Journal of Econometrics 55 (1993) 235–265. North-Holland

The effect of sampling error on the time series behavior of consumption data*

William R. Bell
U.S. Bureau of the Census, Washington, DC 20233, USA

David W. Wilcox
Board of Governors of the Federal Reserve System, Washington, DC 20551, USA

Much empirical economic research today involves estimation of tightly specified time series models that derive from theoretical optimization problems. Resulting conclusions about underlying theoretical parameters may be sensitive to imperfections in the data. We illustrate this fact by considering sampling error in data from the Census Bureau's Retail Trade Survey. We find that parameter estimates in seasonal time series models for retail sales are sensitive to whether a sampling error component is included in the model. We conclude that sampling error should be taken seriously in attempts to derive economic implications by modeling time series data from repeated surveys.

1. Introduction

The rational expectations revolution has transformed the methodology of macroeconometric research. In the new style, the researcher typically begins by specifying a dynamic optimization problem faced by agents in the model economy. Then the researcher derives the solution to the optimization problem, expressed as a stochastic model for some observable economic variable(s). A trademark of research in this tradition is that each of the parameters in the model for the observable variables is endowed with a specific economic interpretation. The last step in the research program is the application of the model to actual economic data to see whether the model

*This paper reports the general results of research undertaken by Census Bureau and Federal Reserve Board staff. The views expressed are attributed to the authors and do not necessarily reflect those of the Census Bureau or the Federal Reserve Board. We gratefully acknowledge the advice and computer code of Bill Cleveland, the excellent research assistance of Steve Helwig and John Golmant, and the support of Ruth Detlefsen, Michael Shimberg, and Carol Veum of the Census Bureau's Business Division, who provided data and information about the Retail Trade Survey. Any errors remain our responsibility.

0304-4076/93/$05.00

conforms with the data. The trend toward this style of research has been especially striking in the consumption literature.

Most investigators – especially those using aggregate data – have paid scant attention to possible shortcomings of available data series as measurements of the theoretical variables appearing in their models. Such shortcomings can have many sources, as we discuss later in section 2, and in detail for the particular case of retail sales data in appendix A.2. In principle, data imperfections could influence the results of empirical investigations, and thereby change the economic interpretation of the findings. Partly in reaction to this possibility, Wilcox (1991) investigated the source data and estimation methods used to construct aggregate U.S. consumption data. Wilcox discussed a number of deficiencies in the data and provided, for some of them, a preliminary assessment of their significance for empirical research on consumption-related issues.

This paper builds on that earlier effort. Here, we focus on data from the Retail Trade Survey conducted by the U.S. Bureau of the Census. Our interest in these data is motivated by two considerations: First, they are an important ingredient in the construction of personal consumption expenditures in the national income accounts; and second, sampling error in the retail sales estimates is one of the problems of potential significance identified in Wilcox (1991). Furthermore, there is some information available concerning the sampling error properties of disaggregated retail sales series.

Our objective is to assess the sensitivity of parameter estimates for time series models of retail sales data to the treatment of sampling error. We carry out our investigation using for illustration the 'airline' model of Box and Jenkins (1976):

$$(1-B)(1-B^{12})S_t=(1-\theta_1)(1-\theta_{12}B^{12})b_t, \tag{1.1}$$

where B is the backshift operator and b_t is a white-noise innovation. In appendix A.1 we give one set of economic assumptions consistent with the prediction that S_t – interpreted as the level of retail sales – should follow the model (1.1). The key assumption turns out to be that tastes follow the stochastic process

$$(1-B)(1-B^{12})T_t=\nu_t,$$

where ν_t is a white-noise shock to tastes. We show in appendix A.1 that each of the parameters in (1.1) can be given a precise interpretation in terms of the underlying economic model we present there. In particular, the innovation variance σ_b^2 measures the variance of 'news', where news involves the shock to tastes as well as the shock to labor income. The nonseasonal MA parameter θ_1 equals one minus the depreciation rate of the durable good – the same interpretation of this parameter as obtained in the model studied by Mankiw (1982). The seasonal MA parameter θ_{12} is a function of the ratio of σ_ν^2, the variance of the innovations in tastes, to σ_ε^2, the variance of the innovations in labor income. As this variance ratio goes to zero, θ_{12} approaches unity, the seasonal MA polynomial cancels with the seasonal differencing operator [see Bell (1987)], and the model simplifies to

$$(1-B)S_t = \text{seasonal dummies} + (1-\theta_1 B)b_t. \tag{1.2}$$

Therefore, in the context of the model set out in the appendix A.1, we would interpret evidence derived from levels data that θ_{12} is close to unity as suggesting that tastes follow a deterministic seasonal pattern.

In deference to properties of the data discussed in sections 3 and 5, we ignore that our economic model, taken literally, applies to the data in levels, and interpret S_t in (1.1) as referring to the logarithm of retail sales in a particular category. In concrete terms, our goal is to investigate the robustness of estimates of θ_1, θ_{12}, and σ_b^2 in an airline model for $\log(S_t)$ to the presence of sampling error in the data. Our strategy for carrying out this investigation is simple: we apply model (1.1) to the (logarithms of the) retail sales data twice, once augmenting it with an explicit model for the sampling error in the retail sales data, and once not. We then examine the sensitivity of the point estimates of each of the above-mentioned parameters of the signal model to the presence or absence of the sampling error component in the overall model for the observed series. We focus on changes in point estimates because of the difficulties involved in making formal probability statements about parameters (particularly θ_{12}) that are subject to boundary constraints, and, more importantly, because of the difficulties in developing inferences that allow for uncertainty in the estimation of the sampling error model.

We should stress that we are not wedded to the airline model either from a statistical or an economic point of view. From a statistical perspective, we believe that the model's capability of allowing a changing pattern of seasonality is useful; however, there are other ways to approach the general problem of changing seasonality, and an alternative approach might ultimately prove more successful than the airline model. As far as the economics are concerned, we acknowledge that other models undoubtedly could be proposed that would endow the parameters of the airline model with different interpretations than the ones we give. Furthermore, other underlying economic structures could, and have been, proposed that imply that spending should take some form other than the airline model. In any such models, however, the issue we study here would be relevant: how robust are the point estimates of the parameters – and hence the economic interpretation of the results – to errors in the data? In our view, the paper should not be seen as hostile to the predominant research approach in macroeconometrics of the past two decades; on the contrary, we are suggesting that explicit modeling of the sampling error may be a step toward placing the inferences drawn from such an approach on a firmer empirical foundation.

The paper is organized as follows. Section 2 provides some background information on repeated economic surveys. Details of the Census Bureau's monthly Retail Trade Survey are deferred to appendix A.2. We use the information given there as a guide in the construction of seasonal time series models for the sampling errors in retail sales estimates in section 3. An analytical exercise in section 4 shows that the sampling errors can have important effects on the seasonal and nonseasonal autocorrelation properties of the observed time series in comparison to those of the true, unobserved (signal) series. In section 5 we analyze seven retail sales time series (monthly sales of grocery stores, eating places, household appliance stores, men's and boys' clothing stores, hardware stores, radio and TV stores, and drinking places). We first fit airline models (with appropriate regression terms for calendar variation and outliers) to the series ignoring the sampling errors, and then refit the models with a sampling error component included, performing this analysis both for nominal and real data. We find that estimates of θ_{12} and σ_b^2 in some cases are profoundly affected even by seemingly moderate amounts of sampling error, but find little effect on estimates of θ_1 for any of the series. Section 6 provides a summary and conclusions.

2. Background

Many published economic time series are estimates derived from repeated surveys. Examples include, but are by no means limited to, time series of retail and wholesale sales and inventories; building permits issued and housing starts; manufacturers' shipments, inventories, and new orders; unemployment statistics; price indexes; and statistics on imports and exports. Many other macroeconomic series are aggregates derived from such series as these. Such data are subject to two general types of errors: *sampling errors*, which result when the estimates are obtained from a sample survey, rather than a complete census, of the relevant universe (of firms, households, etc.), and *nonsampling errors*, which are all other errors. The latter can often be thought of as biases, and include such things as definitional errors, reporting errors, nonresponse errors, sampling frame undercoverage, processing errors, etc. Government agencies conducting repeated economic surveys attempt to minimize nonsampling errors, and also perform special studies to try to assess their magnitude. However, nonsampling errors remain extremely difficult to handle statistically, particularly for those analyzing published data. It is for this reason, and not because they are unimportant, that we will not deal with them in this paper.

There are series (for example, import and export statistics and retail sales of department stores) that are obtained from complete or essentially complete repeated censuses, and thus contain no sampling error, though they certainly are subject to nonsampling errors. Thus, the amount of sampling error in published time series ranges upward from zero, though government statistical agencies tend to avoid publishing series subject to very high levels of sampling error. Fortunately, sampling error is amenable to statistical treatment; in fact, government statistical agencies regularly publish estimates of sampling error variances. In regard to sampling error, therefore, analysis of published time series would seem to present a classical errors-in-variables problem, with the desirable situation of having known, or at least estimated, error variances.

There is, unfortunately, a catch. The catch is that sampling errors from repeated surveys are often correlated over time. In general, sampling errors in repeated surveys can be autocorrelated if (1) time series for individual population units are autocorrelated and (2) samples at different time points are not drawn independently, e.g., if they have specified overlapping segments. Drawing samples independently each time period is generally infeasible operationally for many reasons, including the cost involved in doing this. Instead, repeated survey designs typically use samples that overlap for different time periods. If overlap persists only for some finite number, say q, of time periods, and the nonoverlapping samples are drawn independently, then the sampling errors will follow a moving average model of maximum order q. Hausman and Watson (1985) used this idea in developing a model for sampling errors in the Current Population Survey (CPS), though nonoverlapping samples in the CPS are not quite drawn independently.

The amount of autocorrelation present in sampling errors depends on the amount of sample overlap and the autocorrelation inherent in the individual units in the population. It can also depend on the estimation scheme used if the estimates use both past and present data to estimate current values, as in composite estimation in the Retail Trade Survey [Wolter (1979)]. Unfortunately, available information on autocorrelation of sampling errors for particular economic surveys is spotty. Government agencies do not regularly produce estimates of lagged covariances of sampling errors, though this has been done on occasion as part of special studies. This includes work done previously by the Census Bureau that underlies the sampling error models developed in the next section for some retail sales time series. To cite a few other examples: Bell and Hillmer (1990a, b) consider regional and national single and five or more unit housing starts time series, for which estimates produced by the Construction Statistics Division of the Census Bureau suggested the sampling errors exhibit at most very mild correlation at lag 1. They also consider a time series of teenage unemployment from the CPS, using an ARMA (1, 1) model to approximate the autocorrelation structure of the sampling errors. Train, Cahoon, and Makens (1978) produced the original estimates of sampling error autocorrelations for this and some other series from the CPS. Tiller (1990) discusses some preliminary work of Dempster and Miller to estimate autocorrelations for state level CPS data. These are, unfortunately, isolated examples, in part because information on the autocorrelation structure of sampling errors in data from economic surveys is not regularly produced. This situation may improve in the near future, as in recent years there has been more interest in this within government statistical agencies. The interest has been stimulated mostly by research into use of time series signal extraction techniques to remove some of the sampling error from estimates in repeated surveys, an idea originally suggested by Scott and Smith (1974) and Scott, Smith, and Jones (1977), and pursued more recently by Binder and Dick (1989), Bell and Hillmer (1990a, b), Eltinge and Fuller (1989), and Pfeffermann (1991).

3. Modeling sampling errors in the RTS

Bell and Hillmer (1990a), hereafter BH, develop a time series model for the sampling errors in (unbenchmarked) estimates of sales of eating places and of drinking places from the RTS. We briefly review the model developed in BH, apply it to some additional store categories from the RTS, and then consider some limitations of the model. Appendix A.2 should be consulted for the basics of the survey that are used here. (The references cited in appendix A.2 give further details on the RTS.) Given some of the problems noted in appendix A.2, and some additional problems mentioned in what follows, the models can only be regarded as crude approximations that are hoped to capture the most important aspects of the sampling error autocovariance structure. Despite their limitations, use of such models seems preferable to ignoring the sampling error altogether.

BH begin with a time series model for the sampling errors in the current- and previous-month Horvitz–Thompson (HT) estimates (see appendix A.2) in the RTS. Let $Y_t' = S_t + e_t'$ be the current-month (t) HT estimate and $Y_{t-1}'' = S_{t-1} + e_{t-1}''$ be the previous-month ($t-1$) HT estimate, where S_t is the unobservable true (signal) series and e_t' and e_{t-1}'' are the sampling errors in the HT estimates. BH use identical models for both e_t' and e_{t-1}''. Estimates of $\rho \equiv \operatorname{corr}(e_t', e_{t-1}'')$ are extremely high – typically exceeding 0.95. These values are probably inflated by businesses sometimes reporting the same figure for current- and previous-month sales, and perhaps also by the imputation procedures used (as noted in appendix A.2); in any case, given the current survey estimation procedures it is difficult to distinguish characteristics of e_t' from those of e_{t-1}''. BH use the model

$$(1-\phi^m B^m)(1-\Phi B^{12})e_t' = \nu_{1t}, \tag{3.1}$$

where m is the number of panels in the sample – currently three (four prior to September 1977). The same model is assumed for e''_{t-1} with $\nu_{2,t-1}$ replacing ν_{1t}. $[\nu_{1t}\nu_{2,t-1}]^{\mathrm{T}}$ (where superscript T indicates transpose) is assumed to be bivariate white noise with common variance σ_ν^2 for ν_{1t}. This model assumes independence of estimates from different panels, since it allows for autocorrelation in $[e'_t e''_{t-1}]^{\mathrm{T}}$ only at lags that are multiples of m. The Φ parameter allows for additional correlation at seasonal lags. (3.1) has a convenient property we shall use shortly: if the sampling error in each panel would follow (3.1) with $m = 1$ if the panel were observed every month, then for any number m (that is a divisor of 12) of rotating independent panels reporting successively, e'_t follows (3.1).

To estimate (3.1) we require estimates of lag correlations for e'_t and e''_{t-1}. While estimates of lag covariances and correlations are not regularly produced for the RTS, this was done as part of a special study using data from January 1973 through March 1975, a time when the survey had four rotating list panels. Estimates of such lag correlations can be averaged over time assuming correlation stationarity. Table 1 shows averaged correlations at lags 4, 8, 12, 16, 20, and 24 for e'_t and e''_{t-1} [the averaging was done after applying Fisher's transformation $0.5\log((1+r)/(1-r))$, and then the inverse transformation was applied to the results] for seven series from the RTS. Table 1 shows that the sampling errors exhibit strong positive autocorrelation and evidence of seasonality from the increase in the correlations at lags 12 and 24, justifying the $(1-\Phi B^{12})$ term in (3.1). We estimated ϕ and Φ by minimizing the weighted sum of squared deviations of the correlations from (3.1) with $m = 4$, from those of table 1. (Lags 20 and 24 were ignored, and lag 16 given a weight of 0.5, due to the small number of correlation estimates that were averaged together at these higher lags.) This procedure actually estimates ϕ^4, but assuming $\phi > 0$ estimates of ϕ and ϕ^3 can be computed directly for use with the three-panel samples. The results for $\hat{\phi}^3$ and $\hat{\Phi}$ are given in table 2. The correlations resulting from (3.1) with $\hat{\phi}^4$ and $\hat{\Phi}$ are also shown in table 1, as an indication of the goodness-of-fit of the model. On the whole, the averaged correlations at the lags shown are reasonably well matched by those from the model, though the empirical correlations at lags 20 and 24 appear to die out somewhat less rapidly than the model predicts.

BH then derive a model for the sampling errors of the linear composite estimator [Wolter (1979)], which is given by

$$Y'''_t = (1-\beta)Y'_t + \beta(Y'''_{t-1} + Y'_t - Y''_{t-1}) \quad \text{(preliminary estimator)},$$
$$Y_{t-1} = (1-\alpha)Y''_{t-1} + \alpha Y'''_{t-1} \quad \text{(final estimator)}. \quad (3.2)$$

In the RTS values of $\alpha = 0.8$, $\beta = 0.75$ have been used with the three-panel samples, and values of $\alpha = 0.82$, $\beta = 0.8$ were used with the previous four-

Table 1

Sampling error correlations for Horvitz–Thompson estimates.[a,b]

	Lag					
	4	8	12	16	20	24
	Grocery stores (SIC 541)					
Averaged	0.86	0.85	0.92	0.82	0.81	0.88
From (3.1)	0.87	0.85	0.93	0.80	0.76	0.83
	Eating places (SIC 5812)					
Averaged	0.72	0.71	0.79	0.63	0.65	0.77
From (3.1)	0.75	0.69	0.81	0.60	0.53	0.61
	Household appliance stores (SIC 5722)					
Averaged	0.85	0.81	0.86	0.78	0.74	0.75
From (3.1)	0.85	0.82	0.88	0.74	0.68	0.72
	Men's and boys' clothing stores (SIC 561)					
Averaged	0.75	0.75	0.77	0.72	0.63	0.76
From (3.1)	0.80	0.74	0.81	0.63	0.56	0.60
	Radio and TV stores (SIC 5732)					
Averaged	0.74	0.71	0.86	0.66	0.65	0.81
From (3.1)	0.75	0.71	0.87	0.64	0.59	0.72
	Drinking places (SIC 5813)					
Averaged	0.70	0.67	0.78	0.60	0.60	0.61
From (3.1)	0.72	0.66	0.80	0.56	0.50	0.59
	Hardware stores (SIC 525)					
Averaged	0.76	0.73	0.85	0.74	0.73	0.76
From (3.1)	0.78	0.74	0.87	0.67	0.62	0.72
	Number of correlations of both Y_t' and Y_{t-1}'' averaged					
	23	19	15	11	7	3
	Weights used in determining $\hat{\phi}, \hat{\Phi}$					
	1	1	1	0.5	0	0

[a]Raw estimates of $\operatorname{corr}(e_t', e_j')$ and $\operatorname{corr}(e_{t-1}'', e_{j-1}'')$ available for all pairs of months from January 1973 through March 1975 were averaged by lag for the lags shown after applying Fisher's transformation, and the results were then transformed back.

[b]Correlations are shown from model (3.1) for $m = 4$ with parameter values $\hat{\phi}^4$ and $\hat{\Phi}$ (given in table 2) determined to minimize the weighted sum of squared deviations of the correlations from the model and the averaged correlations using the weights shown.

panel sample. Note that (3.2) also holds for the sampling errors, i.e., with Y replaced by e. BH then use (3.1) and (3.2) to derive the following expression for e_t, the sampling error in the final composite estimates Y_t:

$$(1-\beta B)(1-\phi^m B^m)(1-\Phi B^{12})e_t = (1-\alpha)\nu_{2t} - \beta\nu_{2,t-1} + \alpha\nu_{1t}. \tag{3.3}$$

Table 2
Parameter estimates for (3.2) and related quantities.[a,b]

						Coefficients of variation (%)	
	$\hat{\phi}^3$	$\hat{\Phi}$	$\hat{\eta}$	$\hat{\sigma}_c^2 \times 10^6$	$\hat{\rho}$	HT	Composite
Grocery stores	0.77	0.86	−0.19	1.34	0.994	2.2	1.3
Eating places	0.69	0.72	−0.13	19.48	0.985	4.2	2.5
Household appl. stores	0.77	0.77	−0.10	40.41	0.979	7.8	5.1
Men's and boys' clothing stores	0.73	0.69	−0.02	41.65	0.953	5.1	3.6
Hardware stores	0.70	0.80	−0.04	54.64	0.962	7.0	4.8
Radio and TV stores	0.67	0.81	−0.10	87.60	0.979	9.8	6.1
Drinking places	0.66	0.71	−0.13	93.01	0.986	8.8	5.1

[a]Values for $\hat{\phi}^4$ and $\hat{\Phi}$ were obtained as discussed in the text and in table 1. Then $\hat{\phi}$ and $\hat{\phi}^3$ were obtained assuming $\phi > 0$. The $\hat{\eta}$ and $\hat{\sigma}_c^2$ values were obtained as described in the text. We calculated the coefficients of variation of the composite estimators using the fitted models (3.4); these differ some from published estimates (see text).

[b]Values for $\hat{\rho}$ were obtained by averaging transformed values of $\text{corr}(e_t', e_{t-1}'')$ and then transforming back, and values of the relative variance of the HT estimates were obtained by averaging values of $\text{var}(e_t')/(Y_t')^2$ and $\text{var}(e_{t-1}'')/(Y_{t-1}'')^2$. For all but eating places, drinking places, and hardware stores, data for 1989 were used in the averaging of the lag 1 correlations and the relative variances. For these other three series, data for 1982 through 1986 were used. Also, for these series the logarithms of the relative variance estimates were averaged, added to one half of the sample variance of these, and this was then exponentiated to get the relative variances of the HT estimates. This produced only slightly different results than simply averaging the relative variances.

The right-hand side above is a first-order moving average process whose variance and lag 1 autocovariance can be determined for given values of α, β, σ_ν^2, and $\text{corr}(\nu_{1t}, \nu_{2,t-1})$. From these the corresponding moving average parameter and innovation variance can be obtained. We can use (3.1) to get σ_ν^2 for given values of ϕ, Φ, and $\text{var}(e_t') = \text{var}(e_{t-1}'')$, and we assume that $\text{corr}(\nu_{1t}, \nu_{2,t-1}) = \rho$, at least approximately, which is justified in BH.

Contrary to the above model, however, estimates of sampling variance, $\widehat{\text{var}}(e_t')$ and $\widehat{\text{var}}(e_{t-1}'')$, for retail sales series are highly dependent on the level of the series. Since $\widehat{\text{var}}(e_t')/Y_t'$ and $\widehat{\text{var}}(e_{t-1}'')/Y_{t-1}''$, the estimates of *relative* variance, are much more stable over time, BH turn to a multiplicative decomposition, $Y_t = S_t \cdot u_t$, where $u_t = 1 + e_t/S_t$ is the multiplicative sampling error in Y_t, and $\text{var}(\log(u_t))$ approximately equals $\text{var}(e_t/S_t)$, the relative sampling variance of Y_t, for small e_t/S_t. They then assume that a model of the form derived for e_t actually holds for $\log(u_t)$ in $\log(Y_t) = \log(S_t) + \log(u_t)$. For three-panel samples, the model resulting for $\log(u_t)$ is

$$(1 - 0.75B)(1 - \phi^3 B^3)(1 - \Phi B^{12})\log(u_t) = (1 - \eta B)c_t, \qquad (3.4)$$

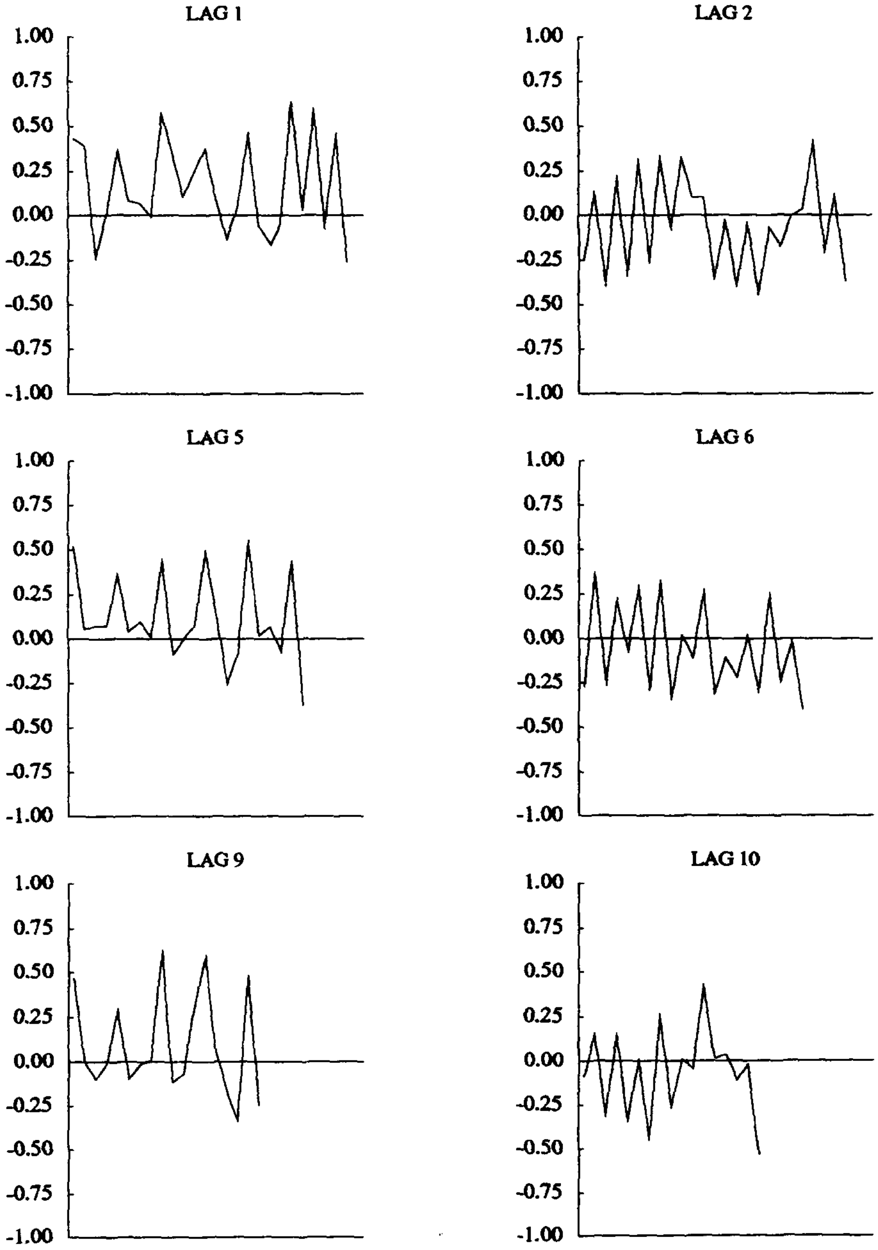

Fig. 1. Estimated autocorrelations for the sampling error; hardware stores.

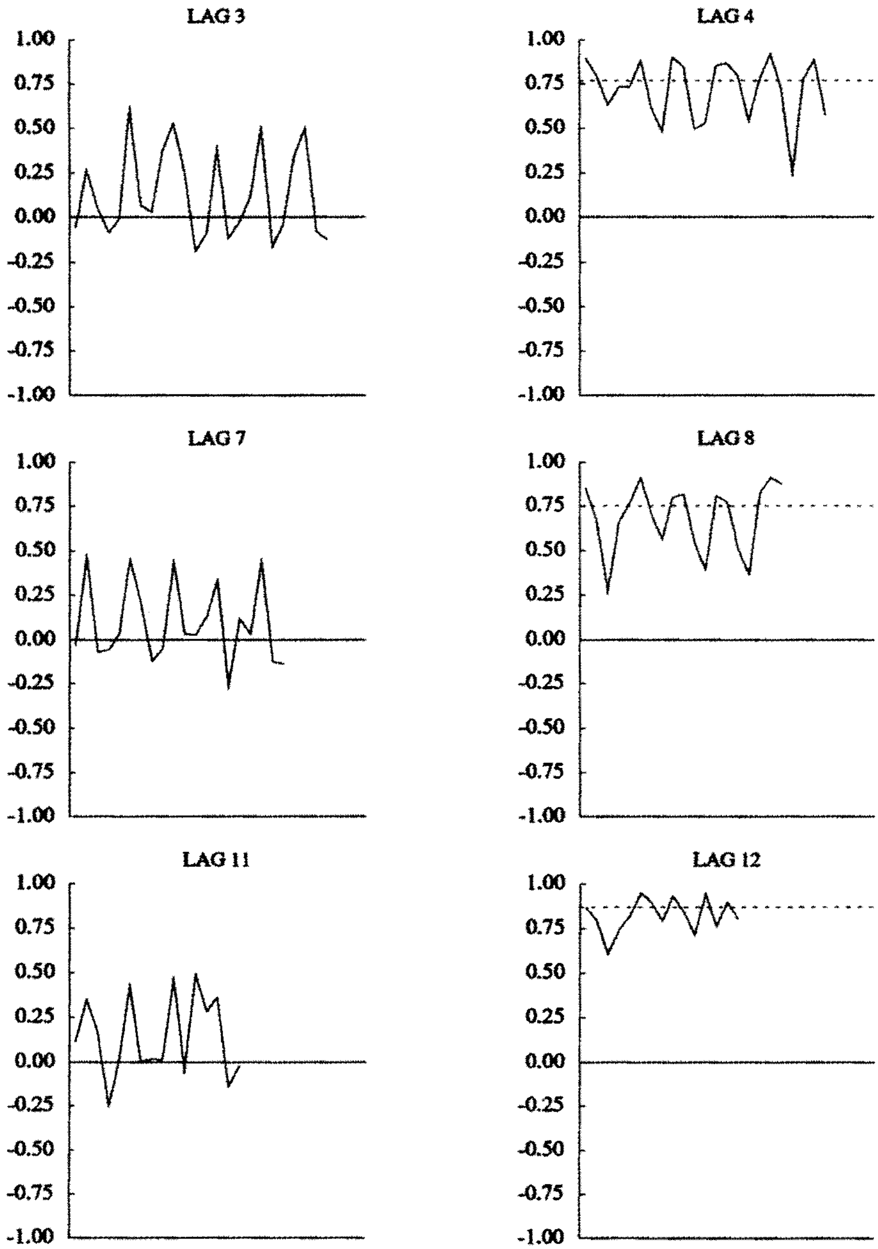

Fig. 1 (continued)

where c_t is white noise. We use our estimates of ϕ, Φ, ρ and of the *relative* sampling variances of Y_t' and Y_{t-1}'' to develop estimates of the variance and lag 1 autocovariance of the analog of the right-hand side of (3.3) for $\log(u_t)$, and hence obtain estimates of η and σ_c^2. Estimates $\hat{\rho}$, Y_t', Y_{t-1}'', $\widehat{\text{var}}(e_t')$, and $\widehat{\text{var}}(e_{t-1}'')$ are regularly produced in the RTS. We used estimates of these quantities for the twelve months of 1989 for all the series except eating places, drinking places, and hardware stores, for which we already had estimates for 1982 through 1986. We averaged the estimates of relative variance and ρ over time (the latter after applying Fisher's transformation), producing the results shown in table 2 under the column headings 'HT' and '$\hat{\rho}$', respectively. (Instead of the relative variance of the HT estimates, we show their square root, the coefficient of variation or CV.) Using these results, we solved for the estimates $\hat{\eta}$ and $\hat{\sigma}_c^2$ given in table 2. We then solved for $\text{var}(\log(u_t))$ in (3.4), which is an estimate of the relative variance of the final composite estimator. The square roots of these quantities, the CV's of the final composite estimates, are also shown in table 2, and are in some cases somewhat lower than, but overall are reasonably close to, published estimates that are obtained more directly.

We now mention some shortcomings of the sampling error model. First, the model (3.1) assumes zero correlation between sampling errors in the HT estimates from different panels, which means the only correlations expected to be nonzero are at lags that are multiples of the number of panels m. We had available estimates of HT sampling error correlations for all pairs of months from January 1973 through March 1975 from the four-panel survey, and so could examine whether correlations at lags not multiples of four appeared to be zero. The results were not uniformly encouraging. Fig. 1 shows plots of the correlation estimates, chronologically by lag for lags 1 through 12, for hardware stores, one of the discouraging examples. The solid line in each plot shows estimates of $\text{corr}(e_t', e_j')$, and the dotted line in the plots at lags 4, 8, and 12 shows the correlations implied by the fitted models given in table 1. [We also examined estimates of $\text{corr}(e_{t-1}'', e_{j-1}'')$. On a plot these were almost indistinguishable from the estimates of $\text{corr}(e_t', e_j')$, and so are not shown here.] The correlations in fig. 1 exhibit some substantial deviations from zero at several lags other than 4, 8, and 12. A possible explanation for this is the effect of monthly noncertainty cases discussed in appendix A.2. Also, many of the lag correlation estimates in fig. 1 exhibit periodic behavior with period 4, often with substantially larger periodic oscillations than the deviation of the average of the correlations from zero. This suggests either that correlation estimates for different panel pairs may not be estimating the same quantity, or that there is a very high correlation between correlation estimates for the same panel pairs, a hypothesis we are unable to assess. Other than this, we presently have no explanation for this phenomenon.

Another slight problem with the model is that the composite estimates are not used at the beginning of a new sample – Wolter (1979) mentions that the (approximate) minimum variance linear unbiased estimates are used for the first three months of a new sample. This introduces a transient effect into the autocorrelations that we shall ignore.

In the following sections we shall use the model (3.4) with parameters given in table 2 for the composite estimate sampling errors in the three-panel samples. Also, as discussed in appendix A.2, because new three-panel samples were drawn independently in September 1977, January 1982, and January 1987, we assume independence of sampling errors from these different samples. While it would be preferable in some respects to use time series data from the four-panel survey, for which the sampling error correlation estimates were directly obtained, efforts we made to model the seven years of data from the four-panel survey (September 1970 through August 1977) yielded rather unstable results, suggesting to us that this was not a long enough stretch of data to support the type of models we were fitting. We would prefer instead to use direct estimates of sampling error lagged covariances in the three-panel survey in developing the sampling error models. Unfortunately, as noted earlier, such estimates are not currently available. We hope we can eventually obtain lagged covariance estimates for the three-panel survey. If so, we can better assess the suitability of the model (3.4), and also try to modify the model (3.4) as seems necessary.

4. Effects of sampling errors on autocorrelation properties of observed series

Consider the additive decomposition $Y_t = S_t + e_t$, where S_t follows the airline model $(1-B)(1-B^{12})S_t = (1-\theta_1 B)(1-\theta_{12}B^{12})b_t$, and the sampling errors follow the model (3.4) but with $\log(u_t)$ replaced by e_t. (We consider the additive decomposition here merely for simplicity of notation, we could similarly use logarithms of a multiplicative decomposition.) In this section we investigate analytically the possible effect of the sampling errors on how the autocorrelation properties of the observed series Y_t differ from those of S_t. We use this exercise as a guide to what might be expected from the empirical results later. Since Y_t requires the same differencing operator, $\nabla\nabla_{12} = (1-B)(1-B^{12})$, as S_t, we let $W_t = \nabla\nabla_{12}Y_t = v_t + \xi_t$, where $v_t = \nabla\nabla_{12}S_t = (1-\theta_1 B)(1-\theta_{12}B^{12})b_t$, and $\xi_t = \nabla\nabla_{12}e_t$ follows the model

$$\begin{aligned}&(1-0.75B)(1-0.7B^3)(1-0.75B^{12})\xi_t\\&=(1+0.1B)(1-B)(1-B^{12})c_t.\end{aligned} \tag{4.1}$$

The values of $\hat{\phi}^3$, $\hat{\Phi}$, and $\hat{\eta}$ in table 2 do not vary greatly over series; in (4.1) we have picked roughly the average parameter values from table 2. The

Table 3a

Autocorrelations from (4.1) with $\phi^3 = 0.7$, $\Phi = 0.75$, $\eta = -0.1$.

Lag	1	2	3	4	5	6	7	8	9	10	11	12
$\rho_\xi(k)$	−0.24	−0.26	0.60	−0.21	−0.21	0.35	−0.15	−0.15	0.18	−0.10	−0.09	.03
Lag	13	14	15	16	17	18	19	20	21	22	23	24
$\rho_\xi(k)$	−0.06	−0.05	0.01	−0.02	−0.02	−0.01	0.00	0.00	−0.05	0.02	0.02	−0.10
Lag	25	26	27	28	29	30	31	32	33	34	35	36
$\rho_\xi(k)$	0.03	0.03	−0.07	0.03	0.03	−0.07	0.04	0.04	−0.08	0.04	0.04	−0.10

model (4.1) reflects overdifferencing [the $(1-B)$ and $(1-B^{12})$ moving average operators] since e_t following (3.4) does not require differencing.

Let $\gamma_w(k)$ and $\rho_w(k)$ denote the autocovariance and autocorrelation of W_t at lag k, with analogous notation for the other series involved. Then one can easily see that

$$\rho_w(k) = \lambda\rho_v(k) + (1-\lambda)\rho_\xi(k), \qquad \lambda = \gamma_v(0)/\big(\gamma_v(0) + \gamma_\xi(0)\big).$$

Thus, the autocorrelation function of the differenced observed series will be a weighted average of those of the differenced signal series and the differenced sampling error. The averaging parameter, λ, depends on the variance in the signal series relative to that in the observed series; it can also be expressed as a function of the signal-to-noise ratio, σ_b^2/σ_c^2, and the parameters θ_1 and θ_{12}. Table 3a gives the autocorrelations of ξ_t arising from (4.1). This shows positive correlations damping out at lags 3, 6, and 9, a correlation at lag 12 of near zero, and small negative correlations at lags 1, 2, 4, 5, 7, 8, 10, 11, 13, and 14 that tend to decrease in magnitude with increasing lag. Autocorrelations beyond lag 14 are mostly close to zero; those at lags 24 and 36 are the largest of these in magnitude. The autocorrelations of W_t are the result of averaging these together with those of v_t, which will be negative at lags 1 and 12 (for positive values of θ_1 and θ_{12}), the product of $\rho_v(1)$ and $\rho_v(12)$ (and so smaller and positive) at lags 11 and 13, and zero at all other lags.

Autocorrelations of W_t were computed for the above models for $\theta_1 = (0, 0.3, 0.9)$ and $\theta_{12} = (0.7, 1.0)$. This was done for a case of low sampling error (a moderately high signal-to-noise ratio of $\sigma_b^2/\sigma_c^2 = 10$) and for a case of high sampling error (a low signal-to-noise ratio of $\sigma_b^2/\sigma_c^2 = 3$). The results are shown as solid lines in figs. 2a and 2b, which also show bar graphs of the autocorrelations of v_t. The presence of sampling error shrinks the spikes in $\rho_v(12)$ towards zero and decreases the values of $\rho_v(11)$ and $\rho_v(13)$. The effects on $\rho_v(1)$ depend on the value of θ_1: for $\theta_1 = 0.9$ the strong negative $\rho_v(1)$ is pulled towards zero; for $\theta_1 = 0$ we have $\rho_v(1) = 0$ but a fraction

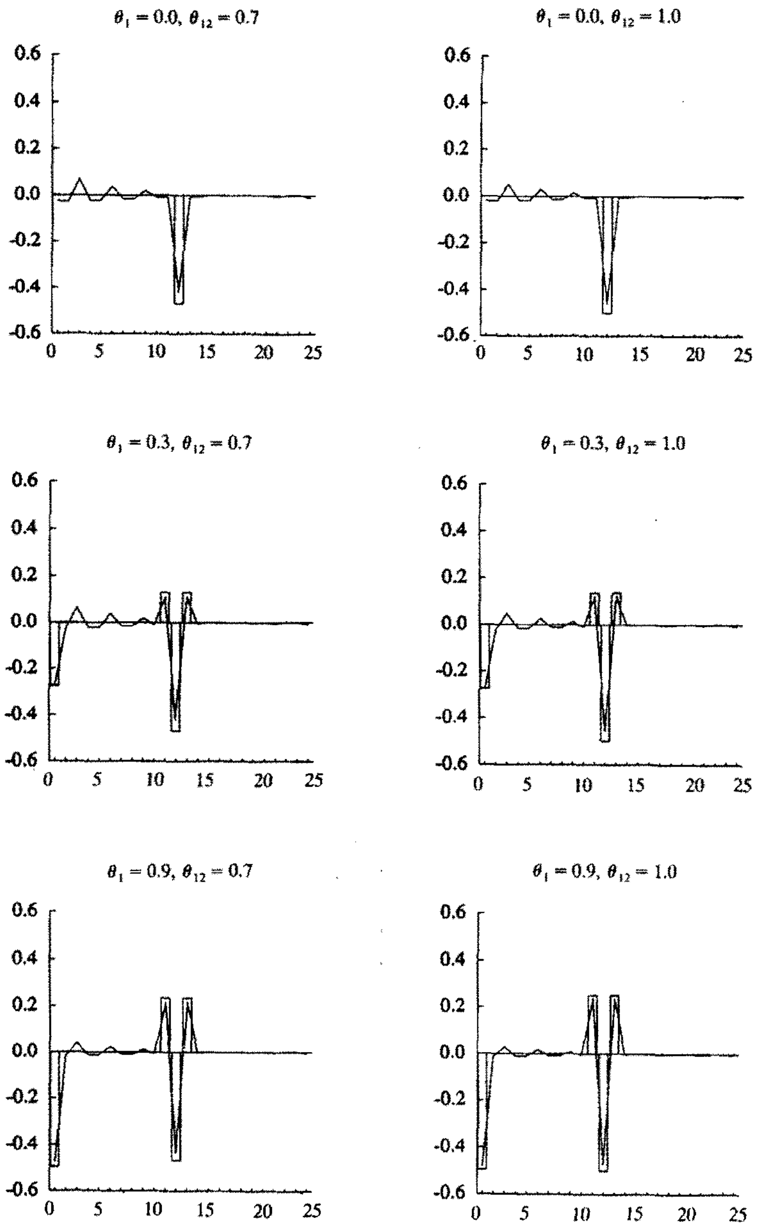

Fig. 2a. Low sampling error.

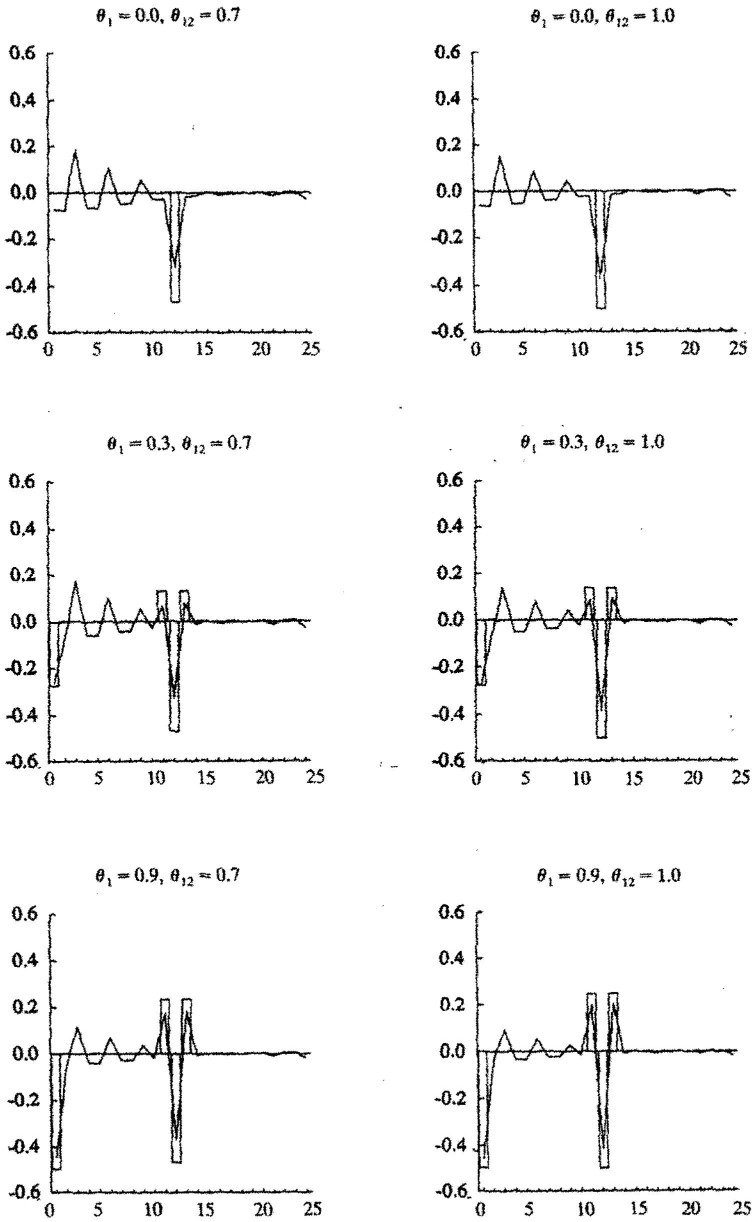

Fig. 2b. High sampling error.

Table 3b

Airline model parameters $(\tilde{\theta}_1, \tilde{\theta}_{12})$ producing the best approximation to autocorrelations of the differenced data, W_t, in the presence of (differenced) sampling error following (4.1).

	Low sampling error		High sampling error	
	θ_{12}		θ_{12}	
θ_1	0.7	1.0	0.7	1.0
0	(0.02, 0.52)	(0.01, 0.63)	(0.05, 0.35)	(0.04, 0.44)
0.3	(0.29, 0.53)	(0.29, 0.64)	(0.28, 0.36)	(0.28, 0.45)
0.9	(0.76, 0.59)	(0.78, 0.72)	(0.62, 0.45)	(0.66, 0.55)

$(1-\lambda)$ of the negative correlation in ξ_t at lag 1 filters through to W_t; and $\theta_1 = 0.3$ is approximately a stationary point, with $\rho_w(1)$ very close to $\rho_v(1)$. At other lags $\rho_w(k) = (1-\lambda)\rho_\xi(k)$. Still, even in the high sampling error case, values of $\rho_w(k)$ for $k \neq 1, 11, 12, 13$ are less than 0.2 in magnitude, and someone identifying a time series model from $\rho_w(k)$ could easily pick an airline model for Y_t. The most likely alternative choice would be to augment this model with additional low-order MA or AR lags, e.g., an $(0,1,3) \times (0,1,1)_{12}$ model instead of the airline model.

To get an idea of what might happen if one were to fit an airline model directly to Y_t in this situation, we picked values of θ_1 and θ_{12} to approximate $\rho_w(k)$ for $k = 1, 11, 12, 13$, in the sense of minimizing the sum of squares of the deviations of the airline model correlations from the $\rho_w(k)$. The results are given in table 3b. As might be expected from figs. 2a and 2b, the presence of sampling error as modeled here biases up the estimate of θ_1 when the true θ_1 equals zero, biases it down when the true value is large, and has virtually no effect for intermediate values (i.e., θ_1 near 0.3). The impact of the sampling error on the estimate of θ_1 is much greater when the true θ_1 is large. The effects of sampling error on estimates of θ_{12} are a dramatic shrinking of the value towards zero in virtually all cases considered. Of course, the innovation variance (not shown in the table) is smaller when the model allows for sampling error, and so attributes some of the variability to this component. In the next section we shall see to what extent these effects show up empirically in fitting models with and without sampling error components to observed retail sales series.

5. Estimating time series models for retail sales

We now estimate models with and without sampling error components for our seven retail sales time series. Plots of the (unbenchmarked) final composite estimates Y_t for these series are given in fig. 3. (These are nominal series; we shall later also analyze constant dollar series.) Our data begins with the

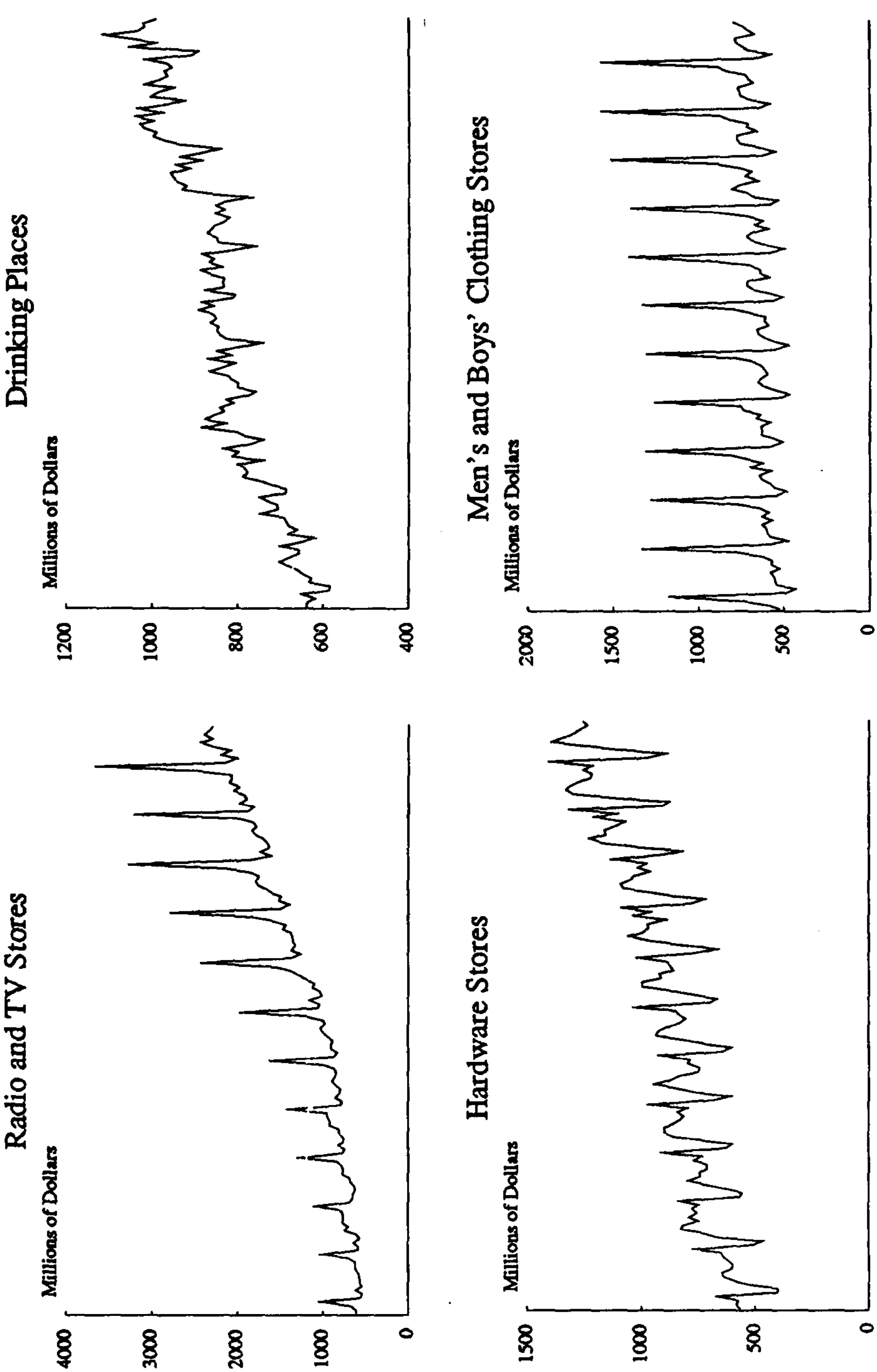

Fig. 3. Monthly final composite estimates (unbenchmarked).

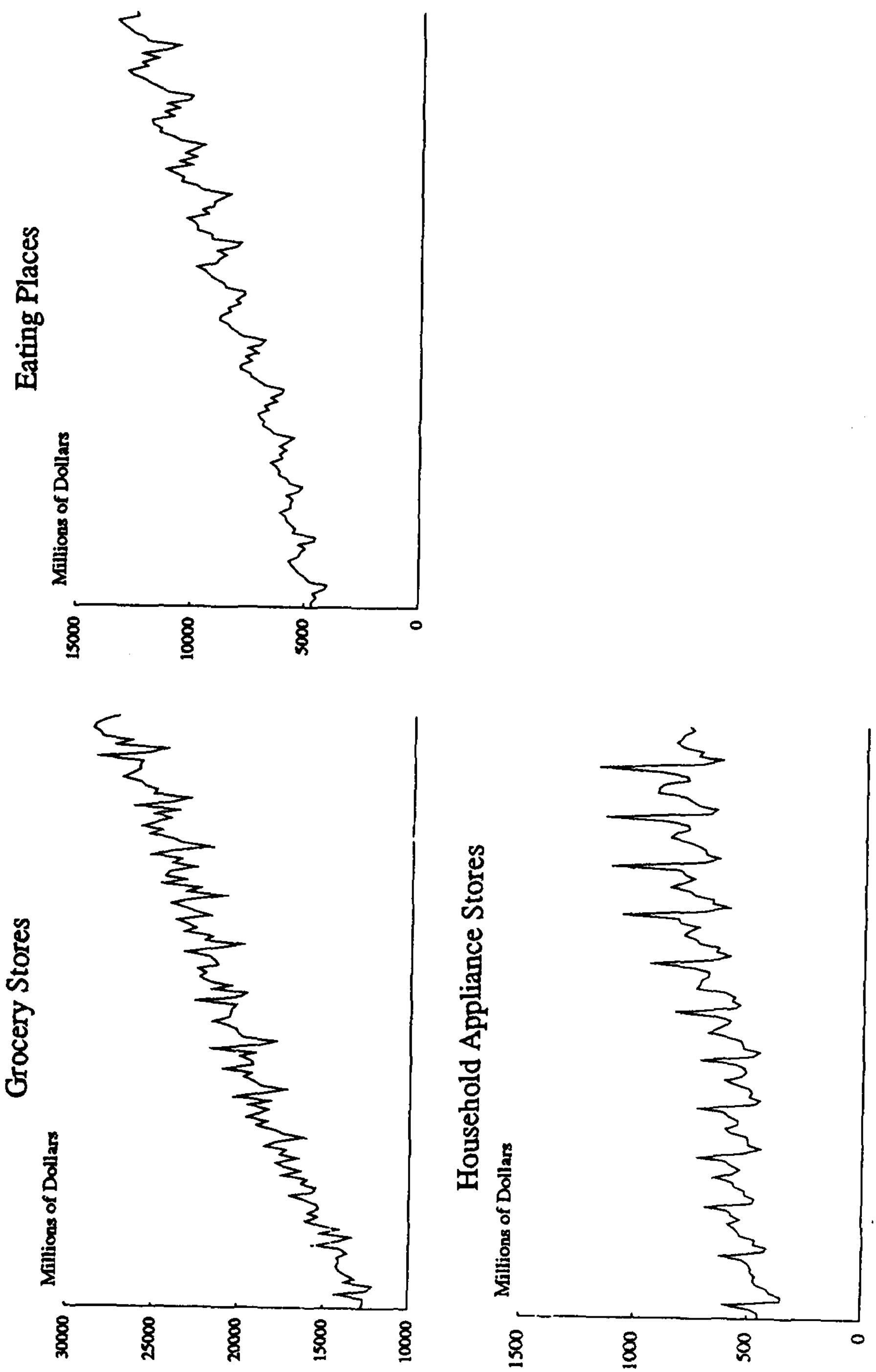

Fig. 3 (continued)

start of the three-panel survey, in September 1977, and ends in October 1989, yielding 146 observations. There were two redrawings of the RTS sample during our observation period: in January 1982 and January 1987.

As fig. 3 makes clear, all seven of our series show strong seasonality, nonstationarity in level, and variability increasing with increasing level, suggesting a need for nonseasonal and/or seasonal differencing, and the transformation $\log(Y_t)$. Taking logarithms of the multiplicative decomposition $Y_t = S_t \cdot (1 + e_t/S_t) = S_t \cdot u_t$ to get $\log(Y_t) = \log(S_t) + \log(u_t)$ is also convenient in terms of the properties of the sampling error, as noted in section 3. BH note that because the Y_t are 'design-unbiased' estimates (unbiased over repeated realizations of the sample), the time series $\log(S_t)$ and $\log(u_t)$ are approximately uncorrelated. All the time series here are known to be affected by trading-day variation, and two of them (grocery stores and men's and boys' clothing stores) by Easter holiday variation. We handle these effects with regression variables as discussed in Bell and Hillmer (1983), with a nine-day Easter effect for grocery stores and a seven-day Easter effect for men's and boys' clothing stores. We also include indicator variables in the models for a few outliers found in the series, using essentially the scheme of Bell (1983); the basic outlier detection methodology is discussed in more detail in Chang, Tiao, and Chen (1988). We did not find many outliers, and those found were not inordinately large in magnitude. For the sake of brevity, we shall omit estimates of regression parameters in what follows. In general, the estimates of the regression parameters in our models are little affected by the presence or absence of a sampling error component in the model.

As discussed earlier, we use the airline model for the signal series with regression effects removed, i.e.,

$$\nabla\nabla_{12}[\log(S_t) - \text{regression terms}] = (1 - \theta_1 B)(1 - \theta_{12} B^{12}) b_t,$$

or we ignore the sampling error and apply this model directly to $\log(Y_t)$. This choice of model can be justified empirically as follows. Sample autocorrelations of $\log(Y_t)$, $\nabla \log(Y_t)$, $\nabla_{12} \log(Y_t)$, and $\nabla\nabla_{12} \log(Y_t)$ suggest taking $\nabla\nabla_{12} \log(Y_t)$ for all seven series. Examining sample autocorrelations of the residuals from a regression of $\nabla\nabla_{12} \log(Y_t)$ on differenced trading-day and (when required) Easter holiday variables then suggests the airline model for most of our series, and for none of the series does the airline model appear to be an unreasonable choice. Also, BH suggest selecting a model for $\log(Y_t)$ via such usual techniques ignoring sampling error and using this as a starting point in modeling $\log(S_t)$, modifying the model if diagnostic checking suggests inadequacies in the model. We note later that diagnostic checking does not suggest inadequacies with the model in most cases. Notice that if θ_1 or θ_{12} is estimated to be approximately one, indicating overdifferencing, we can cancel the ∇ or ∇_{12} and add a trend constant or fixed seasonal regression terms to the model.

When the sampling error, $\log(u_t)$, is included in the model for $\log(Y_t)$, it is assumed to follow the model (3.4) within samples, with parameter values as given in table 2. The breaks in covariance structure of $\log(u_t)$ when a new sample is introduced are handled in estimation by the Kalman filter as discussed in Bell and Hillmer (1990b). We carried out the estimation using software recently developed by the time series staff of the Statistical Research Division, U.S. Bureau of the Census, for Gaussian maximum likelihood estimation of ARIMA component time series models with regression terms. The sampling error models are held fixed in the time series estimation; i.e., the likelihood is maximized over only $\theta_1, \theta_{12}, \sigma_b^2$, and the regression parameters. Table 4a gives estimation results for current dollar data with and without the sampling error component included in the model.

Focusing first on estimates of σ_b (a one-step-ahead prediction coefficient of variation) we see, as must occur, reductions in $\hat{\sigma}_b$ when some of the variation in the series is attributed to sampling error. There are substantial reductions in $\hat{\sigma}_b$ for eating places, hardware stores, and drinking places, so for these series economic interpretations of $\hat{\sigma}_b$ would be importantly affected by the presence of sampling error. The reductions in $\hat{\sigma}_b$ are smaller for the other series. The amount of reduction in $\hat{\sigma}_b$ depends not so much on the absolute amount of sampling error present in the series, which is measured by the composite estimate CV's in table 2, but on the magnitude of the sampling error relative to the signal, as inversely measured by the signal-to-noise ratios $\hat{\sigma}_b^2/\hat{\sigma}_c^2$ given in table 4a. The distinction between absolute and relative magnitudes of the sampling error is important for our series. For example, eating places has the second lowest composite estimate CV in table 2, but also has the second lowest signal-to-noise ratio in table 4a. This occurs because several other series with higher composite estimate CV's also appear to have signals that are inherently much less predictable, as reflected in their higher estimates of $\hat{\sigma}_b$. Thus, the impact of sampling error on $\hat{\sigma}_b$ depends not just on the magnitude of the sampling error, but on its magnitude in relation to the variability in the signal series.

Turning to the estimates of the seasonal moving average parameter θ_{12}, we see that including sampling error in the model brings large increases in $\hat{\theta}_{12}$ for eating places, hardware stores, and drinking places. Again, for these series the economic interpretation that could be placed on $\hat{\theta}_{12}$ would be sensitive to the treatment of sampling error. There is little change in $\hat{\theta}_{12}$ for the other series. For the most part, the behavior of $\hat{\theta}_{12}$ for the series here is consistent with the predictions of the previous section (see table 3b). The exceptions are radio and TV stores, and grocery stores. For radio and TV stores we would expect to see more change in $\hat{\theta}_{12}$ when the sampling error model is included, given its signal-to-noise ratio $\hat{\sigma}_b^2/\hat{\sigma}_c^2$, which is comparable

Table 4a

Estimated parameter values for models with and without sampling error (current dollar data).

	Model ignoring sampling error				Model with sampling error				
	$\hat{\theta}_1$	$\hat{\theta}_{12}$	$100\hat{\sigma}_b$	Q_{12}	$\hat{\theta}_1$	$\hat{\theta}_{12}$	$100\hat{\sigma}_b$	Q_{12}	$\hat{\sigma}_b^2/\hat{\sigma}_c^2$
Grocery stores	0.48	0.59	1.04	35.1	0.46	0.70	0.98	30.4	71.7
Eating places	0.25	0.73	1.52	9.5	0.19	0.94	1.29	10.4	8.5
Household appl. stores	0.45	0.52	4.10	11.1	0.49	0.49	3.98	10.5	39.2
Men's and boys' clothing stores	0.40	0.21	3.32	21.0	0.39	0.21	3.11	23.9	23.2
Hardware stores	0.18	0.68	3.56	16.4	0.19	0.99	2.91	23.8	15.5
Radio and TV stores	0.01	0.65	4.10	22.9	0.02	0.70	3.81	23.6	16.6
Drinking places	0.29	0.56	2.59	15.8	0.23	0.88	2.04	10.4	4.5

to that for hardware stores. For grocery store sales the increase in $\hat{\theta}_{12}$ is larger than expected, given the very small amount of sampling error evident from tables 2 and 3. We have no explanation for these exceptions. In general, the results for $\hat{\theta}_{12}$ here would seem to have important implications for seasonal adjustment, since the sampling error is capable of making what is essentially fixed seasonality ($\hat{\theta}_{12}$ near one) appear stochastic.

The estimates of θ_1 in table 4a change very little when sampling error is included in the model, suggesting that economic interpretations of this parameter should be relatively robust. This is not surprising since none of the $\hat{\theta}_1$'s approach the value of 0.9 for which important effects of sampling error were noted in section 4. (We hope eventually to investigate the possibility that effects might be seen with improved sampling error models developed from micro-data for the three-panel survey.) In terms of the simple theory outlined in the introduction and appendix A.1 the $\hat{\theta}_1$'s do not make much sense; even when sampling error is included in the model taking $1-\hat{\theta}_1$ as an estimated depreciation rate we find implied depreciation rates of at least 50 percent per month for all the store categories considered. Moreover, viewed from this perspective the ordering of the $1-\hat{\theta}_1$ values is not appealing. For example, the point estimates suggest that items sold at men's and boys' clothing stores, hardware stores, and radio and TV stores depreciate more rapidly than grocery store goods.

Finally, table 4a includes Ljung–Box (1978) statistics (Q_{12}) using twelve lags, as a rough check of aggregate model adequacy. These cannot sort out whether the signal and sampling error models are separately adequate, only whether the resulting model for $\log(Y_t)$ may be inadequate for explaining the covariance structure of the observed series. The Q_{12} statistics were computed using standardized residuals produced by the Kalman filter using the maxi-

Table 4b

Estimated parameter values for models with and without sampling error (constant dollar data).

	Model igoring sampling error				Model with sampling error				
	$\hat{\theta}_1$	$\hat{\theta}_{12}$	$100\hat{\sigma}_b$	Q_{12}	$\hat{\theta}_1$	$\hat{\theta}_{12}$	$100\hat{\sigma}_b$	Q_{12}	$\hat{\sigma}_b^2/\hat{\sigma}_c^2$
Grocery stores	0.40	0.76	1.12	35.0	0.35	0.90	1.02	30.3	77.8
Eating places	0.24	0.71	1.52	5.6	0.19	0.98	1.24	6.6	7.9
Household appl. stores	0.40	0.49	4.15	11.6	0.44	0.45	4.01	10.8	39.9
Men's and boys' clothing stores	0.34	0.24	3.37	18.9	0.32	0.25	3.17	21.7	24.2
Hardware stores	0.20	0.73	3.57	9.0	0.19	0.98	2.91	11.8	15.5
Radio and TV stores	0.00	0.64	4.14	19.9	0.02	0.68	3.83	20.6	16.8
Drinking places	0.28	0.54	2.58	13.3	0.25	0.86	2.06	10.3	4.6

mum likelihood estimates of the model parameters. Given the nature of our model, it is not clear that the usual asymptotic theory would apply to these statistics. Nevertheless, we provide them as a rough indication of model fit. If one follows the usual practice of comparing these against chi-squared critical values for 10 degrees of freedom (which are 18.3 for five percent and 23.2 for one percent) there would be mild concerns about some of the model fits, which could probably be alleviated by including additional nonseasonal moving average or autoregressive terms in the model at low lags. The largest Q_{12} values are exhibited by grocery stores, though this is of the least concern for our purposes here since grocery stores has the least sampling error of any of the series. Including sampling error in the model lowers Q_{12} in some cases, raises it in others.

Because consumption theory applies more naturally to real quantities than to current dollar values, we also fit our models to constant dollar series. The price deflators we used were components of the consumer price index (CPI) that we selected in order to obtain the best match between sales series and price series; for all of the sales categories except hardware stores (for which we used a price series for housing maintenance and repair commodities) it was not difficult to find a price series that appeared to provide a good match. These price indices are available in not-seasonally-adjusted form. (Details of our selections are available on request.) If p_t denotes the price index, then notice that we have

$$\log(Y_t/p_t) = \log(S_t/p_t) + \log(u_t), \tag{5.1}$$

so that by deflating the composite estimates Y_t we can model the deflated signal series, S_t/p_t, subject to the same sampling error u_t. Actually, the CPI components are themselves estimates from a sample survey, and thus our price deflators are also subject to sampling error. Since we are ignoring this, we have really omitted another (independent, additive, logarithmic) sampling error component in (5.1). We have not investigated whether information is available on the autocovariance structure of these sampling errors.

In any case, table 4b gives estimation results for the deflated data that convey the same message as those of table 4a, so that it matters little whether we use current or constant dollar data. Indeed, corresponding entries of tables 4a and 4b (other than some Q_{12} values) are quite close for every store category except grocery stores, where the estimate of θ_{12} is a bit higher in the constant dollar data both when the sampling error model is suppressed and when it is included.

6. Summary and conclusions

In this paper we have examined the sensitivity of modeling results for time series from the Census Bureau's Monthly Retail Trade Survey (RTS) to the presence of sampling error. Our interest in these series stems from their relevance to the consumption literature, which in recent years has emphasized the economic interpretation of parameters in time series models for consumption spending. Using available information on sampling error autocorrelations and (relative) variances, we constructed time series models for sampling errors in the RTS. Then, using the 'airline model' for the true, unobserved series as a baseline specification, we estimated models for seven time series from the RTS with and without models for the sampling error components. While there are some identifiable shortcomings in our sampling error models, the results of this effort, and those of the analytical exercise of section 4, suggest that certain modeling results can be sensitive to the presence of even moderate amounts of the type of sampling error present in the RTS. Explicit treatment of sampling error in the model can result in considerably lower estimates of the innovation variance of the true (signal) series and considerably higher estimates of the seasonal moving average parameter, relative to results obtained when sampling error is ignored. In regard to the latter, ignoring (seasonally correlated) sampling error can make seasonality appear much more variable than it appears when the sampling error is accommodated in the model. Estimates of the nonseasonal moving average parameter were not much affected by including sampling error in the model for the series considered here, though the parameter estimates turned out to be in a range where little effect would be expected. Given that estimates of the nonseasonal moving average parameter typically lie in the range 0 to 0.5, economic interpretations of this parameter would appear to be robust to the presence of sampling error from the monthly RTS.

The message from this analysis is that sampling error should be taken seriously in attempts to derive economic implications from modeling of time series data from repeated surveys. The presence of a 'small amount' of sampling error should not affect the analysis much; unfortunately, what constitutes a 'small amount' depends on the magnitude of the sampling error relative to the variability of the true (signal) series, i.e., on the signal-to-noise ratio. Thus, effort beyond looking up a sampling coefficient of variation in a publication is required to determine if sampling error may matter in a particular situation. Further work is needed to obtain better information on the autocovariance structure of sampling errors, and to bring them into an integrated inferential framework for time series analysis.

Appendix A.1: Deriving the airline model from an economic model for durable goods expenditures

This appendix gives one set of assumptions consistent with the level of spending on a durable good following an airline model.

The representative consumer chooses a path for expenditure on the durable good by solving the following maximization problem:

$$\max -\tfrac{1}{2}\,\mathrm{E}_t\left\{\sum_{j=0}^{\infty}\beta^j\left[\left(K_{t+j}-T_{t+j}\right)^2+l_{t+j}^2\right]\right\}, \tag{A1.1}$$

where β is the subjective discount factor, K_{t+j} is the accumulated stock of the durable good at time $t+j$, and T_{t+j} is an exogenous stochastic process describing the evolution of tastes. We discuss the role of l_t below. Maximization takes place subject to the following constraints:

$$K_t = (1-\delta)K_{t-1}+S_t, \tag{A1.2}$$

$$W_t = (1+r)W_{t-1}+Y_t-S_t, \tag{A1.3}$$

$$(1-B)(1-B^{12})T_t = \nu_t, \tag{A1.4}$$

$$l_t = \phi W_t, \tag{A1.5}$$

where δ is the depreciation rate; S_t measures spending on the durable good in time period t; r is the real interest rate; W_t is the level of nonhuman wealth held at the end of period t in some asset other than the durable good;

and Y_t is labor income. There are two sources of uncertainty: shocks to the taste-shift process (variance σ_v^2) and shocks to labor income (variance σ_y^2). We assume that the innovations in tastes are uncorrelated at all leads and lags with the innovations in labor income.

The transversality conditions are given by

$$\lim_{j\to\infty} \beta^j \mathrm{E}_t\left(\lambda_{t+j}^W W_{t+j}\right) = 0, \tag{A1.6}$$

$$\lim_{j\to\infty} \beta^j \mathrm{E}_t\left(\lambda_{t+j}^K K_{t+j}\right) = 0, \tag{A1.7}$$

where λ_{t+j}^W and λ_{t+j}^K are the Lagrange multipliers on W_{t+j} and K_{t+j}, respectively. We assume that the real interest rate is fixed and equal to the rate of time preference.

Many previous authors who have studied optimization problems similar to (A1.1)–(A1.7) have not included the variable l_t in the utility function. Implicitly, such previous authors have set ϕ equal to zero. Hansen and Sargent (1991) interpret l_t as measuring household inputs into production; in the specific case considered here 'production' takes place via a storage technology, and l_t can be interpreted as an adjustment cost. Hansen and Sargent (1991) note that with ϕ equal to zero, the optimal plan for the consumer would involve setting consumption at the bliss level period-by-period, and letting debt accumulate in an unrestricted fashion. Hansen and Sargent show that the transversality conditions would be satisfied under this solution despite the rapid accumulation of debt because the Lagrange multipliers both equal zero when consumption is at the bliss level. They also show, however, that when ϕ is nonzero, the optimal solution involves neither the setting of consumption equal to the bliss level nor the pathological accumulation of debt. Moreover, with ϕ set arbitrarily close to zero, the optimal solution approximates arbitrarily well the (technically incorrect) solution that most previous authors have proposed to problems similar to (A1.1)–(A1.7).

Tedious algebra yields that, for ϕ chosen very small, spending on the durable good S_t follows a univariate process arbitrarily close to

$$(1-B)(1-B^{12})S_t = \left[(1-(1-\delta)B\right]\left(1-\theta_{12}B^{12}\right)\xi_t. \tag{A1.8}$$

Model (A1.8) is the analogue in levels to the model we estimate in logs. We have verified model (A1.8) using the computer programs described in Hansen and Sargent (1991).

The seasonal MA parameter θ_{12} is a function of the real interest rate and the variance ratio σ_v^2/σ_y^2. As this ratio goes to zero, θ_{12} goes to one, and the model for S_t simplifies to an IMA(1, 1) with deterministic seasonal means.

Appendix A.2: The retail trade survey (RTS)

The Census Bureau's Retail Trade Survey (RTS) produces monthly estimates of sales for detailed kinds of retail businesses (defined by SIC codes) at the U.S. and regional level, and for less detailed kinds of retail businesses for some states and metropolitan areas. In this paper we shall deal only with data at the U.S. level. We note that the estimates for states and metropolitan areas typically are subject to much higher levels of sampling error than are the corresponding estimates at the U.S. level, and hence empirical work based on subnational sales estimates would be even more prone to the effects of sampling error investigated in this paper.

In the RTS, estimates are obtained from reports of sales from a monthly sample survey of businesses, and from benchmark adjustments to reports of annual sales from an annual sample survey and from the quinquennial economic censuses. The published estimates differ from the exact, actual values of retail sales because of sampling error and various nonsampling errors. In this appendix we briefly review some basic features of the survey design and estimation, some of which are used in section 3 in developing time series models for the sampling errors. Several other aspects discussed have to do with nonsampling errors, the magnitudes of which are largely unknown and which, therefore, cannot be accounted for in the modeling. Still, it is useful to be aware of these limitations in the data. The discussion here is necessarily brief; for more detailed discussion of the RTS see Wolter et al. (1976), Wolter (1979), and Garrett, Detlefsen, and Veum (1987). Woodruff (1963) is another useful reference, though important changes in the RTS have been made since the time of his writing. Useful summaries of the operation of the RTS, and notices of major changes, are provided in appendices to the data publications, the Census Bureau's Monthly Retail Trade Reports. Waite (1974) investigated nonsampling errors in the RTS.

Since 1971 the RTS has primarily relied on a list sample drawn from the Standard Statistical Establishment List (SSEL), with information relating to firm births and deaths obtained from the Social Security Administration and the Internal Revenue Service. A separate geographic area sample is used to cover businesses not within the list frame, mainly new businesses not yet entered into the SSEL and businesses without payroll (e.g., some family businesses). In recent Retail Trade Reports the area sample is reported as contributing only about six percent to overall retail sales, though this amount varies by kind of business. Because of its generally small contribution, we shall not consider the area sample in detail here.

The RTS consists of a panel of larger businesses selected into the sample with certainty, rotating panels of list sample businesses selected by stratified simple random sampling without replacement, and the rotating panel area sample cases selected in a multistage procedure. New samples are independently redrawn and introduced about every five years; since the move to the list sample, this has occurred in September 1977, January 1982, and January 1987. When a new sample is instituted, certainty status is assigned to those businesses whose sales in the most recent economic census exceeded cutoff points specific by kind of business. Certainty cases report monthly on sales for the current reporting month. Presently three list sample rotating panels report current and previous month sales every three months (three-panel design); prior to September 1977 an analogous four-panel design was used. This basic design has important implications for the time series properties of the sampling errors. Since the rotating panels are drawn independently (or approximately so, to the extent that the sampling fraction of the noncertainty cases is small), the sampling errors in estimates arising from different panels should be independent. Also, since the redrawing of the samples about every five years is done independently, sampling errors from these different samples will be independent.

Each month, unbiased Horvitz–Thompson (HT) estimates [Cochran (1977, pp. 259–261)] of current and previous month sales are constructed. These estimates are weighted totals of the sample observations, where the weights are the inverses of the probabilities of selection – one for the certainty cases. The HT estimates are then used in producing 'composite' estimates, as discussed in section 3. Sampling variances are estimated using the random group method [Wolter (1985)] for the list sample (with 16 random groups), and the collapsed stratum method for the area sample. In principle, by linking random group totals for pairs of months, covariances of the corresponding sampling errors can be estimated by the random group method in the same way that sampling variances are estimated. This is how the lag covariances were estimated in the special study mentioned in section 3 that produced the sampling error autocorrelations used in this paper. Unfortunately, random group totals are not saved when the monthly survey estimation processing is done, so that current estimates of sampling error autocovariances are not readily available.

Large observation procedures [see Woodruff (1963) and Wolter et al. (1976)] are used to reduce the variances of the HT estimates while still retaining unbiasedness. Basically, reported sales of noncertainty sample units are compared against cutoff values and, if the cutoff values are exceeded, the units are designated as monthly noncertainty cases, either temporary (cutoff exceeded in one month) or permanent (cutoffs exceeded for a sequence of six months). These units are then canvassed and tabulated one additional month (temporary case) or each month thereafter (permanent case) but with sample weights reduced to maintain unbiasedness. The important consequence of these procedures is that some of the businesses selected in one rotating panel will, over time, start contributing to the estimates of the other rotating panels, thus inducing correlation in the sampling errors of the estimates from different panels. When a new sample is implemented there are no monthly noncertainty cases. Ruth Detlefsen of the Business Division at Census reports that the monthly noncertainty cases typically build up fairly quickly and stabilize at some level that is significant for many kinds of business.

Calendarization and imputation procedures are needed to address two important response problems. The former deals with the fact that some businesses do not keep their books on a calendar month basis. In such cases the preferred procedure is for the business to provide an estimate of calendar month sales. However, some such businesses file RTS reports for periods other than calendar months (e.g., four- or five-week periods); the Census Bureau then applies calendarization procedures to convert these data to a calendar month basis. Such procedures obviously are necessary to produce calendar month sales figures; however, either possibility leads to nonsampling error of an essentially unknown magnitude in the published estimates. Imputation procedures are needed to deal with missing data arising from nonresponse, late response, and edit failures (detected bad data). Imputations are based on past values for the same business and other businesses of the same kind in the same panel. While this procedure will not induce correlation across panels, it can inflate correlation within a panel. The extent of this effect is difficult to know, though it certainly depends on the level of nonresponse in the RTS, which varies over time and by kind of business. Garrett, Detlefsen, and Veum (1987) report that as of August 1987 roughly 25 percent of total retail sales by value was being imputed, though since 1989 the imputation rate has dropped to about 17 percent. Imputation rates are higher at the start of a new sample due to difficulties in getting some new sample units to respond initially, and difficulties in determining that some units selected for the new sample are no longer in business. In any case, there has been a significant deterioration from previous years; Wolter et al. (1976) report a nonresponse rate of only nine percent.

As mentioned earlier, monthly survey estimates are benchmarked to annual totals from the annual RTS and the quinquennial economic censuses. The data from these sources are believed to be more reliable than those from the monthly RTS because the response rates are higher (possibly due to mandatory reporting requirements), and because businesses generally have book figures (rather than just estimates) available on an annual basis (for calendar or fiscal years). Benchmarking can remove significant amounts of nonsampling error, at least in regard to annual totals, but month-to-month movements are still determined by the monthly survey.[1] As benchmarking involves a filtering of the time series of estimates, it will affect their autocorrelation properties. To avoid such problems in this paper, we use data that are not benchmarked. For this reason, the data used here do not agree with published estimates.

Another adjustment of importance to the monthly estimates is made to avoid sudden shifts in level whenever a new sample is introduced. Data in the old sample are multiplied by the geometric mean of the ratios of new sample to old sample estimates obtained for two months for which the old and new samples overlap. Since the models we shall use involve taking logarithms and (regular and seasonal) differencing, the effect of this adjustment will be limited to the first month of the new sample and the same month a year later.

References

Bell, W.R., 1983, A computer program for detecting outliers in time series, Proceedings of the American Statistical Association, Business and Economic Statistics Section, 634–639.

Bell, W.R., 1987, A note on overdifferencing and the equivalence of seasonal time series models with monthly means and models with (0,1,1) 12 seasonal parts when $\Theta = 1$, Journal of Business and Economic Statistics, 5, 383–387.

Bell, W.R. and S.C. Hillmer, 1983, Modeling time series with calendar variation, Journal of the American Statistical Association 78, 526–534.

Bell, W.R. and S.C. Hillmer, 1990a, The time series approach to estimation for repeated surveys, Survey Methodology 16, 195–215.

Bell, W.R. and S.C. Hillmer, 1990b, Applying time series models in survey estimation, Unpublished manuscript (Bureau of the Census, Washington, DC).

Binder, D.A. and J.P. Dick, 1989, Modelling and estimation for repeated surveys, Survey Methodology 15, 29–45.

Box, G.E.P. and G.M. Jenkins, 1976, Time series analysis: Forecasting and control (Holden-Day, San Francisco, CA).

Chang, I., G.C. Tiao, and C. Chen, 1988, Estimation of time series parameters in the presence of outliers, Technometrics 30, 193–204.

Cochran, William G., 1977, Sampling techniques (Wiley, New York, NY).

Eltinge, J.L. and W.A. Fuller, 1989, Time series random component models for sample surveys, Paper presented at the winter meeting of the American Statistical Association, San Diego, CA.

Garrett, J.K., R.E. Detlefsen, and C.S. Veum, 1987, Recent sample revisions and related enhancements for business surveys of the U.S. Bureau of the Census, American Statistical Association, Proceedings of the Business and Economic Statistics Section, 141–149.

Hansen, L.P. and T.J. Sargent, 1991, Recursive linear models of dynamic economies, Unpublished manuscript (University of Chicago, Chicago, IL and Hoover Institution, Stanford, CA).

Hausman, J.A. and M.W. Watson, 1985, Errors in variables and seasonal adjustment procedures, Journal of the American Statistical Association 80, 531–540.

Ljung, G.M. and G.E.P. Box, 1978, On a measure of lack of fit in time series models, Biometrika 65, 297–304.

Mankiw, N. Gregory, 1982, Hall's consumption hypothesis and durable goods, Journal of Monetary Economics 10, 417–426.

[1]Monsour and Trager (1979) describe benchmarking in the RTS, and Trabelsi and Hillmer (1990) relate the benchmarking procedure to a general statistical approach using time series models.

Monsour, N.J. and M.L. Trager, 1979, Revision and benchmarking of business time series, American Statistical Association, Proceedings of the Business and Economic Statistics Section, 333–337.

Pfeffermann, D., 1991, Estimation and seasonal adjustment of population means using data from repeated surveys, Journal of Business and Economic Statistics 9, 163–175.

Scott, A.J. and T.M.F. Smith, 1974, Analysis of repeated surveys using times series methods, Journal of the American Statistical Association 69, 674–678.

Scott, A.J., T.M.F. Smith, and R.G. Jones, 1977, The application of time series methods to the analysis of repeated surveys, International Statistical Review 45, 13–28.

Tiller, Richard, 1990, Am application of time series methods to labor force estimation using CPS data, Proceedings of the American Statistical Association, Section on Survey Research Methods, 284–289.

Trabelsi, A. and S.C. Hillmer, 1990, Benchmarking time series with reliable benchmarks, Applied Statistics 39, 367–379.

Train, G., L. Cahoon, and P. Makens, 1978, The Current Population Survey variances, inter-relationships, and design effects, American Statistical Association, Proceedings of the Survey Methods Research Section, 443–448.

Waite, P.J., 1974, An evaluation of nonsampling errors in the monthly retail trade sales data, American Statistical Association, Proceedings of the Business and Economic Statistics Section, 602–607.

Wilcox, David W., 1991, What do we know about consumption?, Unpublished paper (Federal Reserve Board, Washington, DC).

Wolter, K.M., 1979, Composite estimation in finite populations, Journal of the American Statistical Association 74, 604–613.

Wolter, K.M., 1985, Introduction to variance estimation (Springer-Verlag, New York, NY).

Wolter, K.M., C.T. Isaki, T.R. Sturdevant, N.J. Monsour, and F.M. Mayes, 1976, Sample selection and estimation aspects of the Census Bureau's monthly business surveys, American Statistical Association, Proceedings of the Business and Economic Statistics Section, 99–109.

Woodruff, R.S., 1963, The use of rotating samples in the Census Bureau's monthly surveys, Journal of the American Statistical Association 58, 454–467.