

# AN ANALYSIS OF THE ADVANCE MONTHLY RETAIL SALES SURVEY

Julia L. Bienias, William Davie, Jr., Howard Hogan, Carl A. Konschnik, Bureau of the Census  
Carl A. Konschnik, Bureau of the Census, Washington, DC 20233 ckonschnik@ccmail.census.gov

**Key Words:** Monthly surveys, Nonresponse bias, Reporting bias, Rotating panels

## 1. Overview

Nine working days following the end of each month, the Bureau of the Census publishes its first estimate of retail sales for that month. This Advance estimate is widely reported in the press and examined closely by financial markets and policy analysts. At the same time, the Bureau publishes a Preliminary estimate of retail sales for the previous month and a Final estimate for the month before that. The Advance estimate will normally differ from both the Preliminary and the Final estimates for the same month. Keen interest by policy analysts and others in using the Advance estimate to reliably predict the change in month-to-month sales (to be later determined more accurately by the Preliminary and Final estimates) causes us to work to reduce these differences.

We looked at several possible reasons for the differences. These included added variance due to a smaller Advance sample, nonoverlapping units between the samples, the effects of early reporting, nonresponse, different estimators, different imputation methodology and different seasonal adjustment factors. In this paper, we present the results of our investigation of these reasons and their impact. In summary, the Advance estimate generally predicts the Preliminary and Final estimates quite well. Better early reporting and reducing nonresponse can improve the Advance estimates. Nonetheless, the primary reason for differences appears to be the variance of the smaller Advance sample. We also discuss steps that the Census Bureau is taking to reduce differences in the future. We begin by describing the two voluntary surveys that give rise to the Preliminary and Final monthly retail sales estimates and to the Advance monthly retail sales estimates.

## 2. Description of the Monthly Retail Trade Survey and Advance Monthly Retail Sales Survey

To describe the relationship between the Monthly Retail Trade Survey (MRTS) and the Advance Monthly Retail Sales Survey (Advance), we need to look at relationships between their samples, estimation methodology, data collection methods and periods, imputation practices, seasonal adjustment, and other factors.

## 2.1 Description of the Samples

### 2.1.1 The MRTS Sample

The MRTS sample is drawn every five years from the Census Bureau's Standard Statistical Establishment List (SSEL). The SSEL is a universe list of employer businesses built and periodically updated with the administrative records of the Federal Government (chiefly those of the Internal Revenue Service and the Social Security Administration), along with the results of current surveys and censuses.

For the MRTS, selected company and Employer Identification Number (EIN) certainty (weight=1) units are assigned to a fixed panel (panel 0) that is canvassed every month. In contrast, the noncertainty (weight > 1) selected EIN units are assigned to three rotating panels, panels 1, 2, and 3, and are asked to report data for two consecutive months, four times a year. We refer to these two reference months as the current month and the previous month. One other panel of selected EIN units, panel 5, represents noncertainty cases that, like panel 0 cases, are canvassed each month. Since August 1993, birth employers (new employer business EINs) and nonemployers are represented through a benchmarking factor approach. Prior to that time, they were represented by an area sample. See Konschnik (1994) for more details on this representation.

The number of selected sample units changes over time. The latest sample was selected in 1991, but by late 1995, the number of certainty company units dropped significantly from about 3,000 to about 2,200 units due to mergers, closings, and selling of stores, with their successors sometimes becoming certainty EINs. As a result of these successors and because of birth and death processing and variance reduction procedures, counts in late 1995 showed 1,603 certainty EINs in panel 0 and about 7,900 noncertainty EINs in each rotating panel. In addition, at this point in the sample life cycle, panel 5 had grown to contain 1,446 units.

### 2.1.2 The Advance Sample

The Advance sample is drawn from the selected MRTS sample, usually every two or three years. For its sampling frame, the Advance sample uses units that report in the MRTS. Units with irregular response patterns, or those that are out of business, out of scope, or

refusals, are excluded from the Advance frame. The frame used for the sample initiated in early 1993 and used until April 1996 contained about 9,000 units, from which a sample of about 3,400 units was selected. At the beginning of each month, selected units are asked to report sales for the month just ended without regard to what their MRTS panel had been. Over time, the Advance sample diminishes due to units going out of business, refusing to participate, etc., necessitating a sample reselection. Over the roughly three year span since the sample was selected, through attrition (births are not added and out-of-business cases, refusals, etc., drop out) the mailout sample was reduced to about 2,300 units. The results of this paper are based mostly on this 1993 sample that had an area sample component covering nonemployers and birth employers. A new Advance sample was drawn and initiated in May of 1996. Nonemployers and the most recent birth employer units are not represented in this new sample because the MRTS sample from which it was drawn no longer has a sample representation for them.

## **2.2 Estimation Methodology**

### **2.2.1 MRTS Estimation Methodology**

By way of example, we obtain two simple weighted estimates (unbiased estimates) for January sales, one from panels 0, 1 and 5 (where January is considered the current month), and another a month later from panels 0, 2 and 5 (where January is considered the previous month). These January estimates are linked to the prior month's estimate through a composite estimation procedure that results in a Preliminary composite estimate for January (after using results from the latest data collection for panels 0, 1 and 5), and a final composite estimate for January (after using results from the latest data collection for panels 0, 2 and 5). The Preliminary composite estimate for the current month is a weighted average of two estimates: (1) the current month unbiased estimate (weight 0.25) and (2) a ratio estimate (weight 0.75) obtained by multiplying the current-to-previous month ratios developed from the two months of data on the current month questionnaires by the Preliminary composite estimate for the previous month.

The Final composite estimate, produced one month later, is also a weighted average of two estimates: the Preliminary composite estimate for a given month (weight 0.8) and the unbiased estimate (weight 0.2) for the same month as obtained from the next month's reporting panel. So each month we obtain a Preliminary composite estimate for the current month and a Final composite estimate for the previous month. Their ratio yields the current month's Preliminary trend estimate.

### **2.2.2 The Advance Estimation Methodology**

The Advance survey provides estimates of month-to-month change based on those selected units that responded in both the current and previous months and passed edits. Other selected sample units are not included in the tabulations, and no explicit imputation for nonresponse is done. No attempt is made to get a monthly level-of-sales estimate from the Advance sample alone. The month-to-month change estimates are multiplied by the previous month's Preliminary composite estimates to yield Advance level estimates for the current month. Thus the Advance survey yields, in effect, only estimates of month-to-month change. The key question, then, is how well do the Advance estimates of month-to-month change predict the Preliminary estimates and Final estimates that come later.

## **2.3 Data Collection and Imputation Practices**

### **2.3.1 MRTS Data Collection and Imputation Practices**

The MRTS is a mail-out, mail-back survey. We begin telephone follow-up for nonresponse to the mail survey about the middle of the month following what we refer to as the current (or latest reference) month. The data collection, processing and analysis continue for about the next three weeks with Preliminary composite estimates for the current month and Final composite estimates for the previous month available around the end of the first week in the next month (in time to be used for the Advance survey estimation for the month following the current month).

In all, about 60% of the reporting units respond by mail and about 20% respond via telephone follow-up. About 20% of the remaining units do not respond; their data are imputed by using their most recent sales levels and the trend of all reporters. Using these additional imputed data, we obtain total sales levels for the month.

### **2.3.2 The Advance Survey Data Collection and Imputation Practices**

Five days before the end of the month, we mail an Advance survey questionnaire to sample units, but we usually receive only about 10 to 15 percent back through mail and FAX prior to beginning telephone follow-up. Thus we obtain most response by telephone. Calling for data runs from the 3rd through the 6th working day of the month, with final calls completed by the morning of the 7th working day. Publication is scheduled for the 9th working day. However, if the 9th working day is a Monday, the data are released on the 10th working day.

Of the approximately 2,300 Advance sample units contacted in recent months, only about 60%, or about 1,800 units, respond and have data in the current and previous months. These responding units are edited and those that pass edits are included in the Advance month-to-month change estimate. There is no imputation because the level estimate comes from using this Advance ratio applied to the previous month's Preliminary composite estimate.

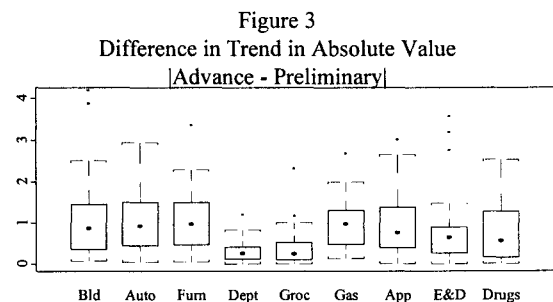
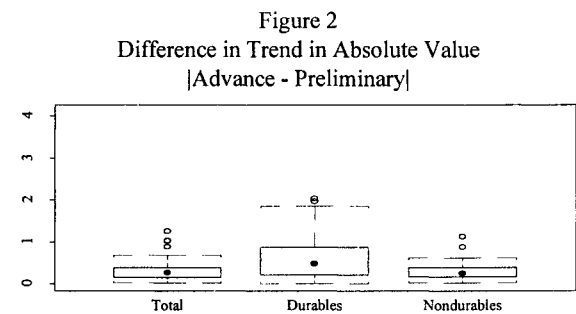
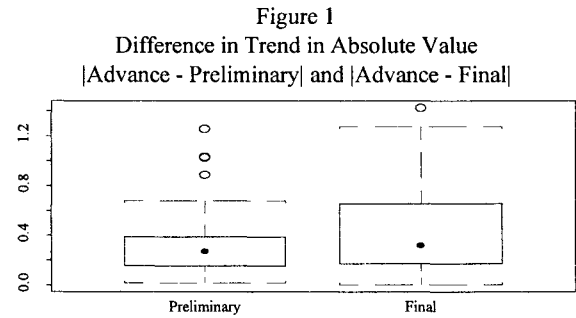
## 2.4 Adjustment Factors for Seasonal and Other Variations

We use essentially the same methods to adjust both the estimates from the MRTS and from the Advance for seasonal, trading day, and holiday patterns. Both use the X-11 ARIMA program and concurrent seasonal adjustment using all available unadjusted estimates. However, one difference between the adjustment factors for a given month's Advance estimate and its Preliminary estimate is that, for input into the seasonal adjustment program, the Advance uses the Advance estimate for the given month, the Preliminary estimate for the previous month and the Final estimate for the month before the previous month. In contrast, the Preliminary estimate for a given month uses seasonal factors derived from the Advance estimate for the month after the given month, the given month's Preliminary estimate, and the previous month's Final estimate. Further, for some SIC's, the Advance estimate's seasonal factors are computed at broader levels (in line with the Advance publication levels) than those used for the Preliminary. The Final estimate for a given month uses seasonal factors generated (at the same levels as those of the Preliminary) by using as input the Final estimate for the given month, the Preliminary estimate for the month after the given month, and the Advance estimate for the month two months after the given month.

### 3. Differences in Estimates of Trends for a Recent 38 Month Period

In Figures 1, 2 and 3, we show box plots that describe the trend differences between Advance estimates and MRTS estimates for the 38 month period November 1992 through December 1995. Seasonally adjusted data were used to prepare these plots. In Figure 1, the median absolute difference (in percentage points) between the Advance and Preliminary estimates of total retail sales trend is 0.27. Between the Advance and Final, the median absolute difference is 0.32. However, because of space limitations, in Figures 2 and 3 and in the rest of this paper we focus on the differences in trends between the Advance and Preliminary estimates. Figure

2 is at broad levels (durable and nondurable goods) and Figure 3 shows plots for the major kind of business groups within retail. From left to right, these plots are for Building materials (Bld), Automotive (Auto), Furniture (Furn), Department stores (Dept), Grocery (Groc), Gasoline (Gas), Apparel (App), Eating and Drinking (E&D) and Drug (Drugs) store sales. Differences are generally higher than those for total retail, except for department stores and grocery stores, where the differences are about the same as those for total retail.



### 4. The Major Reasons for Differences and Their Impact

The different estimators used by the Advance survey and the MRTS have little impact on the differences in estimates of trend--the Advance estimate relies on a ratio of cases that reported in the current and prior month ("identical reporters"), and the composite estimator

essentially preserves the estimate of trend based on a ratio of identical reporters in the MRTS. Nor do the different imputation methodologies have an impact, because the Advance does not impute for nonresponse, and the MRTS imputes nonrespondents so that they assume the same trend as the responding units.

We found several principal reasons for differences in estimates of trend between the Advance and MRTS. These are: (1) sample variance, and the degree of non-overlap of the samples, (2) early Advance reporting, (3) survey nonresponse, and (4) seasonal adjustment factors. We discuss each of these and their impact in the following sections.

#### 4.1 Sample Variance and the Degree of Non-overlap of the Advance and MRTS Samples

The Advance sample produces a median estimated coefficient of variation (cv) of about 0.8% for month-to-month trend at the total retail level. By contrast, the median estimated cv on the MRTS Preliminary estimate of month-to-month change at the total retail level is about 0.3%, indicating that much greater precision is achieved by the much larger sample.

Not only is the Advance sample that is tabulated in a given month much smaller than the MRTS sample, but it also consists of different cases because of the rotating panels in the MRTS. Thus, although the Advance sample is drawn from the MRTS sample, in any particular month it is not a proper subsample. For example, an Advance reporter for the January data month might be a panel 2 or 3 in MRTS, in which case it would not have both December and January data from the MRTS. Generally, only cases in panels 0, 1 and 5 in the MRTS would overlap the Advance reporters for January. Other reasons for sample differences exist. For example, the Advance sample is drawn from a MRTS sample at a specific point in time, and a selected unit in the Advance sample may not be in the MRTS sample at a later point in time. Area sample units are examples of these types of cases. Looking at the actual overlap shows that about 900-1,000 cases typically overlap between the two surveys in any given month. Table 1 includes the degree of overlap between the two surveys for the eight month period September 1995 through April 1996. Data are given at the total retail sales level only. Because only half or so of the cases in a given month's Advance estimate are also in the MRTS sample for that month, it has been difficult to do direct comparisons to determine any specific causes for differences. Nevertheless, for the cases that overlap both surveys in a given month, we looked at their contributions and discuss problems with the data quality of early reports in the next section.

**Table 1**  
Measures of Advance and MRTS Overlap and Effects of Data Revision on the Advance Estimates of Current Month to Previous Month Trend for Total Retail Sales  
 (Source Data are Unadjusted for Seasonality)

Mon/Yr	Pub Adv		# Adv Cases Also in MRTS	# Adv Cases w/ Revised MRTS Data	Adv Using Rev MRTS Trend	Pub Prelim Trend
	#	Trend				
Sep 1995	1849	- 5.2	933	554	- 5.2	- 5.4
Oct 1995	1828	0.8	917	545	1.2	0.4
Nov 1995	1818	4.1	915	547	5.9	4.0
Dec 1995	1984	16.9	997	449	18.8	17.4
Jan 1996	1925	-26.4	962	506	-28.5	-26.5
Feb 1996	1905	2.6	985	544	3.0	3.9
Mar 1996	1680	10.0	893	539	10.5	10.9
Apr 1996	1679	- 0.8	859	521	- 0.4	- 0.5

#### 4.2 Early Reporting Effects

In an effort to determine other possible causes for the differences in estimates of trend between the Advance and Preliminary, we looked at whether early reporters for the Advance survey would later in the month provide revised data for the MRTS. Further, we measured whether these revisions, had they been reported for the Advance tabulation, would have significantly improved the Advance estimate. Table 1 shows that of the cases in the Advance that are also in the MRTS for a given month, often more than half revise their data between the earlier Advance collection and the MRTS collection about 3 weeks or so later. However, comparing the Advance reported current month to previous month trend and the Revised Advance trend using MRTS data to the published Preliminary trend, we see that the revised data usually does not improve the Advance estimate much (and sometimes makes it worse) relative to the published Preliminary trend estimate. (Note that both Table 1 and subsequently Table 2 which follows use data unadjusted for seasonality.) For the September through April period, the improvement is either small or would take the Advance estimate farther from the Preliminary.

Table 2 expands this analysis to individual kinds of business. It shows absolute differences in trends between the Advance and Preliminary estimates (column (1), |A-P|) and between the revised (with later MRTS data) and Preliminary estimates (column (2), |R-P|) for total retail and for the major components of retail sales. In Table 2, where |A-P| > |R-P|, the revision in the data would have improved the Advance estimate. Further, where |R-P| is close to 0 while |A-P| is substantial, the revision essentially adjusted well for the difference between the Advance as originally reported and the Preliminary. Furniture for September is an example of this, but this doesn't adjust for the difference completely, nor should it be expected because any early reporting effect would not be expected to account for the entire difference.

Obviously, where  $|R-P| > |A-P|$ , or, equivalently, where column (2) > column (1), the revision would have made the Advance estimate worse as a prediction of the Preliminary.

From studying Tables 1 and 2 we conclude that although data reported early that are later revised do sometimes improve the Advance estimates for total retail or for some major components, this early reporting effect is not the dominant factor between the differences in the Advance and MRTS trend estimates.

Further discussion of the quality of early retail sales data used for the Advance survey and subsequent data revision in the MRTS is given in the paper by Davie (1996). Obviously, knowing exactly when reliable data can be provided by sample units is a key element of producing reliable early estimates of retail sales. We hope to take advantage of this in our upcoming sample designs.

tabulated in the MRTS. In essence, then, S includes data for the Advance nonresponse cases.

The degree to which  $|S-P|$  is smaller than  $|R-P|$  indicates the gain in estimating P, separate from accounting for the early reporting effect, if the nonresponding units in the Advance survey had reported in time for the Advance estimate tabulation. Comparing columns (2) and (3) in Table 2 leads us to conclude that in some cases accounting for the nonresponse helps to improve the estimate relative to P. However, in many cases S is farther from P than R. In this latter instance, if we had tabulated data for the entire Advance sample for which MRTS data were available for that month, we would not have improved the estimate over and above what we would have obtained from using the Advance sample (only units tabulated for that month's Advance estimate) as revised with MRTS data.

For example, looking at two specific kinds of business and excerpting from Table 2, Figures 4 and 5 show graphically that early Advance reporting and Advance nonresponse have more of an effect on the grocery estimates than on the estimates from auto sales establishments. From this we conclude that the sample size for grocery stores is probably adequate but we should increase the sample for autos.

**Table 2**  
Absolute Difference in Trend for Advance vs. Preliminary Revised Advance vs. Preliminary, and Revised Plus Advance Nonresponse vs. Preliminary for the Eight Month Period September 1995 through April 1996 for Selected Kinds of Business  
 (A = Advance Trend, R = Revised Advance Trend, S = Revised Plus Advance Nonresponse Trend, P = Preliminary Trend)  
 (1) =  $|A-P|$ , (2) =  $|R-P|$ , (3) =  $|S-P|$   
 (Source Data Are Unadjusted for Seasonality)

Kind of Business	September 1995	October 1995	November 1995	December 1995
	(1),(2),(3)	(1),(2),(3)	(1),(2),(3)	(1),(2),(3)
U.S. Total	0.2,0.2,0.2	0.4,0.8,0.7	0.1,1.9,0.4	0.5,1.4,1.4
Build Mat	0.6,1.0,0.8	2.0,1.6,1.2	0.8,1.0,0.6	1.7,2.2,1.9
Auto	1.5,0.8,0.2	1.4,1.9,1.8	0.8,0.6,2.5	1.3,1.3,0.8
Furniture	0.7,0.2,0.4	0.7,1.0,0.4	0.8,0.3,2.1	0.8,1.4,0.2
Dept. Strs	0.3,0.1,0.1	1.2,1.0,0.8	0.3,0.3,1.2	0.2,1.1,0.2
Grocery	0.5,0.7,0.5	0.6,0.5,0.5	0.4,0.3,0.0	0.2,0.5,0.1
Gasoline	1.8,0.8,0.7	0.6,1.0,1.0	1.7,0.9,0.7	0.8,0.7,0.5
Apparel	1.5,1.6,0.3	0.1,1.1,0.6	0.0,0.6,0.7	0.0,1.1,1.5
Eat & Dr	0.7,0.5,0.2	0.6,0.3,0.0	1.7,2.0,1.0	0.3,0.3,0.8
Drugs	1.2,1.5,1.0	0.7,3.0,2.3	0.3,1.2,0.5	0.4,2.5,0.1

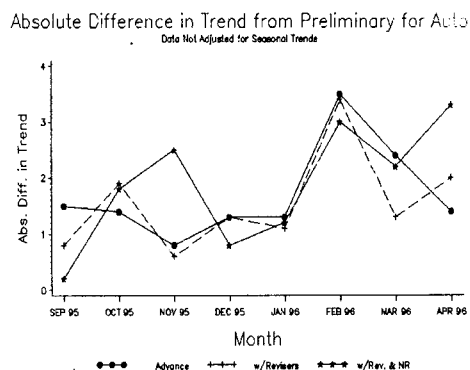
  

Kind of Business	January 1996	February 1996	March 1996	April 1996
	(1),(2),(3)	(1),(2),(3)	(1),(2),(3)	(1),(2),(3)
U.S. Total	0.1,2.0,1.8	1.3,0.9,1.1	0.9,0.4,0.7	0.3,0.1,0.4
Build Mat	2.0,2.4,0.6	0.8,0.4,0.7	2.3,2.8,2.2	5.2,6.5,4.6
Auto	1.3,1.1,1.2	3.5,3.4,3.0	2.4,1.3,2.2	1.4,2.0,3.3
Furniture	1.4,0.7,1.3	0.3,0.1,1.1	0.8,0.5,0.5	1.5,1.2,2.0
Dept Str	0.0,0.1,0.8	0.1,0.0,0.9	1.8,0.7,0.4	1.2,0.9,0.7
Grocery	0.2,0.1,0.1	0.1,0.3,0.0	0.2,0.1,0.1	0.1,0.1,0.1
Gasoline	2.2,0.4,0.1	0.2,0.3,0.4	2.4,0.8,0.2	0.3,0.4,0.3
Apparel	0.1,0.3,0.5	4.2,2.2,1.9	1.9,0.2,2.8	1.8,2.1,1.5
Eat & Dr	1.2,1.6,1.1	0.9,0.5,0.3	1.5,1.4,0.7	0.9,1.4,0.3
Drugs	0.2,0.8,1.6	0.1,0.4,0.8	0.7,0.5,0.3	0.0,1.0,0.9

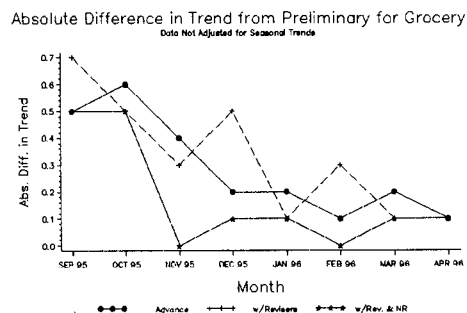
### 4.3 Advance Survey Nonresponse

In Table 2, in column (3) for each month we show the absolute difference between the Preliminary (P) and the Advance sample (S) trends, where S uses data from the Advance sample as of the MRTS tabulation. Thus, S includes the cases in the Advance estimate that are revised for the MRTS, as well as cases that are not in the Advance estimate but are in the Advance sample and are

**Figure 4**



**Figure 5**



#### 4.4 Concurrent Seasonal Adjustment

We have just begun to look more closely at the impact of the concurrent seasonal adjustment process on the difference between the Advance trend and Preliminary trend, although analysts have felt that it generally helps to minimize the difference. Comparisons were done for U.S. total retail sales and for some major components. For the U.S. total retail sales for the February 1996 data month, the published difference between the adjusted Advance and Preliminary trend estimates was 1.1 percentage points. (February was not a good month as this is much larger than the median difference of about 0.3 percentage points.) Had the same seasonal adjustment factors been used for the Preliminary trend estimate as for the Advance estimate, the differences would have been 1.3 percentage points. Similar results held for the kind-of-business components for the January to February trend.

These results also held for the other months we have examined--March, April, and May, 1996. The concurrent seasonal adjustment improved the trend differences by 0.5, 0.2, and 0.3 percentage points, respectively, at the U.S. total retail sales level.

#### 5. Conclusion and Future Research Plans

Were the difference in trend between the Advance and Preliminary or Advance and Final to remain each month around 0.3 percentage points or less for total retail sales, there would be a much greater level of data user satisfaction. However, the months like February 1996, where the difference is 1.1 percentage points, are disturbing to data users and to statistics producers as well. Thus, we are seeking ways to avoid such large differences.

From section 4, we conclude that the chief source of difference between the Advance and Preliminary trends for a given month appears to be the variance of the smaller (and somewhat different) Advance sample, because we have investigated response bias and nonresponse bias and found them to be generally small contributors to the difference.

In May of 1996 we began a new Advance sample that has an increased sample size in the key autos component. Early results suggest that this should help to improve the Advance estimates from this point on. This Advance sample will be in place when we introduce a new fixed panel design for the MRTS beginning in early 1997. Caldwell and Cantwell (1996) discuss this fixed panel design for the MRTS relative to the currently-used rotating panel design. A new Advance sample will be drawn from the fixed panel MRTS sample and introduced probably late in 1997 or early 1998. At that point we will

have the Advance sample as a proper subset of the MRTS sample. Because every Advance sample case will also have MRTS data, we will be able to study in detail, and without confounding, the reporting patterns, data differences, response and nonresponse bias of the Advance data. This should allow us to test and select the most appropriate model for improving the Advance estimates as predictors for the MRTS estimates. At that time we expect to have the best opportunity to improve the quality and reliability of the important Advance estimates of retail sales.

#### References

- Konschnik, C. (1994), "The Impact of Dropping the Business Area Sample as a Means of Covering Birth Employers and Nonemployers in the Current Business Surveys", Proceedings of the Section on Survey Research Methods, American Statistical Association, Vol. II, 995-1000.
- Caldwell, C., and Cantwell, P., (1996), "Comparing a Fixed-Panel Sample Design with a Rotating-Panel Sample Design for Monthly Trade Surveys", Proceedings of the Annual Research Conference, Bureau of the Census, (to appear).
- Davie, W.C., Jr., (1996), "Measuring the Effect of Early Reporting on Response Errors in an Economic Survey", Proceedings of the Section on Survey Research Methods, American Statistical Association, (to appear).

This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

We thank Linda Moore and Michael Shimberg for their assistance with programming, James Burton for his assistance with Figures 1-3 and Kathy Trickey for typing this paper.