the true one. Interestingly, the extent of this distortion depends on the time-series properties of the true (but unobserved) series. A third example considers the possible implications of sampling-error-induced autocorrelation in the sales data for tests of the random-walk hypothesis. Such tests can be structured so as to ignore possible spurious autocorrelation at the first lag; even if the test is conducted in this manner, however, the autocorrelation of the sampling error at higher lags may come into play.

Section III focuses on the product-composition problem and notes that the main effect of current methodology is to exaggerate the coherence of the various components of PCE. This, too, may cause incorrect inferences to be drawn from the published data. For instance, one could observe the relatively high coherence between various detailed categories of PCE and conclude that taste shocks play a relatively small role in determining the allocation of consumption among types of goods; a look at the retail-sales data, however, might yield a substantially different conclusion. A fourth example presented in the paper examines the behavior of outlays for apparel and furniture—the key pair of goods as far as the product-composition issue is concerned—and presents circumstantial evidence that the PCE data overstate the coherence between these two categories of spending by a fairly wide margin.

The methodologies of the Census Bureau and the Bureau of Economic Analysis (BEA) are carefully designed and reflect both a foundation in statistical theory and decades of experience. Nevertheless, the retail-sales and PCE estimates are not perfect. Accordingly, this paper is intended primarily as a cautionary note to empirical researchers: published data should not automatically be assumed to correspond exactly to their theoretical analogues. The paper may draw attention to areas in which the data could be improved; however, its more immediate purpose is to characterize existing imperfections in the data, explore the circumstances in which such imperfections might have implications for empirical work, and suggest strategies for ensuring that empirical work is robust to data imperfections.

It is worth emphasizing that the analysis in this paper focuses on the problems involved in measuring only a subset of total consumption, namely, outlays for goods other than motor vehicles. It should not be inferred from this focus that the measurement problems in other areas of PCE (or GNP, for that matter) are less important. Future work could usefully examine the characteristics of the data from these other areas.

I. The Estimation of Personal Consumption Expenditures and Retail Sales

The National Income and Product Accounts are constructed by the Bureau of Economic Analysis (BEA) in the Department of Commerce. BEA constructs the accounts using data from a multitude of sources including the Census Bureau, the Departments of Agriculture, Energy, Labor, and Transportation, the Social Security Administration, the Internal Revenue Service, the unemployment-insurance system, the Motor Vehicle Manufacturers Association, and the Edison Electric Institute, to name a few.[1] PCE accounts for roughly two-thirds of total GNP. PCE consists of spending in three major categories: durable goods, nondurable goods, and services. These categories make up about one-sixth, one-third, and one-half of PCE, respectively. The information for the estimation of PCE comes from a collection of sources no less diverse than the one underlying the GNP estimates. Of these sources, probably the single most important one is the monthly Retail Trade Survey, from which the estimates of retail sales are derived. BEA uses the retail-sales estimates to construct a portion of PCE that is known as "PCE control." Roughly speaking, PCE control consists of spending on durables excluding motor vehicles, plus nondurables.[2] This paper focuses on issues related to the construction of PCE control and the characteristics of the underlying retail-sales data.

The retail-sales estimates are compiled by the Bureau of the Census (also in the Department of Commerce) and are released monthly, both seasonally adjusted and not seasonally adjusted. The first estimate of retail sales for any given month, known as the "advance estimate," is released roughly two weeks after the close of the month in question. The second, or "preliminary," estimate is released one month later, and the third, or "final," estimate is released one month after that.[3]

The retail-sales figures are prepared from a probability sample that has three components.[4] The first component consists of companies that had sales in 1982 greater than a specified cutoff amount (the cutoff amount varied by kind of business). These large companies were included in the sample with probability 1 and hence are referred to as "certainty cases." In this component of the sample, the company was

[1]See Carol S. Carson (1987) for a useful introduction to the estimation of GNP.

[2]Certain pieces of retail sales are not mapped into PCE. For example, most sales at building material and supply stores are ignored because the items sold at these stores mainly relate to home-ownership, and such spending is captured in the residential-investment category. Also, most retail sales at automotive dealers are ignored because spending on autos is estimated using data from other sources (notably *Ward's Automotive Repairs*). Prior to 1987, BEA also was stripping out most of retail sales at gasoline stations; now, however, the gasoline-station component of retail sales effectively is included in the control category, owing to a change in BEA's methodology for estimating nongasoline sales at gasoline stations.

[3]Ultimately, even the so-called final estimates are subject to further revision at the time of the annual benchmark revisions (usually in March); the benchmark revisions take into account new information from the Annual Retail Survey and the quinquennial Census of Retail Trade (when available). In addition, other adjustments sometimes are introduced along with the benchmark revisions. For example, the revised figures released in March 1988 incorporated new information (derived from examination of payroll registries) on business deaths. The new information showed that some firms originally thought to have been active at the time of the introduction of the current monthly sample (in January 1987) in fact had already gone out of business; the revised figures showed faster growth for 1987 than previously had been estimated.

[4]This section draws heavily on the material contained in the appendixes to the Census Bureau's *Monthly Retail Trade Report*. The discussion in the text does not include a detailed description of the methodology for the advance estimates, because they play no role in determining the behavior of the historical time series.

taken as the sampling unit. That is, data are solicited for the company as a whole, although multiestablishment companies might be asked to provide data for each of their establishments separately as an aid to the estimation of regional and kind-of-business components of total retail sales. The certainty cases are asked to report on sales every month. Roughly 2,000 companies were selected as certainty cases in the 1987 sample redraw; these companies account for about 40 percent of total retail sales.

Companies whose sales in 1982 fell below the certainty cutoff were placed in a separate group. In this group, the employer identification (EI) number was taken as the sampling unit.[5] The EI numbers were stratified by major kind of business and estimated sales in 1982, and within each stratum a simple random sample of EI's was drawn. The sampling rates in the different strata ranged from one in three to one in 1,120. Roughly 30,000 noncertainty cases were selected into the sample in 1987.[6] The sampling units in this component of the sample are divided into three panels. Each panel is asked to report once every three months; each report is intended to consist of data for two consecutive months. Prior to September 1977, the Census Bureau used a four-panel rotation.

The first two components of the sample account for about 94 percent of retail sales. The other 6 percent is accounted for by the "area sample," which is a probability sample of land segments and is intended to pick up businesses not represented in the first two components, mainly recent EI births. Because of its limited importance, the area sample is ignored for the rest of this paper.

Each month, the information from the reporting panel's responses is combined with the data from the certainty cases to create two "unbiased" estimates: one of sales in the "current" month and one of sales in the "previous" month. These unbiased estimates are constructed by weighting up the reported sales of the individual sampling units. (Each weight is calculated as the multiplicative inverse of the sampling rate that was applied to the stratum from which the sampling unit was drawn; certainty cases receive a weight of unity.) The unbiased estimates are then used to construct the preliminary and final estimates for the current and previous months, respectively. The method by which this is accomplished is the "composite" estimation method; the composite method will be discussed in detail below.[7]

There are two general sources of measurement error in the unbiased estimates: sampling error and nonsampling error. Sampling error results from the fact that the monthly sales estimates are based upon a sample of stores rather than a complete enumeration of all retail establishments. One common measure of sampling error is known as the coefficient of variation (CV), which is defined as the standard deviation of the estimate (over all possible samples of a given size drawn from the total population according to a given sample design) divided by the quantity being estimated (see Kirk M. Wolter [1985] for a discussion of CV's). Each month, the Census Bureau estimates CV's by type of store for both levels and growth rates of sales; the results are reported in the *Monthly Retail Trade Report*. In general, the estimated CV's are smaller for the more aggregated store categories, and they are also smaller for growth rates

[5] Employer identification numbers are issued to employers who make Social Security payments for their employees under the Federal Insurance Contributions Act. A single company can have more than one employer identification number.

[6] Collectively, the first two components of the sample are referred to as the "list sample," because they are drawn from the Census Bureau's Standard Statistical Establishment List. The probability sample is redrawn every five years, taking account of the results of each new Census of Retail Trade.

[7] The unbiased estimates are not publicly available. Although it would be possible, in principle, to recover the unbiased estimates from the time series of preliminary and final estimates, the Census Bureau advises that users not do so because corrections to the preliminary estimate often are incorporated into the final estimates; failure to account for these corrections would make the recovery of the unbiased estimates inaccurate.

TABLE 1—COEFFICIENTS OF VARIATION FOR SELECTED COMPONENTS OF RETAIL SALES (PERCENT)

Category of sales	Final composite (level)	Ratio of consecutive months (growth rate)
Total	0.7	0.3
Durable-goods stores	1.2	0.8
Automotive dealers	1.3	1.1
Nondurable-goods stores	0.6	0.2
Department stores	0.0	0.0
Food stores	1.3	0.3
Apparel stores	1.9	0.9
Women's	2.9	1.6
Men's and boys'	4.2	1.6
Family	3.2	1.4
Shoe stores	3.1	2.0

Note: The ratio of consecutive months is calculated as the ratio of the current-month preliminary estimate to the previous-month final estimate.

Source: Bureau of the Census, *Monthly Retail Trade Report*, May 1985. The figures reported in the text are the medians of the CV's estimated over the six-month period running from August 1984 through January 1985.

than for levels. The latter reduction in error results from the fact that the sampling errors turn out to be highly autocorrelated and, therefore, tend to cancel in monthly changes.

Table 1 presents the median of the CV's estimated for the six months running from August 1984 through January 1985 for selected categories of retail sales. The table shows that the median CV for total retail sales was 0.7 percent in the level of the final estimate and 0.3 percent in the growth rate. (Note that the latter CV is applied to the gross growth rate; thus, a two-CV interval bracketing an estimate of 0.5-percent growth in total retail sales would extend from -0.1 percent to 1.1 percent.) The estimated CV for growth in total spending at apparel stores was 0.9 percent; for several of the component categories of apparel sales the estimated CV's were substantially larger. By way of comparison, the mean absolute deviation in the percentage change in total retail sales for the period 1967–1989 was about 1.0 percent; for apparel stores the comparable figure is 1.6 percent.[8] The sampling error for department stores is shown as zero, reflecting that all retail establishments meeting the definition of a department store are selected into the sample.

Nonsampling error results from a variety of circumstances, including errors in reporting and coding, misunderstanding of definitions, and failures to report.[9] Of these problems, nonresponse appears to be the most serious. Participation in the monthly survey is voluntary, and response rates have varied over time. Joseph K. Garrett et al. (1987) report that, as of August 1987, nonrespondents were contributing roughly 25 percent of total retail sales by value; the February 1988 *Monthly Retail Trade Report* puts the figure at "about 20 percent." Both figures represent a significant deterioration from years past; Wolter et al. (1976) estimate the nonresponse rate at only 9 percent. Participation in the annual survey is "mandatory," and response rates typically have been

[8]Note that some of the variation in the observed retail-sales series reflects sampling error; a more informative comparison would be between the variation in true (but unobserved) retail sales and the CV of the observed series.

[9]Preston Jay Waite (1974) investigates the sources and magnitudes of nonsampling error.

somewhat higher than in the monthly survey.[10]

Some nonsampling errors, such as errors in reporting, sometimes are corrected eventually, but others likely reflect fundamental inadequacies (relative to the reporting requirements) in the accounting systems within stores, as for example when a store reports exactly the same level of sales for two consecutive months. The Census Bureau has established procedures for editing the incoming data, screening out "unreasonable" observations, and imputing replacement observations where necessary.

The Census Bureau's task in deriving the monthly estimates is complicated by the fact that some retailers do not tabulate their sales by calendar month. Many of these noncalendar reporters compile their data according to an accounting calendar that is built around four-week and five-week periods. Each quarter in this alternative calendar consists of two four-week periods separated by a five-week period. Thus the accounting year, made up of four such quarters, consists of 52 weeks, or 364 days.[11] Most period-reporters define the accounting periods as beginning on a Sunday and ending on a Saturday. For the largest chain stores, data are reported in *The Wall Street Journal* and elsewhere on a reporting-period basis. The motivation for this method of reporting is not clear, except as it may be an elementary method for dealing with day-of-week variation in the sales data.[12] The pervasiveness of this practice of reporting on a four-week basis apparently varies by type of retailer; for example, department stores frequently are period-reporters, whereas automotive dealers usually are calendar-month-reporters.

From the Census Bureau's point of view, the period data are an inconvenience, because they must be adjusted to a calendar-month basis before the rest of the data-construction process can get under way. The Census Bureau performs this adjustment using day-of-week factors derived from the historical time series for the category as a whole. The production of not-seasonally-adjusted data from the period data is somewhat problematic, given that day-of-week variation is not identified in the period data. There is no publicly available information on the share of sales in each category that is due to period-reporters.

Also related to the data-construction process is the adjustment of the data for seasonal and other calendar-related variation. The Census Bureau uses the X-11 ARIMA computer program to adjust for seasonal and day-of-week variation.[13] Each month, the Census Bureau estimates the seasonal factors for the advance, preliminary, and final estimates using the entire time series of available observations. At the same time, they also publish revised factors for the year-earlier estimates corresponding to the advance and preliminary months. Seasonal factors for the other months are held fixed from month to month, however, and a complete revision of historical data to reflect a fully consistent set of seasonal factors is undertaken only at the time of each annual revision.

II. Sampling Error in the Retail-Sales Estimates

As was described above, the Census Bureau uses each month's incoming informa-

[10]Note, however, that the Census Bureau uses the annual survey only to benchmark the monthly series (indeed the Census Bureau only asks for annual totals); month-to-month movements are determined by the monthly surveys.

[11]As a result, the accounting year shifts each year with respect to the ordinary calendar, so an adjustment is made periodically to bring the accounting year back into line with the ordinary calendar. In 1980, a one-week interlude was introduced between the end of the January period and the beginning of the February period; in 1985 some stores treated January as a five-week period while others introduced another one-week interlude.

[12]Day-of-week variation is induced by differences in the composition of months by type of day. In the case of sales at food stores, for example, months with five Fridays tend to have relatively high sales (other things equal), while months with five Mondays tend to have relatively low sales.

[13]The Census Bureau also adjusts several of the components of retail sales for holiday variation (such as that induced by Easter); this adjustment is handled outside the X-11 ARIMA problem.

tion to produce two so-called "unbiased estimates": one of sales in the "current" month and one of sales in the "previous" month. Both of these unbiased estimates contain sampling error. To fix notation, let z_t be the true (but unobserved) value of sales in month t; let z'_t be the first unbiased estimate of sales in month t, and let z''_{t-1} be the second unbiased estimate of sales in month $t-1$.[14] Recall from the earlier discussion that z'_t and z''_{t-1} both are derived in month $t+2$ from information supplied by one of the panels in the sample rotation scheme.

The sampling errors ξ'_t and ξ''_{t-1} can be defined implicitly as the differences between the observed values of sales and the true values (ignoring other potential deficiencies in the data):

$$\xi'_t = z'_t - z_t$$
$$\xi''_{t-1} = z''_{t-1} - z_{t-1}.$$

In the simplest of all possible worlds, ξ'_t and ξ''_{t-1} would be mutually and serially uncorrelated. Unfortunately, this turns out not to be the case: calculations performed by the Census Bureau using store-level data for the 27 months from January 1973 through March 1975 show that ξ'_t and ξ''_{t-1} are highly autocorrelated at lags 4, 8, 12, 16, and so on (i.e., at lag lengths equal to multiples of the number of panels in the rotation scheme). The Census Bureau calculations also give clear evidence of additional autocorrelation at the seasonal lags: the autocorrelation at lag 12 is greater than the autocorrelation at lag 8 (and similarly at lag 24 relative to lag 20), despite the overall tendency for the autocorrelations to damp out with increasing lag length.[15] Correlations at lags other than multiples of 4 are much smaller, but in some cases they appear to be different from zero. Finally, ξ'_t and ξ''_{t-1} are extremely highly correlated.

[14] In part, the notation in this section is borrowed from Wolter (1979).

[15] This additional autocorrelation at the seasonal lags may reflect that the seasonality of sales at stores excluded from the sample is different from the seasonality of sales at stores included in the sample.

Motivated in part by the evidence from the Census Bureau calculations, William R. Bell and Steven C. Hillmer (1990) propose the following bivariate model for the sampling error in the pre-1977 data:

$$(1-\rho_4 L^4)(1-\rho_{12}L^{12})\xi'_t = \eta'_t$$
$$(1-\rho_4 L^4)(1-\rho_{12}L^{12})\xi''_{t-1} = \eta''_{t-1}$$
$$\operatorname{corr}(\eta'_t, \eta''_{t-1}) = \Lambda.$$

For post-1977 data, the model would be modified by replacing $\rho_4 L^4$ with $\rho_3 L^3$. The model allows for nonzero autocorrelations at lags that are multiples of the number of panels in the rotation (either 4 or 3 depending on whether pre- or post-1977 data are being examined) and assumes that the autocorrelations at the other lags will be zero. The latter assumption is appealing because it would hold if the sampling errors from different panels were uncorrelated. Although there is some evidence against this hypothesis in the Census Bureau calculations using the store-level data, it appears to be consistent with the gross characteristics of most sales series. (See Bell and Wilcox [1993] for a discussion of factors that may have induced correlation between panels.) The multiplicative factor at lag 12 in the Bell-Hillmer specification allows for additional autocorrelation at seasonal lags. Finally, the specification allows for correlation between the current-month and previous-month estimators.

The Census Bureau must use the noisy unbiased estimates to construct a single official estimate of sales in any given month. Ideally, the construction method should minimize the sampling variation in the published data and should be computationally simple. At present, the Census Bureau is using the "composite" estimation procedure to accomplish this task (see Wolter [1979] for a discussion of composite estimation). The preliminary composite estimator is given in equation (1):

$$(1) \quad P_t = (1-\beta) z'_t + \beta(P_{t-1} + z'_t - z''_{t-1}).$$

In words, the preliminary composite estimate is calculated as a weighted average of the first unbiased estimate for the current month, and the preliminary composite estimate for the previous month scaled up by the current panel's estimate of the change in sales between the previous month and the current month. The final composite estimator is given by

$$(2) \qquad F_{t-1} = (1-\alpha) z''_{t-1} + \alpha P_{t-1}.$$

Prior to the 1977 switch from a four-panel to a three-panel rotation, β was set at 0.80 and α at 0.82; since 1977, β has been set at 0.75 and α at 0.80.

Given Bell and Hillmer's (1990) specification of the model for the sampling error in the unbiased estimates, it is straightforward to use equations (1) and (2) to derive the model for the final composite estimate F_t:

$$F_t = z_t + \xi_t^{\mathrm{F}}.$$

For data from the pre-1977 period, ξ_t^{F} follows the model

$$(3) \quad (1-\beta L)(1-\rho_4 L^4)(1-\rho_{12} L^{12})\xi_t^{\mathrm{F}} = (1-\theta_1 L)\nu_t.$$

Again, for data from the post-1977 period, the model is modified by substituting $\rho_3 L^3$ for $\rho_4 L^4$. Bell and Wilcox (1993) apply this model to data for seven categories of stores and report that the typical autocorrelation function can be fairly well approximated by setting $\rho_4 = 0.62$ for the pre-1977 data (correspondingly, $\rho_3 = 0.70$ for the post-1977 data), $\rho_{12} = 0.75$, and $\theta_1 = -0.10$.

With this background, one can now evaluate the potential impact in a variety of contexts of failure by the econometrician to account for the presence of sampling error in the final composite estimates. As a first example, suppose that one were interested in estimating the coherence of monthly sales in different cities in the United States, with a view toward assessing the relative importance of common stocks and idiosyncratic shocks in generating city-level fluctuations in sales. Evidence on this point might be useful for understanding the sources of business-cycle fluctuations and in particular whether common shocks or idiosyncratic shocks are more important.

Of course, sampling error would tend to cause the role of idiosyncratic factors to be exaggerated. As a framework for analysis, consider the following simple setup:

$$s_{1t} = \alpha_1 \gamma_t + \varepsilon_{1t} + \Delta\xi_{1t}$$

$$s_{2t} = \alpha_2 \gamma_t + \varepsilon_{2t} + \Delta\xi_{2t}$$

where s_{it} is growth in sales in city i, γ_t is the common shock, ε_{it} is the economically interesting idiosyncratic variation in sales, and $\Delta\xi_{it}$ is the sampling error (and hence is economically uninteresting).[16] The variables γ_t, ε_{1t}, ε_{2t}, $\Delta\xi_{1t}$, and $\Delta\xi_{2t}$ are assumed to be mutually uncorrelated.

It is easy to show that the correlation between s_{1t} and s_{2t} has a simple interpretation: it equals the geometric mean of the proportion of the total variation in each series that is accounted for by the variation in the common component:

$$\operatorname{corr}(s_{1t}, s_{2t}) = \left[\left(\frac{\alpha_1^2\sigma_\gamma^2}{\alpha_1^2\sigma_\gamma^2 + \sigma_{\varepsilon 1}^2 + \sigma_{\Delta\xi 1}^2}\right) \times \left(\frac{\alpha_2^2\sigma_\gamma^2}{\alpha_2^2\sigma_\gamma^2 + \sigma_{\varepsilon 2}^2 + \sigma_{\Delta\xi 2}^2}\right)\right]^{1/2}$$

A more economically relevant measure would be one that removed the effect of the sampling error; such a measure is easily constructed:

$$\operatorname{corr}^*(s_{1t}, s_{2t}) = \frac{\operatorname{corr}(s_{1t}, s_{2t})}{\left[\left(\frac{\alpha_1^2\sigma_\gamma^2 + \sigma_{\varepsilon 1}^2}{\alpha_1^2\sigma_\gamma^2 + \sigma_{\varepsilon 1}^2 + \sigma_{\Delta\xi 1}^2}\right)\left(\frac{\alpha_2^2\sigma_\gamma^2 + \sigma_{\varepsilon 2}^2}{\alpha_2^2\sigma_\gamma^2 + \sigma_{\varepsilon 2}^2 + \sigma_{\Delta\xi 2}^2}\right)\right]^{1/2}}$$

$$= \frac{\operatorname{corr}(s_{1t}, s_{2t})}{\left[\left(1 - \frac{\sigma_{\Delta\xi 1}^2}{\sigma_{s1}^2}\right)\left(1 - \frac{\sigma_{\Delta\xi 2}^2}{\sigma_{s2}^2}\right)\right]^{1/2}}$$

[16] The notation is chosen to emphasize that, according to the Bell-Hillmer specification, the sampling error is stationary in the log levels and hence overdifferenced in the growth rates.

TABLE 2—THE IMPACT OF SAMPLING ERROR ON THE VARIANCES OF MONTHLY RETAIL SALES IN SELECTED U.S. CITIES

City	Variance of sales	Coefficient of variation	$1-\frac{\sigma^2_{\Delta\xi j}}{\sigma^2_{sj}}$
New York	0.000842	0.051	0.542
Los Angeles	0.001141	0.073	0.308
Chicago	0.001178	0.091	—[a]
Philadelphia	0.004060	0.087	0.724
Detroit	0.001079	0.049	0.670

Notes: The variances shown in the second column are computed by regressing the log difference of each non-seasonally-adjusted sales series on seasonal dummies and differenced day-of-week variables and then squaring the standard error of the regression (sample period: 1978:2–1987:12). The values reported for the coefficient of variation are the medians of CV's for August 1984–January 1985 (Source: *Monthly Retail Trade Report*, May 1985). The last column shows the estimated fraction of the total variation in sales accounted for by elements other than the sampling error.

[a]The estimates imply that more than 100 percent of the variation in the series is induced by sampling error.

where σ^2_{si} is the variance of sales in city i. In words, the raw correlation should be boosted by a factor that is increasing in the proportion of the total variation of sales accounted for by sampling error.

Table 2 presents relevant information constructed from monthly sales data for five U.S. cities. The first column shows the variance of sales in each city, calculated by regressing the log difference of each sales series (unadjusted for seasonal variation) on 12 seasonal dummies and seven day-of-week variables and then squaring the standard error of the regression. The middle column gives coefficients of variation for each series. The last column displays the proportion of the variation in sales in each city generated by sources other than the sampling error; this quantity was calculated by dividing the square of the second column by the first column, after adjusting the CV to take account of the fact that the data have been differenced.

The figures in Table 2 show that, in the sales series for Detroit and Philadelphia, more than a quarter of the variation is estimated to have been induced by sampling error; in New York, nearly half; and in Los Angeles, about two-thirds. As for Chicago, the figures taken literally would imply that sampling error accounts for more than 100 percent of the city-level variation in sales.[17]

Table 3 presents correlations between sales in these five cities at the monthly frequency, unadjusted for sampling error (above the diagonal) and adjusted (below the diagonal). Given the results of Table 2, it is not surprising to find that some of the raw correlations are boosted considerably by the correction for sampling error. For example, the correlation between sales in New York and Los Angeles is more than doubled, from 0.230 to 0.563, and the correlation between Detroit and Los Angeles is increased from 0.249 to 0.548. Thus, in the case of the monthly city-level sales data, sampling error appears to be a first-order concern in computing both the variance of the individual series and the covariances among them.

How much would use of quarterly averages rather than monthly data alleviate the difficulties noted here? Results for quarterly averages are shown in Tables 4 and 5 and

[17]This result probably reflects that the CV's themselves are noisy estimates; the one published for Chicago in the May 1985 *Monthly Retail Trade Report* simply may have been too high.

TABLE 3—CORRELATIONS BETWEEN SALES IN SELECTED U.S. CITIES: MONTHLY DATA

City	City				
	New York	Los Angeles	Chicago	Philadelphia	Detroit
New York	1	0.230	−0.008	0.153	0.127
Los Angeles	0.563	1	0.179	0.084	0.249
Chicago	—[a]	—[a]	1	−0.199	0.093
Philadelphia	0.244	0.178	—[a]	1	0.216
Detroit	0.211	0.548	—[a]	0.310	1

Notes: The correlations are computed using the residuals from the same regressions described in the notes to Table 2. Raw correlations are shown above the diagonal; correlations adjusted for sampling error are shown below the diagonal.
[a]Figures could not be computed; see Table 2, note a.

TABLE 4—THE IMPACT OF SAMPLING ERROR ON THE VARIANCES OF QUARTERLY RETAIL SALES IN SELECTED U.S. CITIES

City	Variance of sales	$1-\frac{\sigma^2_{\Delta\xi j}}{\sigma^2_{sj}}$
New York	0.001294	0.833
Los Angeles	0.001561	0.716
Chicago	0.001374	0.498
Philadelphia	0.006392	0.901
Detroit	0.001132	0.823

Notes: The variances are computed by running the quarterly analogues to the regressions described in Table 2. The last column shows the estimated fraction of the total variation in sales accounted for by elements other than the sampling error.

can be summarized as follows. The estimated unconditional variance of growth in sales is larger in the quarterly data than in the monthly data, by between 5 and 60 percent. The unconditional variance of the *level* of the sampling error, given the specification of Bell and Hillmer (1990), is about the same at either frequency; however, the variance of the *differenced* sampling error (which is the statistic relevant for computing the corrected correlations) is about half as large in the quarterly data as in the monthly data. Therefore, as shown in the right-hand column of Table 4, the proportion of the total variation induced by the signal is substantially higher for all cities in the quarterly averages than it is in the monthly data. The most dramatic improvement occurs in the case of Chicago, where nearly 50 percent of the variation in the quarterly data is estimated to have been generated by the signal.

Table 5 shows that the raw correlations (given above the diagonal) in most cases are higher at the quarterly than at the monthly frequency. Pairwise comparison with the entries below the diagonal shows that even at the quarterly frequency the adjustment of the cross-city correlations for sampling error can be substantively important. Whereas only one of the unadjusted correlations is greater than 0.50, four of the adjusted correlations are. Excluding the two city-pairs with unadjusted correlations less than 0.05, the average of the unadjusted correlations is 0.34, compared with 0.45 for the adjusted correlations. The unadjusted average would suggest that idiosyncratic shocks are twice as important as common shocks; the adjusted average indicates that idiosyncratic shocks are only about 25-percent more important than common shocks. Clearly, use of quarterly averages alleviates the difficulties discovered with the use of monthly data, but it does not eliminate them.

These results must be regarded as only suggestive, because they lean heavily on the assumed structure of the model for the sampling error; in particular, the calculations depend on the assumption that the parameter values identified in Bell and Wilcox (1993) as typical in the national estimates disaggregated by type of store also are valid for total sales at the city level. Definitive

TABLE 5—CORRELATIONS BETWEEN SALES IN SELECTED U.S. CITIES: QUARTERLY DATA

	City				
City	New York	Los Angeles	Chicago	Philadelphia	Detroit
New York	1	0.442	−0.035	0.226	0.430
Los Angeles	0.572	1	0.272	0.309	0.512
Chicago	−0.054	0.456	1	0.026	0.360
Philadelphia	0.261	0.385	0.039	1	0.129
Detroit	0.519	0.667	0.562	0.150	1

Notes: The correlations are computed using the same residuals as those underlying the figures reported in Table 4. Raw correlations are shown above the diagonal; correlations adjusted for sampling error are shown below the diagonal.

work would require that sampling autocorrelations be estimated for the metropolitan-area data.

As a second example of the potential consequences of failure to account for sampling error, consider the possibility of using noise-contaminated data to estimate the depreciation rate of a composite consumption good. Under the joint null hypothesis of rational expectations, complete nondurability, and perfect coincidence of the consumer's decision period with the sampling period, changes in spending would be predicted to follow a martingale process; but noise-contaminated data would not follow such a process even if the underlying signal did so. Therefore, sampling error could, in principle, lead to incorrect inference about the depreciation rate, just as it did about the coherence of cross-city fluctuations in sales.

To shed further light on the influence that sampling error might have on inferences about depreciation rates, I specified models for the signal and the sampling error, derived the implied model for the observed series, and then compared the true depreciation rate with the one that would be estimated from the observed data if the sampling error were ignored. Two alternatives were considered for the signal model: a random walk (interpreted as representing the sales of a hypothetical nondurable good) and an IMA(1,1) with moving-average parameter equal to 0.9 (interpreted as representing the sales of a durable good). The sampling error was assumed to follow the monthly specification of Bell and Hillmer (1990) for the three-panel survey with ρ_3 set equal to 0.70, ρ_{12} set equal to 0.75, and θ_1 set equal to -0.10. Three cases were considered for the variance of the sampling error: a no-sampling-error case, a low-sampling-error case, and a high-sampling-error case. The high-error case assumes a sampling error roughly comparable to that present in sales at men's clothing stores (CV = 4.2 percent in the composite estimates), building material and supply stores (CV = 5.3 percent), and home appliance stores (CV = 5.9 percent). The calculations were performed using the software described in William P. Cleveland (1986).

The results were as follows. In the cases in which sales are assumed to be estimated without sampling error, the true depreciation rate is recovered without distortion. However, the estimated depreciation rate diverges from the true rate once sampling error is added to the model. In the nondurable-goods case, the estimated depreciation rate falls from 100 percent when no sampling error is present to 95 percent in the low-sampling-error case and 84 percent in the high-sampling-error case. The estimate in the low-sampling-error case would correspond to a moving-average parameter of 0.05—probably small enough to be passed off as differing from zero only insignificantly. However, the estimate in the high-sampling-error case would correspond to a moving-average parameter of 0.16—prob-

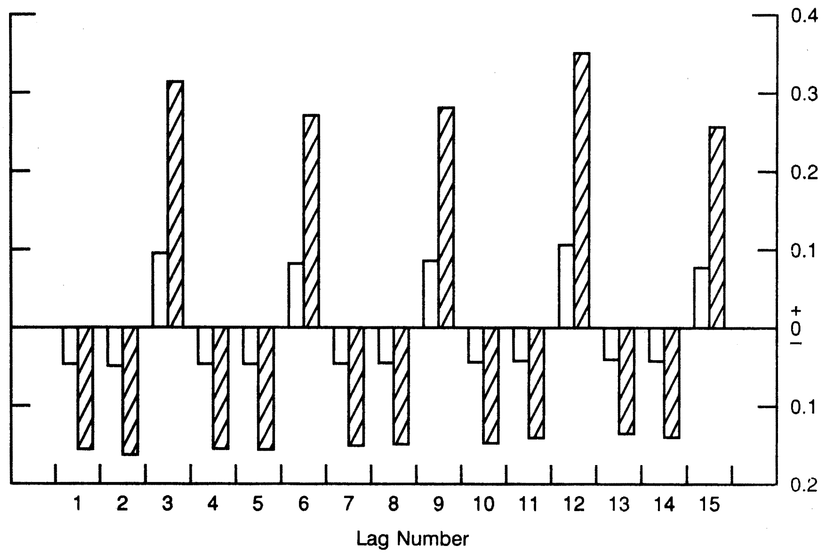

FIGURE 1. AUTOCORRELATIONS OF THE SAMPLING-ERROR-CONTAMINATED SERIES ON NONDURABLE-GOODS PURCHASES

Key: Open bars show the autocorrelations in the low-sampling-error case; striped bars show the autocorrelations in the high-sampling-error case.

ably high enough to be declared significantly different from zero (the asymptotic standard error in a sample of 150 observations is roughly 0.08).

In the case of the durable good, the estimated depreciation rate rises from 10 percent in the absence of sampling error to 20 percent in the low-sampling-error case and 34 percent in the high-sampling-error case. Thus, sampling error is predicted to blur the distinction between the behavior of the durable good and the nondurable good, making the former look less durable than it really is and making the latter look more durable than it really is.

For this example as well, one can ask how the results would be affected if quarterly averages were used instead of monthly data. To answer this question, I time-averaged the monthly signal and noise models underlying the calculations reported above and then computed the model for the resulting quarterly series. The implied depreciation rate of the nondurable good turns out to be quite insensitive to sampling error in quarterly averages, provided that proper allowance is made for the effects of the time-averaging: even when the data are relatively noisy, the naive estimate, calculated without adjustment for the sampling error, would differ from the true value by only 0.02. However, estimates of the durable good's depreciation rate are biased upward in the quarterly data even more severely than in the monthly data. The MA coefficient at lag 1 (equal to 0.73 in the signal model after time-averaging) falls to 0.52 in the low-sampling-error case and to 0.26 in the high-sampling-error case.

Thus, whether one gains or loses from the use of quarterly averages turns out to depend on the characteristics of the underlying true series. If the monthly signal series is highly persistent (as appears to be the case in most categories of retail sales), quarterly averaging of the data will reduce the influence of sampling error. Indeed, sampling error of the magnitude typically encountered in the national-level estimates of retail sales probably can be ignored in estimating composite depreciation rates from national-level quarterly averages. However, if the monthly signal series contains a large transitory component, sampling error may have an even greater influence in the quarterly data than in the monthly data, the high-frequency variation in the signal being reduced by the quarterly averaging, leaving the fluctuation in the sampling error with a relatively more prominent role.

In closing this section, it is worth noting that sampling error of the Bell-Hillmer type affects the autocorrelations of the observed series at lags higher than 1 as well as at lag 1. Figure 1 illustrates this point for the case of the nondurable good at the monthly frequency by showing the first 15 theoretical autocorrelations of the observed series (after differencing) in both the low-sampling-error (open bars) and high-sampling-error cases (striped bars). Recall that the autocorrelations of the signal series are all zero. As might be expected from the specification of the model, the influence of the sampling error is greatest at lags that are multiples of 3.

Thus, the chart suggests that tests of the random-walk hypothesis, as applied to the sampling-error-contaminated data, will reject too often the null hypothesis of no autocorrelation in the first differences even if the autocorrelation at the first lag is ignored (as is commonly done). Nonetheless, these results do not impugn rejections of the random-walk hypothesis (such as the one reported by Hall [1978]) based upon regressions of the observed consumption series on lagged values of another series, such as an index of stock prices, because the sampling-error-induced fluctuation should be purely idiosyncratic.[18]

III. The Product Composition of Sales

The retail-sales estimates pertain to sales by type of store and not by type of product.[19] Thus, for example, the Census Bureau receives monthly reports on sales at grocery stores and sales at department stores, but they do not, strictly speaking, receive monthly reports on sales of food. Nevertheless, BEA produces monthly estimates of personal consumption expenditures by type of product. How are these estimates prepared?

For some categories of PCE, BEA does have direct measures of spending by type of product. For example, the Energy Information Administration (EIA) in the Department of Energy produces estimates of electricity consumption based on monthly billings of a sample of electrical utilities; Ward's Communications publishes data on unit sales of domestically produced automobiles based on reports from the automakers, who in turn are receiving daily results from the universe of franchised dealers. Thus, spending in categories such as these can be measured fairly directly.[20]

[18]Given that the sampling error is autocorrelated, the orthogonality test could be contaminated even in the case of lagged stock prices if there is feedback from the sampling error onto the stock market. The validity of the orthogonality test is more difficult to assess in the case of disposable personal income, because some employment data are used as indicators for the construction of both the personal-income and the PCE estimates.

[19]It is important to note in this regard that the "nondurable-goods" figure in the *Monthly Retail Trade Report* actually refers to sales at nondurable-goods stores, that is, stores whose principal line of business is the sale of nondurable goods. Department stores are classified as nondurable-goods stores; thus, given that department stores (by definition) sell furniture, they are an example of a "nondurable-goods store" selling "durable goods."

[20]This is not to suggest that there are not problems in measuring spending on either electricity or automobiles. For example, the EIA figures on electricity consumption are computed from a sample of 225 utilities out of a universe of more than 3,200; hence, the estimates include some sampling error. A second complication arises from the fact that utilities typically divide their customers into separate "billing cycles," which are processed on a rotating basis. As a result, the monthly figure reported to the EIA (consisting of the sum of amounts billed across billing cycles) does not pertain to a uniform time period. Thirdly, the billing data are reported on a cash basis rather than an accrual basis; the distinction between cash and accrual arises, for example, when a customer is on a billing plan that smooths payments over the year. The data reported from the utility will pertain to the smoothed

However, for several other categories of PCE (notably, food, clothing and shoes, and furniture), no such specialized data sources exist, so BEA estimates spending in these categories from the retail-sales data, using a "product-composition matrix" to map sales by kind of business into spending by type of product.[21] The product-composition matrix is constructed using information from the quinquennial Census of Retail Trade. (In this census, stores are asked to estimate the composition of their sales.) BEA estimates the relevant components of seasonally adjusted consumption by first seasonally adjusting the disaggregated retail-sales data and then passing them through the product-composition matrix. The outputs from the product-composition matrix are used as extrapolators for the "control" components of seasonally adjusted PCE. Similarly, BEA constructs the non-seasonally-adjusted consumption data by passing the non-seasonally-adjusted retail-sales data through the same product-composition matrix as was used in the case of seasonally adjusted data.[22]

This method for solving the product-composition problem has the advantage of producing estimates of PCE by type of product that will be robust to shifts in preferences for shopping at, say, department stores relative to specialty apparel stores. However, the estimates will not be robust to variation in the composition of spending within any given type of store. On the contrary, the methodology assumes that each dollar spent at, say, department stores will be allocated to the various products sold within those stores in fixed proportions in nominal terms.[23] Not surprisingly, there exist conditions that would guarantee the constancy of nominal spending shares at each type of store, namely, that all goods have the same durability, that utility is homothetic, that the elasticity of substitution between goods equals 1, and that tastes are constant.

Of course, violations of these assumptions will cause the estimated split of spending between goods to differ from the actual split. For example, consider the case in which goods differ in their durability but all the other assumptions are satisfied. Standard theories predict that spending on durable goods should initially overshoot its permanent rate after a shock to permanent income, as consumers adjust their holdings of durables. Spending on nondurable goods will adjust at the same time but will not overshoot. Subsequently, spending on durable goods will reverse its overshooting, while spending on nondurables remains at its new permanent level. Therefore, during

payments and not to the actual usage by the customer during the billing period.

As for cars, the unit-sales data should be essentially exact; measurement of prices, however, is exceedingly difficult because the typical transaction consists of many elements, often including trade-in of a used automobile. Also, the allocation of total sales between consumers and businesses does not appear to be exact, especially for foreign-produced automobiles.

[21]For example, results from the 1977 Census of Retail Trade showed that spending on food or alcohol constituted 87.6 percent of sales in that year at grocery stores, 98 percent of sales at "other food stores," 0.3 percent at "other home-furnishings stores," 0.1 percent at household-appliance stores, 0.2 percent at hardware stores, 0.4 percent at nurseries and lawn and garden stores, 2.1 percent at gasoline stations, 4.6 percent at department stores, 8.3 percent at variety stores, 3.4 percent at mail-order stores, 5.8 percent at "other general merchandise" stores, 0.1 percent at women's-apparel stores (but none at men's-apparel stores), 7.5 percent at drug stores, 98.9 percent at eating places, 97.6 percent at drinking places, 0.6 percent at sporting-goods stores, 0.8 percent at book stores, 0.1 percent at flower stores, 1.6 percent at "other nondurable-goods" stores, 26.2 percent at "nonstore" stores, 95.8 percent at liquor stores, 82.7 percent at military commissaries, and 9.9 percent at military exchanges.

[22]Non-seasonally-adjusted consumption data are available only on a quarterly, current-dollar basis; such data are not available at the monthly frequency, nor are they available on a constant-dollar basis. The discussion in the text should make it clear that it is simply wrong to suppose that there must exist a non-seasonally-adjusted analogue for every seasonally adjusted component of PCE.

[23]It is possible to make the alternative identifying assumption, namely, that shopping patterns between types of stores are fixed for a single type of good but that the composition of spending within a given type of store may vary. Samuel S. Kortum (1990) pursues this strategy.

the adjustment of consumption to news about permanent income, the actual composition of spending will fluctuate. The extent to which this fluctuation is reflected in the estimated PCE data will be sharply limited by the product-composition matrix, which will impose a fixed composition of sales for each type of store.

Next, assume that there are only two goods and consider the consequences of the two goods having an elasticity of substitution other than 1; for the sake of concreteness, suppose that the elasticity is less than 1. In this case, the apparent price elasticity of demand for both goods will exceed the actual price elasticity.[24] In addition, the level of spending on the two goods taken together will be overstated in real terms whenever the relative price differs from its base-period value.[25] Furthermore, this overstatement will increase in size as the relative price moves further away from its base-period value; that is, the change in total spending will be overestimated in real terms. If the relative price were to move back toward its base-period value, the change in total spending would be understated while the price was moving, but the estimated level of spending would remain above the actual level until the base-period value of the relative price had been achieved. Not surprisingly, all of these predictions would be reversed if the elasticity of substitution were greater than 1.

Finally, consider the consequences of fluctuations in relative tastes for various consumption goods. Such fluctuations will cause the actual composition of spending by type of store to vary (except in special circumstances). Again, however, the estimated composition of spending by type of store will be imposed by the fixed product-composition matrix, and the coherence of the various estimated series on spending will be too great.

Theoretical considerations aside, what can be said about the composition of nominal spending at any given type of store? Among the categories of stores, department stores are the most problematic in this regard, because spending at those stores is relatively evenly distributed between durable goods (mainly furniture) and nondurable goods (especially apparel). Current methodology rests on the assumption that spending on furniture at department stores is proportional to spending on apparel at department stores. Indirect evidence on the validity of this assumption can be developed by examining the behavior of sales at furniture stores relative to sales at apparel stores.[26]

Figure 2 pursues this strategy. The solid line in the top panel depicts quarterly observations on the ratio of non-seasonally-adjusted sales at furniture stores to non-seasonally-adjusted sales at apparel stores for the period 1967–1988, and shows that this ratio has often varied by more than 20 percentage points from seasonal peak to seasonal trough. (At the monthly frequency, the variation in this ratio has been on the order of 50 percentage points.) The solid line in the bottom panel shows the same ratio calculated from seasonally adjusted series. Evidently, some of the variation in the ratio is nonseasonal. Indeed, some of the variation appears to be nontrend and, therefore, unlikely to be fully accounted for by using the various Censuses of Retail Trade to provide benchmarks for the consumption data.

Variation in the composition of spending within department stores will have impor-

[24] If the relative price of the first good increases, the actual nominal share of spending on the first good will also increase. BEA's methodology, however, would assume that the budget shares had not changed. As a result, the level of spending on the first good would be underestimated, both in nominal terms and in real terms.

[25] As argued in the previous footnote, current procedure will assign too few nominal dollars of spending to the first good, the deflator for which had moved above its base-period value of unity, and too many dollars to the low-deflator category. Moreover, the error in the low-deflator category, being of the same size as the error in the high-deflator category in nominal terms, must be the larger of the two in real terms.

[26] According to the 1977 Census, sales of apparel accounted for more than 95 percent of sales at apparel stores, and sales of furniture accounted for roughly 90 percent of sales at furniture stores.

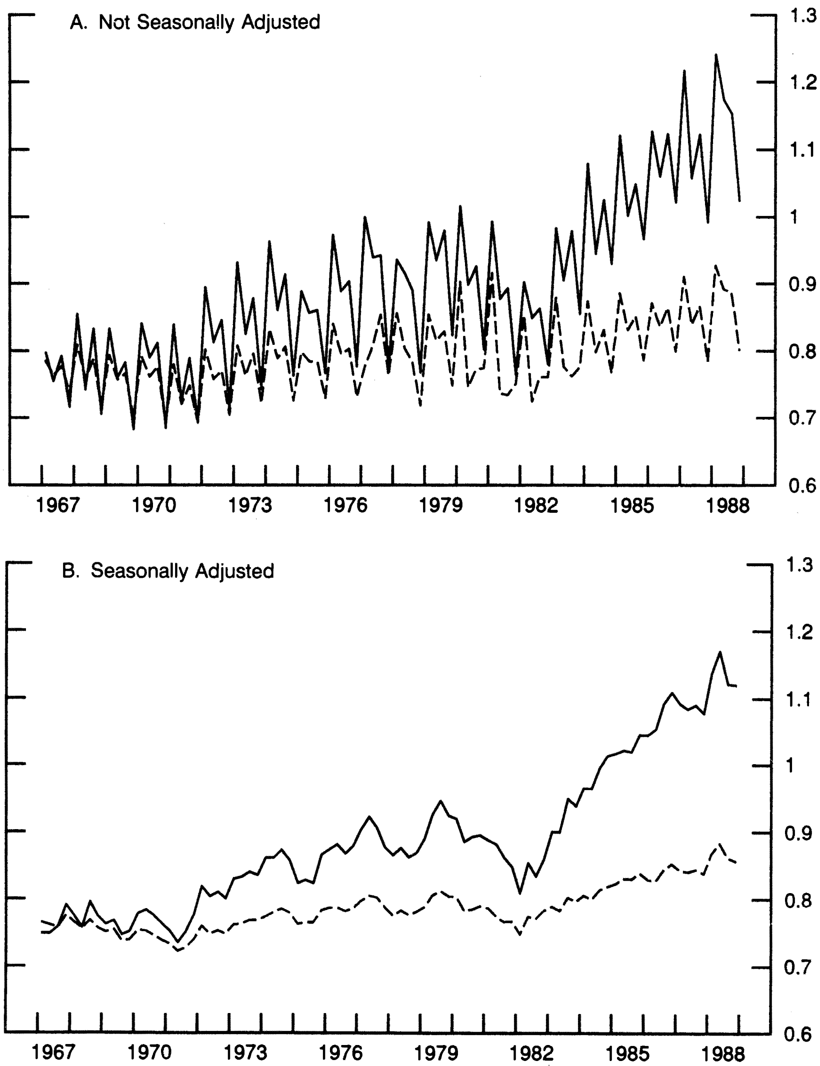

FIGURE 2. FURNITURE AND APPAREL IN THE PCE ACCOUNTS AND IN RETAIL SALES (QUARTERLY, 1967–1988)

Key: Solid lines denote the ratio of sales at furniture stores to sales at apparel stores; dashed lines denote the ratio of PCE on furniture to PCE on clothing and shoes.

tant implications for the behavior of the PCE estimates only if purchases of furniture at department stores constitute a significant fraction of total spending on furniture and likewise for apparel. Results from the 1977 Census of Retail Trade confirm that this is so: sales of furniture at department stores accounted for roughly 30 percent of total spending on furniture, while sales of apparel at department stores accounted for 35 percent of total spending on clothing and shoes. Thus, an important fraction of estimated outlays in both categories is calculated as a fixed proportion of the same quantity: spending at department stores.

As a result, it is not surprising to find that the ratio of PCE on furniture to PCE on clothing and shoes is much less variable than the ratio of retail sales at furniture stores to retail sales at apparel stores. Figure 2A shows that the non-seasonally-adjusted PCE ratio (shown as the dashed line) is one-third to one-half less variable than the non-seasonally-adjusted retail-sales ratio, suggesting that the PCE estimates may be substantially overstating the coherence between these two categories of spending. Figure 2B shows that the same appears to be true of the seasonally adjusted data: the PCE ratio moves much less than the retail-sales ratio.

These visual impressions can be corroborated by way of simple regression analysis. Regressing the log difference of the NSA retail sales ratio against seasonal dummies, I obtain

$$\Delta \ln\left(\frac{\text{sales at furniture stores}}{\text{sales at apparel stores}}\right)_t$$

$$\begin{aligned} =\; & \underset{(0.007)}{0.209\,\mathrm{D1}_t} - \underset{(0.007)}{0.093\,\mathrm{D2}_t} + \underset{(0.007)}{0.038\,\mathrm{D3}_t} \\ & - \underset{(0.007)}{0.144\,\mathrm{D4}_t} + e_t \end{aligned}$$

(sample period = 1968:1–1988:4, $\bar{R}^2 = 0.94$, DW = 2.50), whereas for the PCE ratio, I obtain

$$\Delta \ln(\text{PCE on furniture/PCE on apparel})_t$$

$$\begin{aligned} =\; & \underset{(0.009)}{0.130\,\mathrm{D1}_t} - \underset{(0.009)}{0.074\,\mathrm{D2}_t} + \underset{(0.009)}{0.021\,\mathrm{D3}_t} \\ & - \underset{(0.009)}{0.073\,\mathrm{D4}_t} + e_t \end{aligned}$$

(sample period = 1968:1–1988:4, $\bar{R}^2 = 0.81$, DW = 2.28). The left-hand-side variable in both regressions can be interpreted as the discrepancy between the growth rates of furniture and apparel. Comparison of the coefficient estimates confirms that these discrepancies are much more pronounced in the non-seasonally-adjusted retail-sales data than they are in the non-seasonally-adjusted PCE data.

As for the seasonally adjusted data, the log difference of the seasonally adjusted retail-sales ratio (using official seasonally adjusted data for both numerator and denominator) has a standard deviation of 2.4 percent; the standard deviation of the log-differenced PCE ratio is only 1.3 percent. Apparently, the current methodology for translating the retail-sales data into PCE also suppresses a substantial portion of the nonseasonal idiosyncratic fluctuation in the components of spending.

The following example further illustrates the potential importance of the product-composition issue for the characteristics of the PCE data. Suppose that one is interested in determining the relative procyclicity of various consumption goods.[27] A simple method for estimating relative procyclicity would involve regressing the growth in each of the various components of spending on the growth in total spending:

$$\Delta \ln C_{it} = \alpha_0 + \beta_i \Delta \ln C_t + \varepsilon_{it}.$$

The coefficient β_i would measure the procyclicity of the ith spending component.

For retail sales at women's apparel stores (the reason for selecting this component will be apparent shortly), I find:[28]

$$\Delta \ln(\text{retail sales at women's apparel stores in constant dollars})_t$$

$$\begin{aligned} = & \underset{(0.006)}{-0.003} + \underset{(0.65)}{2.32}\,\Delta \ln C_t \\ & + \underset{(0.028)}{0.051}\,\mathrm{DUM1980Q2} + \varepsilon_{it} \end{aligned}$$

[27] We might be interested in using the relative procyclicity of the various consumption categories to distinguish durable goods from nondurable goods, in view of the previously noted inability of autocorrelation-based methods to do so.

[28] The source for the retail-sales series is "Revised Monthly Retail Sales and Inventories: January 1978 through December 1987." The observation for 1978:4 was used in calculating the *change* in sales in 1979:1; the observations for 1978:1–1978:3 were dropped from the sample to leave an evenly balanced number of observations for each quarter. The dependent variable in the first regression is calculated as the log difference of the quarterly average of retail sales at women's-apparel stores deflated by the implicit deflator for PCE for women's apparel. The dependent variable in the second regression is the log difference of real PCE for women's apparel. The independent variable in both regressions is the log difference of total real PCE.

(sample period = 1979:1–1987:1, $\bar{R}^2 = 0.26$, DW = 2.66), where DUM1980Q2 is a dummy variable taking the value 1 in 1980:2 and zero otherwise. According to these data, spending on women's apparel increases more than twice in proportion to increases in total spending. However, the results for PCE for women's apparel are quite different:

$$\Delta \ln(\text{PCE, women's apparel in constant dollars})_t = \underset{(0.005)}{0.004} + \underset{(0.48)}{1.31}\,\Delta \ln C_t + \underset{(0.021)}{0.035}\,\text{DUM1980Q2} + \varepsilon_{it}$$

(sample period = 1979:1–1987:1, $\bar{R}^2 = 0.14$, DW = 1.96). These data imply that women's apparel is only slightly more procyclical than the average. Moreover, the difference between the slope coefficients from the two regressions is highly significant: running the difference between the retail sales and PCE variables on the growth in real PCE, I find a slope coefficient of 1.02 with a standard error of 0.40.

It is not difficult to account for the difference between these two sets of results. Department-store sales contribute about as much to PCE for women's apparel as do sales at specialty women's-apparel stores (hence, their selection for this example). As for the relative procyclicity of department-store sales (again using the implicit deflator for women's apparel as the price index) I find:

$$\Delta \ln(\text{department-store sales in constant dollars})_t = \underset{(0.004)}{0.008} + \underset{(0.47)}{0.77}\,\Delta \ln C_t - \underset{(0.020)}{0.005}\,\text{DUM1980Q2} + \varepsilon_{it}$$

(sample period = 1979:1–1987:1, $\bar{R}^2 = 0.10$, DW = 1.22). That is, department-store sales exhibit less-than-average procyclicity.

How should these results be interpreted? One possibility is that PCE for women's apparel is being measured exactly correctly. It could simply be that the women's apparel sold in department stores is a wholly different commodity than the women's apparel sold in specialty stores and that when total consumption is increasing rapidly the composition of spending on women's apparel shifts away from department stores and toward specialty stores (precisely the kind of shift to which current methodology is robust).

It seems unlikely that this possibility is true, however, given that roughly 40 percent of PCE for women's apparel (80 percent of the 50 percent that comes from department stores) in fact represents sales of items other than women's apparel. It is certainly reasonable to entertain the hypothesis that the time-series characteristics of department-store sales reflect the aggregation of spending on a wide variety of goods and may be relatively uninformative about the behavior of outlays for women's apparel. An interesting challenge for future research would be to model explicitly the impact of stochastic shifts in product composition on department-store sales.

IV. Conclusion: A Few Prescriptions

The paper argues, through the use of examples, that imperfections in the U.S. data on personal consumption expenditures may be important enough to influence the conclusions of empirical work as typically conducted. Fortunately, quite a bit is known about several of these imperfections, and the paper has illustrated that strategies often can be developed for improving the robustness of empirical work to particular problems in the data. At a more general level, the factors investigated here suggest the following guidelines for empirical practice. First, PCE for goods and services should be treated separately in any study using disaggregated components of consumption. From the perspective of the measurement system, the usual two-part disaggregation (motivated from theory) into (i) durables and (ii) nondurables plus services does not make much sense. Considerations described above argue for a three-part disaggregation into (i) motor vehicles, (ii) other goods, and (iii) services. A further distinction could be made in category (ii) between (iia) durable goods other than motor vehicles and (iib) nondurable goods; however, this distinction cuts across the heart of the product-composition-related problems discussed above, and it should be shown that substantive conclusions are not sensitive to these problems.

Second, it may be that, for some purposes, the retail-sales estimates (excluding sales at automotive dealers and at building-material and supply stores) should be preferred to the corresponding elements of PCE. Of course, the coverage of total PCE is much broader than the coverage of retail sales, because the latter does not include information about spending on services. Also, the PCE estimates of spending on motor vehicles probably are more reliable than retail sales at automotive dealers as a measure of household outlays for motor vehicles. However, the information content of the PCE and retail-sales estimates for the remaining categories of consumer spending (consisting of outlays on goods other than motor vehicles) is essentially the same, because the PCE estimates are calculated as fixed linear combinations of the retail-sales data. Within this category, the presumption in favor of retail sales should be especially strong whenever non-seasonally-adjusted data are involved, since the assumption in that case of a fixed product composition of sales is especially unappealing.

Third, for extremely disaggregated data such as the metropolitan-area sales data considered above, sampling error must be treated explicitly (both in monthly and in quarterly data) using a realistic model for the sampling error such as the one proposed by Bell and Hillmer (1990). Unfortunately, available information about the time-series behavior of the sampling error is rather incomplete; probably the best that can be done is to assume that the Bell-Hillmer specification with the prototypical parameter values described above can be applied to all series, tailoring the variances to the particular case using the CV's published in the *Monthly Retail Trade Report*.

Finally, given that most of the variation in typical consumption and sales series appears to be permanent, researchers may want to use quarterly data to check that conclusions derived from monthly data are not importantly affected by sampling error. If the use of monthly data is essential to the issue at hand and sampling error seems to be influencing the results, it may be worthwhile to adopt the unobserved-components approach of Bell and Hillmer (1990) and to introduce an explicit model for the sampling error.

In sum, the paper has argued that researchers should not automatically identify the theoretical construct of consumption with its empirical counterpart. Authors in the consumption literature may have been especially tempted to make this identification because the simplest theoretical models deliver such strong implications for the stochastic behavior of the series. Future work should be carried out with a view toward distinguishing between what the data can reasonably be expected to tell us and what they cannot.

REFERENCES

Bell, William R. and Hillmer, Steven C., "The Time Series Approach to Estimation For Repeated Surveys," *Survey Methodology*, 1990, *16* (2), 195–215.

Bell, William R. and Wilcox, David W., "The Effect of Sampling Error on the Time Series Behavior of Consumption Data," *Journal of Econometrics*, 1993 (forthcoming).

Carson, Carol S., "GNP: An Overview of Source Data and Estimating Methods," *Survey of Current Business*, July 1987, *67*, 103–27.

Cleveland, William P., "Calendar Adjustment and Time Series," Special Studies Paper No. 198, Board of Governors of the Federal Reserve System, Washington, DC, 1986.

Garrett, Joseph K., Detlefsen, Ruth E. and Veum, Carol S., "Recent Sample Revisions and Related Enhancements for Business Surveys of the U.S. Bureau of the Census," *Proceedings of the Business and Economic Statistics Section, American Statistical*

Association, 1987, pp. 141–9.

Hall, Robert E., "Stochastic Implications of the Life Cycle–Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, December 1978, *86*, 971–87.

Kortum, Samuel S., "Least Squares Disaggregation of Consumption: Alternative Estimates of Personal Consumption Expenditure by Type of Product," mimeo, Boston University, 1990.

Waite, Preston Jay, "An Evaluation of Nonsampling Errors in the Monthly Retail Trade Sales Data," *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 1974, pp. 602–7.

Wolter, Kirk M., *Introduction to Variance Estimation*, New York: Springer-Verlag, 1985.

______, "Composite Estimation in Finite Populations," *Journal of the American Statistical Association*, September 1979, *74*, 604–13.

______, Isaki, Cary T., Sturdevant, Tyler R., Monsour, Nash J. and Mayes, Fred M., "Sample Selection and Estimation Aspects of the Census Bureau's Monthly Business Surveys," *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 1976, pp. 99–109.

Ward's Automotive Repairs, Detroit: Ward's Communications, various years.