BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: Census/SRD/RR-87/18

A COMPARISON OF ALTERNATIVE ESTIMATORS FOR REVISED MONTHLY IMPORT STATISTICS

by

Edward Gbur
Statistical Research Division
Bureau of the Census

Recommended by: Nash J. Monsour

Report Completed: August 1987

Report Issued: August 1987

A Comparison of Alternative Estimators for Revised Monthly Import Statistics

Edward Gbur*
Statistical Research Division
Bureau of the Census
Washington, D.C. 20233

Abstract

The import carryover problem and the estimator of revised monthly import level are described. Two alternative estimators are developed. The performance of these estimators and the estimators of month to month change derived from them are investigated.

1. Introduction

Each month the Census Bureau issues a report containing the U.S. foreign trade statistics. The statistics in these reports are used in a broad range of settings from determining the nation's trade balance to monitoring import quotas to providing detailed information of interest to individual American industries. In these reports the Census Bureau's goal is to produce accurate statistics on imports and exports at both the aggregate and detailed level on a timely basis. Data for monthly tabulations are received on a flow basis from the U.S. Customs Service. Delays in collection of the required documents and in their processing and shipment to the Census Bureau can extend the time between actual importation and inclusion in tabulations to as much as a year. This carryover problem was thought to be relatively minor, involving perhaps 15 to 20% of the documents. However, it was discovered that, in fact, in 1985 the carryover was closer to 45% each month.

*Present address: Agricultural Statistics Lab, University of Arkansas, Fayetteville, AR 72701

The Census Bureau's usual monthly foreign trade press release (U.S. Bureau of the Census, 1985) in September 1985 included a description of the carryover problem along with relevant tables showing the carryover for total imports and exports. Two new policies were announced. First, revised estimates of total imports and exports would be published one month after the original estimates had appeared and second, "true" monthly levels for each calendar year would be published after the fact when all of the necessary data were finally available. Detailed series would not be revised.

At the same time the problem was attacked from two directions; obtaining closer cooperation with the Customs Service to speed up the flow of documents and delaying the initial publication of the statistics to allow additional time to obtain the necessary data. As a result, revisions were discontinued in the Spring of 1987.

The Census Bureau's selected method of revision was oriented toward the estimates of import level and did not fully consider the estimation of the month to month change in the import level. In many economic applications the estimation of change is equally, or perhaps even more important. Our purpose here is to critically examine the Census Bureau's procedure for revising the initial estimates and to develop and evaluate an alternative derived from the viewpoint of estimating change.

In this paper we first describe the Census Bureau's revised estimator of the total value of imports. Two alternative estimators are then developed. The estimators of the month to month change in the import level corresponding to each of the estimators of the revised level are also studied. The development is sufficiently general to be applicable to any level of aggregation.

Since the final estimates are obtained, at least in principle, from a complete census of the population, sampling variability does not enter into consideration. A stochastic approach to the problem would necessarily be based on the assumption of a random mechanism generating the "true" monthly levels. We have chosen to utilize both deterministic and stochastic elements in our development. The problem of nonsampling errors is not addressed in our development and analysis of the estimators.

Two aspects of the general problem of revising an economic time series which will not be discussed in our context are alternative methods for constructing preliminary estimates and modeling the revision system in combination with the data series to produce improved estimates. The former is discussed in Rao, Srinath, and Quenneville (1986) and the latter in, for example, Howrey (1978) and Harvey, McKenzie, Blake, and Desai (1983).

2. Motivation and Expressions for the Estimators

In this section we describe the estimator of revised monthly import level and the resulting estimator of change in levels used by the Census Bureau. Two alternative estimators are also presented.

To establish notation, let

T_{ij} = the dollar value of imports for actual month of importation j which are first included in the report for month i,

T_{+j} = "true" import level for month j (the total which will eventually be reported),

T_{i+} = import level initially reported for month i,

C_i = carryover for month i (the data reported in month i which actually belong in some month prior to month i),

L_j = late reported data for month j (the data which belong in month j but are first available for tabulation and reporting in some month subsequent to month j).

We shall assume that $T_{ij} = 0$ for $j > i$; i.e., imports in a reported month cannot belong to any future actual month of importation. Note that

$$C_i = \sum_{j<i} T_{ij} \quad ,$$

$$L_j = \sum_{i>j} T_{ij} \quad ,$$

$$T_{i+} = \sum_{j} T_{ij} = T_{ii} + C_i \quad ,$$

$$T_{+j} = \sum_{i} T_{ij} = T_{jj} + L_j \quad .$$

We shall refer to T_{ii} as the correctly reported data for month i and $T_{i+1,i}$ as the one month late data for month i. Figure 1 illustrates the format of the data. Our goal is to construct a revised estimator of the true import level T_{+i} for month i which is better than the initial estimate T_{i+} by utilizing the data obtained in month i+1, $T_{i+1,j}$ for $j \leq i+1$.

In the above notation the estimator of the revised monthly import level for month i (published in month i+1) used by the Census Bureau is given by

$$T_i^{CB} = T_{ii} + T_{i+1,i} + \sum_{j<i} T_{i+1,j} \quad , \qquad (1)$$

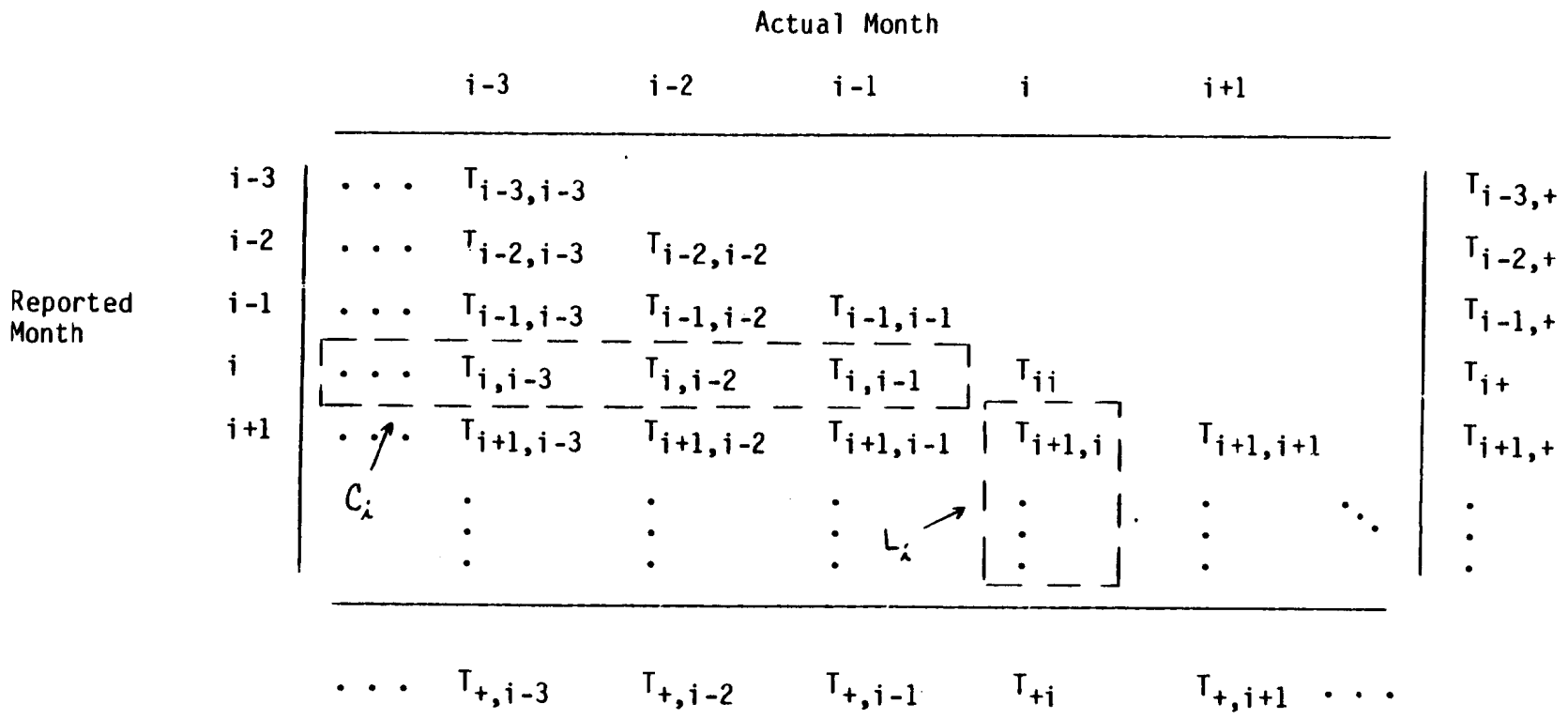

Figure 1. Import carryover data by actual month of importation and month reported.

which can be reexpressed as

$$T_i^{CB} = T_{i+1,+} + (T_{ii} - T_{i+1,i+1}) \quad .$$

Thus, the revised estimator for month i can be viewed as an adjustment of the original estimator for month i+1. In terms of the carryover we have

$$T_i^{CB} = T_{ii} + C_{i+1} \quad .$$

This formulation of T_i^{CB} shows that, in effect, the carryover in month i+1, C_{i+1}, is used as an estimate of the late reported data for month i, L_i.

In month i, the initial estimated import level for month i and the revised estimate for month i-1 are published. A natural estimator of change in the monthly level from month i-1 to month i is the difference in these estimates. Using the Census Bureau's revised estimator, we have the following estimator of change.

$$\begin{aligned} \Delta T_{i,i-1}^{CB} &= T_{i+} - T_{i-1}^{CB} \\ &= T_{ii} - T_{i-1,i-1} \quad . \end{aligned} \tag{2}$$

That is, the estimator of change is the difference between the correctly reported data for months i-1 and i. In 1985, this estimator of change utilized approximately 65% of the data which was eventually collected for each of the two months.

An alternative revised estimator is motivated by examining the assumptions underlying the estimator of change in levels. The true change in monthly import levels is given by

$$T_{+i} - T_{+,i-1} = (T_{ii} + L_i) - (T_{i-1,i-1} + L_{i-1}) \quad (3)$$
$$= (T_{ii} - T_{i-1,i-1}) + (L_i - L_{i-1}),$$

where

$$L_i - L_i = (T_{i+1,i} - T_{i,i-1}) + \left(\sum_{k>i+1} T_{ki} - \sum_{k>i} T_{k,i-1}\right) . \quad (4)$$

In month i, T_{ii} and $T_{i-1,i-1}$ are available but only the term $T_{i,i-1}$ in the difference $L_i - L_{i-1}$ is available. Hence, the difference in the late reported data $L_i - L_{i-1}$ must be estimated. Since $\Delta T^{CB}_{i,i-1} = T_{ii} - T_{i-1,i-1}$, the Census Bureau's estimator of $L_i - L_{i-1}$ is zero.

During the period when revisions were made, the one month late data $T_{i+1,i}$ accounted for approximately 30% of the actual level T_{+i} or approximately 75% of L_i. In month i the value of $T_{i,i-1}$ is available but that of $T_{i+1,i}$ is not until month i+1. If a good estimator $\hat{T}_{i+1,i}$ of $T_{i+1,i}$ can be obtained, then a change estimator of the general form

$$\Delta T_{i,i-1} = (T_{ii} - T_{i-1,i-1}) + (\hat{T}_{i+1,i} - T_{i,i-1}) \quad (5)$$

should be better than $\Delta T^{CB}_{i,i-1}$.

To this end, let γ_i be the proportional change in the one month late data from month i to month i+1. Then

$$\gamma_i = (T_{i+1,i} - T_{i,i-1}) / T_{i,i-1} , \quad (6)$$
$$\hat{T}_{i+1,i} = (1 + \gamma_i) T_{i,i-1} ,$$

and from (5),

$$\Delta T^{A}_{i,i-1} = (T_{ii} - T_{i-1,i-1}) + \gamma_i T_{i,i-1} . \quad (7)$$

Since $\Delta T^{A}_{i+1,i} = T_{i+1,i} - T^{A}_{i}$, the revised estimator of the import level for month i (published in month i+1) associated with $\Delta T^{A}_{i+1,i}$ is given by

$$T_i^A = T_{i+1,+} - \Delta T_{i+1,i}^A$$
$$= T_{i+1,+} - (T_{i+1,i+1} - T_{ii}) - \gamma_{i+1} T_{i+1,i} \qquad (8)$$
$$= T_i^{CB} - \gamma_{i+1} T_{i+1,i} \quad .$$

However, in month i+1, the value of $T_{i+2,i+1}$ is not available so that γ_{i+1} cannot be calculated. Thus, it must be replaced by an estimate $\hat{\gamma}_{i+1}$.

For any month i, the sequence of γ_k computed from previous months, say $\gamma_1,\ldots, \gamma_{i-1}$, can be used to predict γ_i. Since the γ_k-series is short, fitting an ARIMA model by Box-Jenkins techniques would be difficult. Alternatives include applying an exponential smoothing procedure such as Holt-Winters (Gardner, 1985), fitting a simple structural model (Harvey and Todd, 1983), and stepwise fitting of autoregressive models (Newbold and Granger, 1974). The latter reference compares the performance of several of these methods.

Two alternative estimators of γ_i are considered here. The first is based on a Holt-Winters exponential smoothing and the second on the fitting of simple regression models with lagged variables. The models

$$\gamma_k = b_0 + b_1\gamma_{k-1} + e_k \qquad \text{for } k=2,\ldots, i-1,$$

and

$$\gamma_k = b_0 + b_1\gamma_{k-1} + b_2\gamma_{k-2} + e_k \qquad \text{for } k=3,\ldots, i-1,$$

where the e_k are iid$(0,\sigma^2)$, were originally entertained.

The natural extension of the idea of estimating $L_i - L_{i-1}$ by $\hat{T}_{i+1,i} - T_{i,i-1}$ is to estimate the entire set of differences $T_{ki} - T_{k-1,i-1}$, k=i+1, i+2,... . While this extension would provide an estimate of the entire difference $L_i - L_{i-1}$ rather than some portion of it, the resulting estimator may not be an improvement. To estimate the

difference in the two month late data $T_{i+2,i} - T_{i+1,i-1}$, a two step ahead prediction would be required from a series of proportional changes which is one time period shorter than the γ_k-series. Each subsequent term of $L_i - L_{i-1}$ requires further ahead prediction from a shorter series. Estimation of the tail terms is made more difficult by two practical considerations; the number of nonzero terms in each L_i need not be the same and the change in character of the data generating the terms. The terms of L_i which represent the longest lags are obtained primarily from documents which are unreconcilable edit failures returned to Customs, shipments held by Customs for inspection and testing, and delays related to classification schedule changes rather than from the (apparently random) set of documents which are simply delayed by the processing burden at Customs. The behavior of these processes is more volatile and hence, more difficult to model.

3. A Numerical Comparison of the Estimators

An empirical comparison of the relative performance of the estimators described in Section 2 was conducted. Monthly data on the customs value of total imports cross-classified by month of importation and month of initial reporting were available for January 1983 through December 1986. Since the true totals were required for comparison purposes, only data through June 1986 were included. Data prior to July 1983 were used only to fit the lagged variable regression models. Hence, 36 months of data were used in the study. The data for the customs value of imports are similar to those for c.i.f. (customs value, insurance, and freight) which can be found in U.S. Bureau of the Census (1985, 1986).

Plots of the true and originally reported import levels and

changes in level are given in Figures 2 and 3, respectively. The customs values are given in millions of dollars. From these figures it is clear that, especially for month to month change, the originally reported estimates can be substantially different from the true values and that some form of revised estimates are needed. Figure 4 shows the correctly reported data T_{ii} and the carryover C_i as a percentage of the reported level. From Figure 4, for each month in the study period a large proportion of the initially reported estimate consisted of carryover data. The problem reached its peak in the first two months of 1985 when the majority of the initial estimates were obtained from carryover rather than correctly reported data.

Four estimators were included in the study; the originally reported estimates (no revision), the Census Bureau revised estimator, and the two alternative estimators. A plot of the γ_k-series is shown in Figure 5. The entire series was fit to the lagged regression models described in Section 2. Based on these initial fits, the model

$$\gamma_k = b_1\gamma_{k-1} + b_2\gamma_{k-2} + e_k \quad , \qquad k=3,\ldots, i-1, \tag{9}$$

was selected. For each of the 36 months in the study, the model (9) was fit and the one step ahead prediction $\hat{\gamma}_i$ was used to construct the estimators T_i^{LR} and $\Delta T_{i,i-1}^{LR}$. The estimators T_i^{ES} and $\Delta T_{i,i-1}^{ES}$ were based on an exponentially smoothed estimate of γ_i from SAS's (1984) PROC FORECAST.

The basis for comparison of the competing estimators is the absolute relative error

$$ARE(\hat{\tau}) = | \hat{\tau} - \tau | / | \tau | \ ,$$

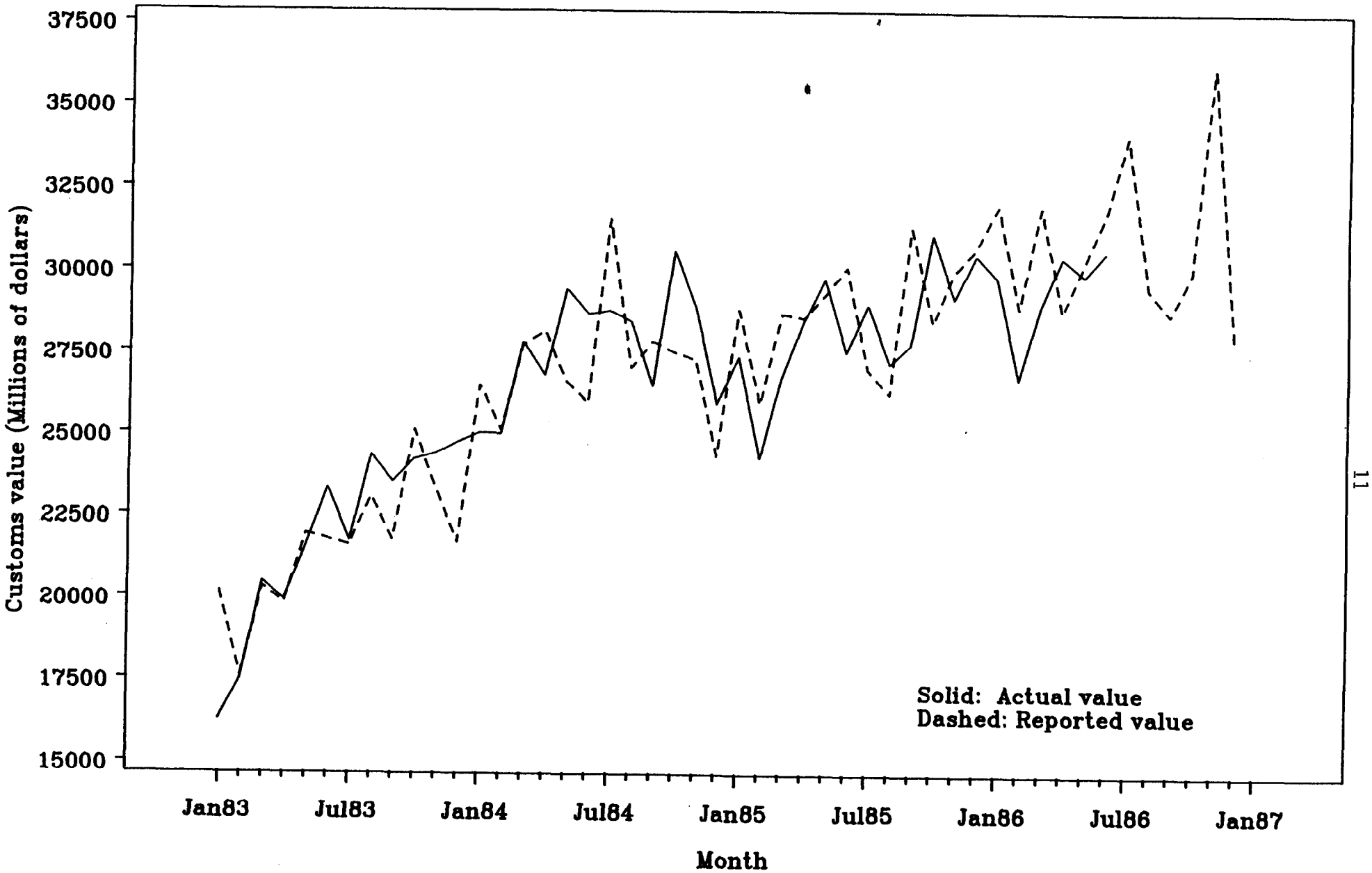

Figure 2. Import levels for customs value.

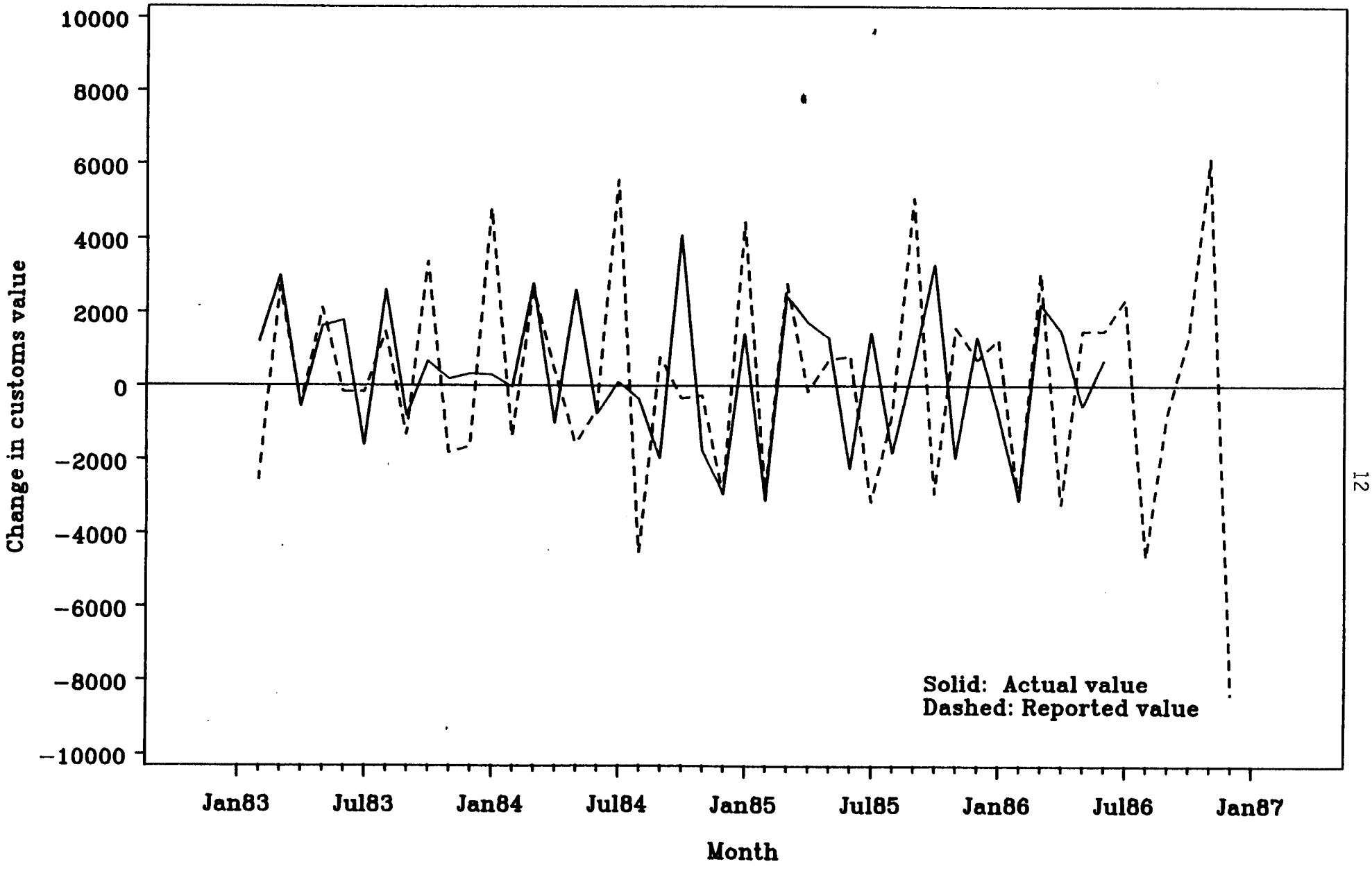

Figure 3. Change in import levels for customs value.

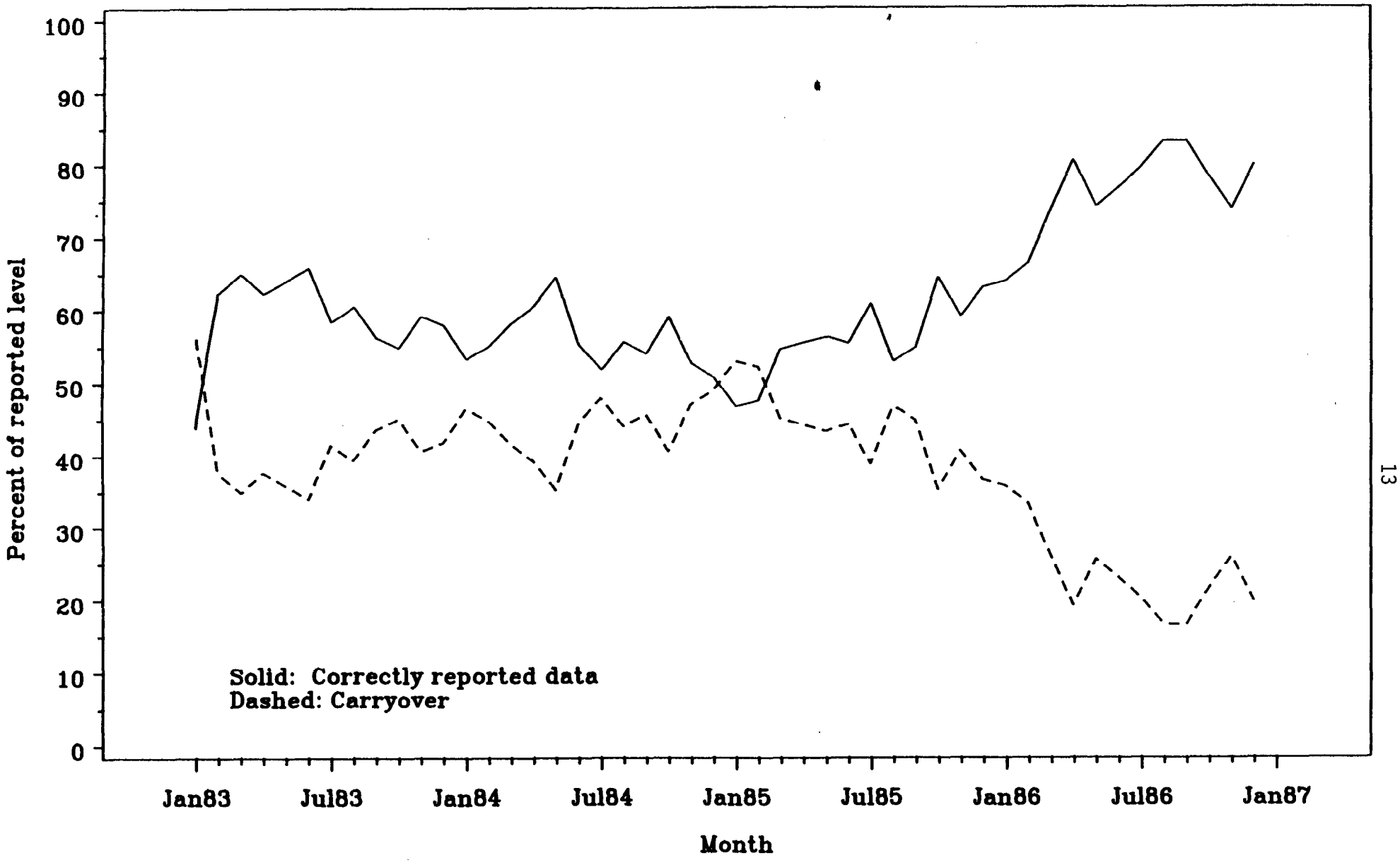

Figure 4. Correctly reported data and carryover as percent of reported import level for customs value.

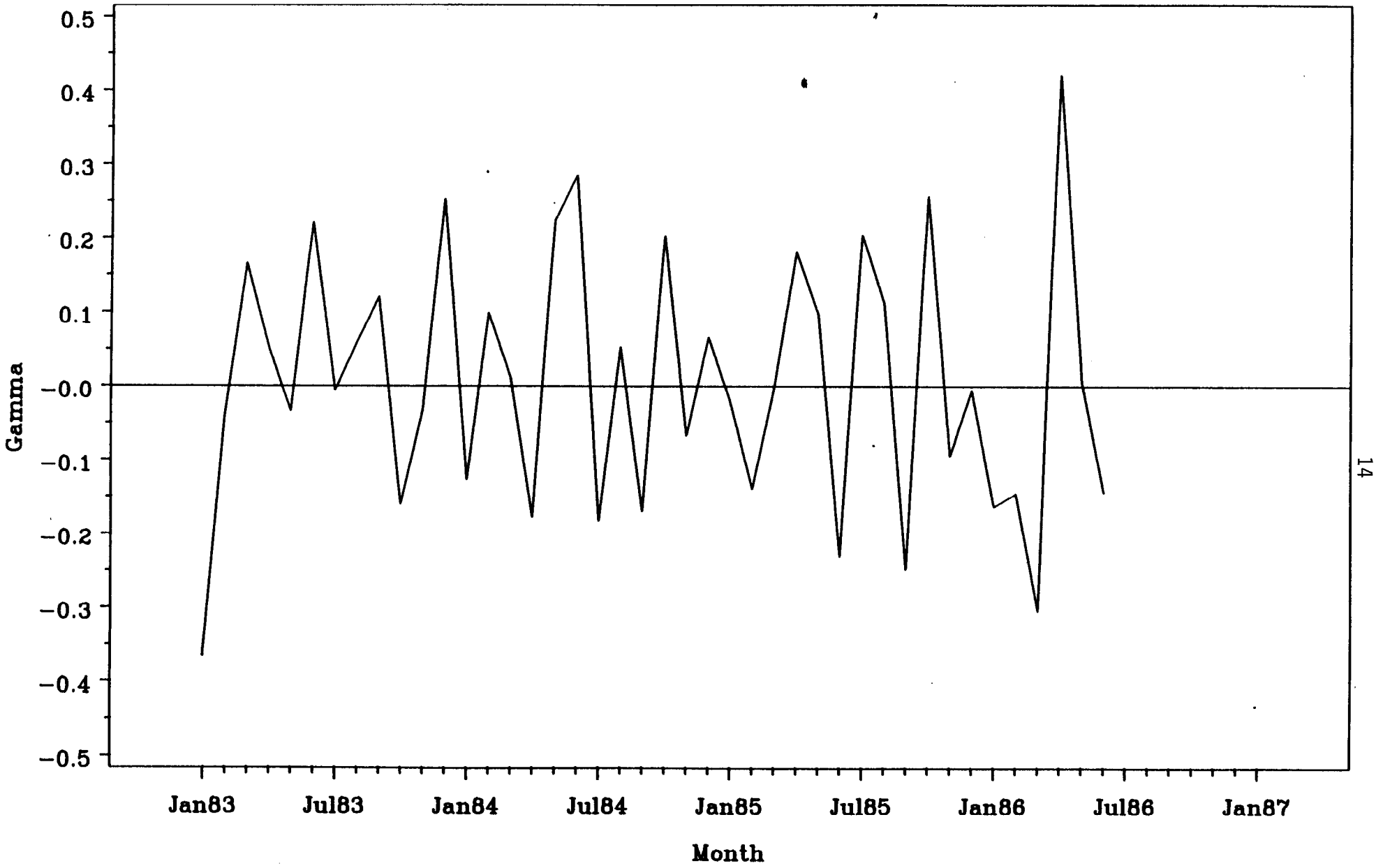

Figure 5. Proportional change in one month late data for customs value.

where $\hat{\tau}$ is the estimator of the parameter τ . A small absolute relative error indicates a good estimator. The ARE criterion is applied to both the estimators of level and change in level.

Boxplots of the distributions of the AREs are shown in Figure 6 for the estimators of the import level and in Figure 7 for the estimators of change in level.* For estimation of both level and change, each of the three methods of revision is an improvement over the unrevised statistics. From Figure 7, both alternative estimators are comparable to the Census Bureau's estimator of change. Although all of the change estimators are outlier prone, the two alternatives appear to be the least affected. For estimation of import level, the Census Bureau and the exponentially smoothed estimators are the best.

Since neither of the alternatives were clearly superior to the Census Bureau's estimator, which at first glance is not intuitively appealing, the assumptions underlying it were analyzed. For the estimator of import level, the carryover in month i+1 is used as an estimate of the late reported data for month i. For the estimator of change, the difference in the late reported data for consecutive months is estimated by zero. The carryover C_{i+1} and late reported data L_i are plotted in Figure 8, which indicates close agreement between the two series.

Let

$$\delta_i = \frac{C_{i+1} - L_i}{L_i} \qquad \text{for } i=1,\ldots, T-1$$

*A number of extremely large ARE for the change estimators were not plotted but were used in the construction of the plots. They are: Unrevised 10.6, 15.1, 39.8, 45.4, 12.1; Census 15.7; Regression 21.1; Smoothed 21.2.

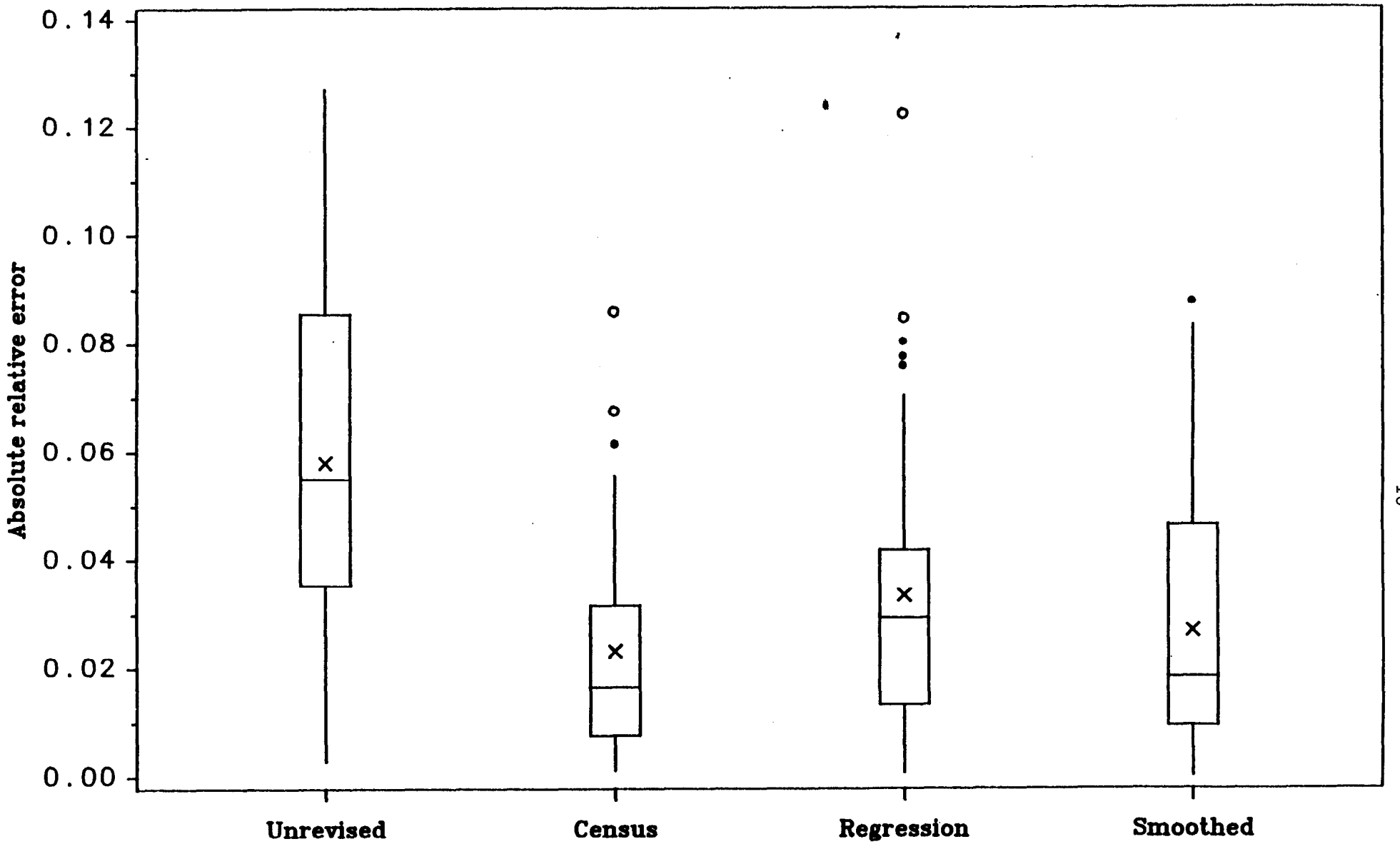

Figure 6. Absolute relative errors of the estimators of import level.

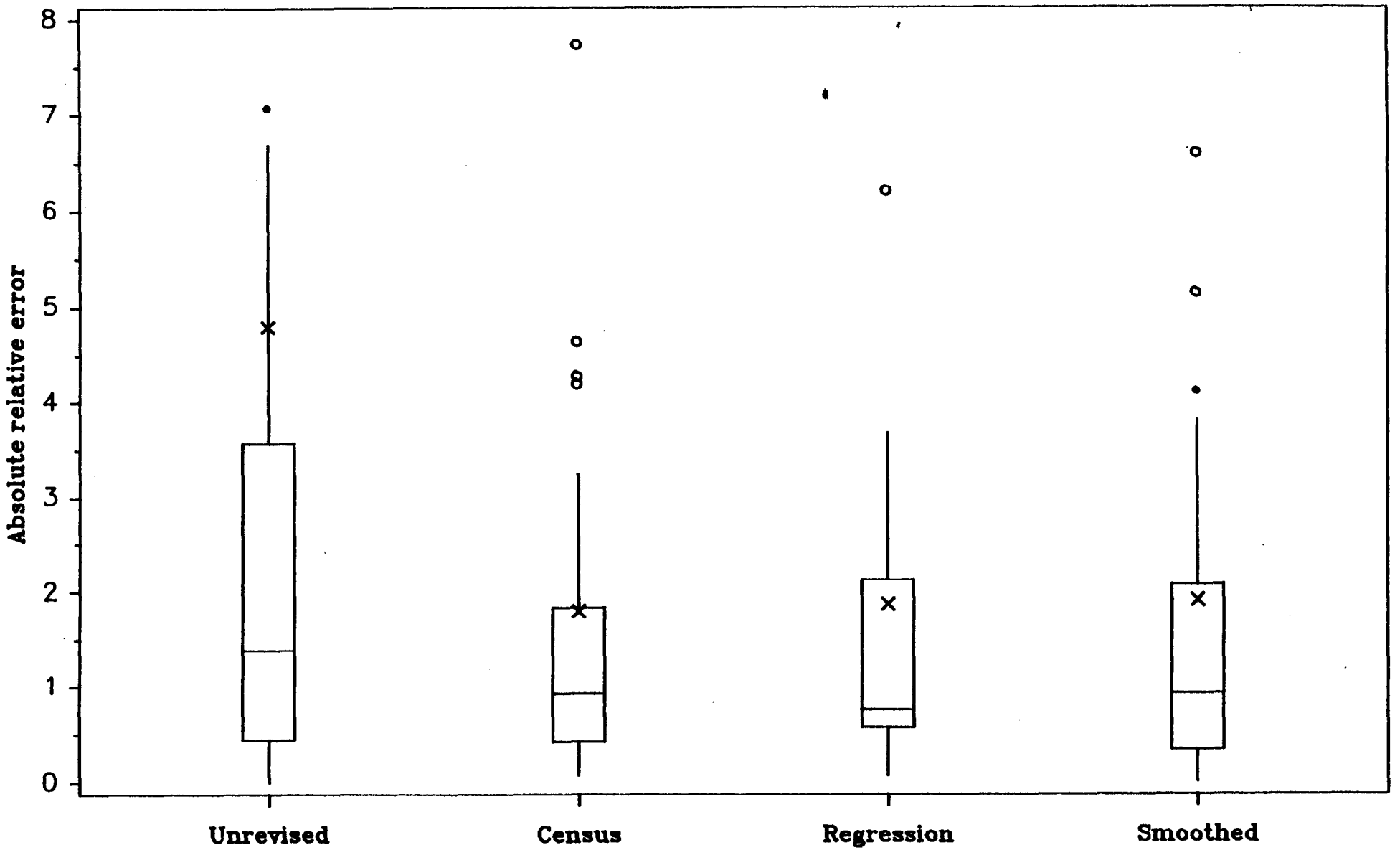

Figure 7. Absolute relative errors of the estimators of change in level.

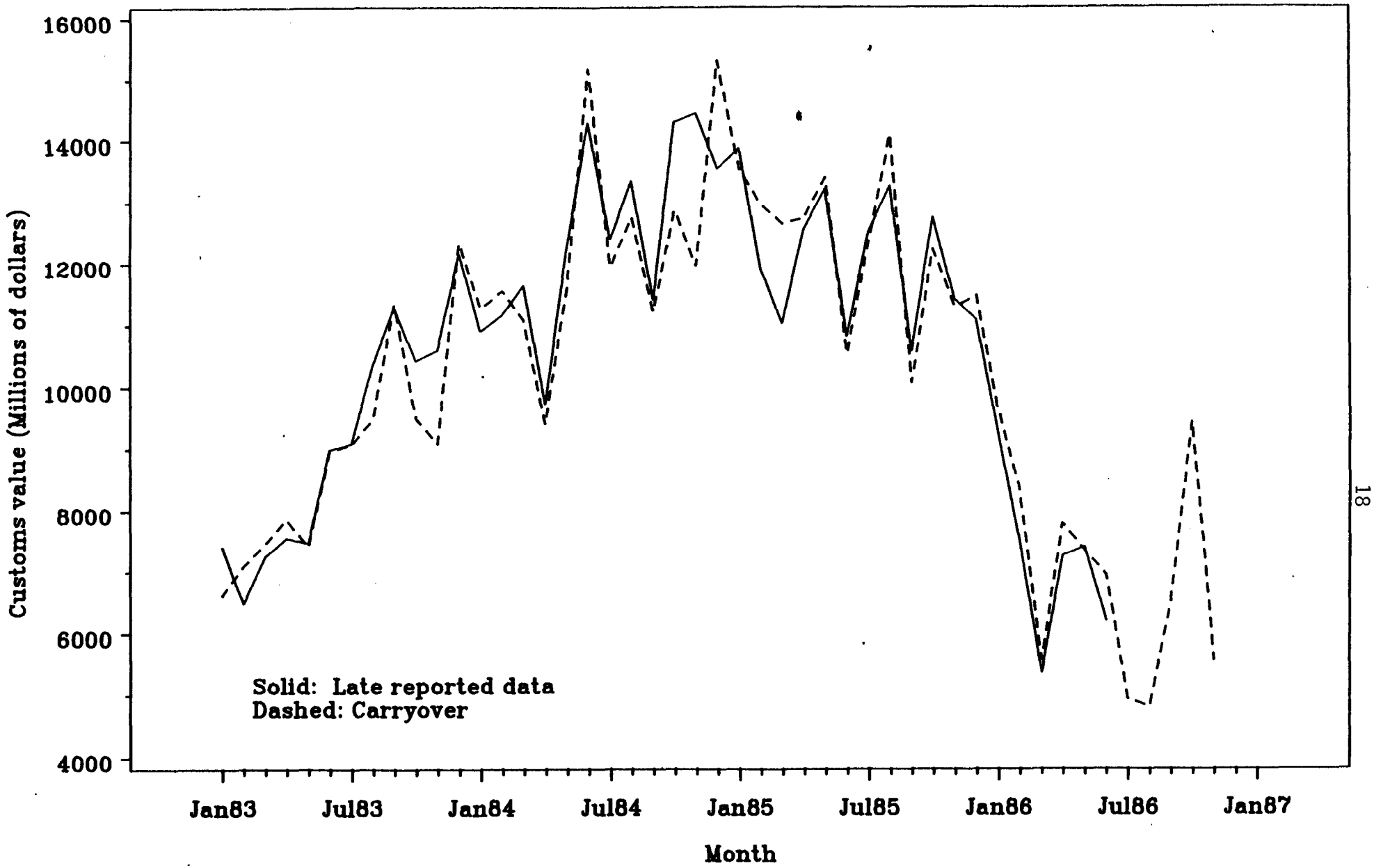

Figure 8. Carryover and late reported data for customs value.

and

$$\lambda_i = \frac{L_{i+1} - L_i}{L_i} \qquad \text{for } i=1,\ldots, T-1 \ .$$

Boxplots of the δ_i and λ_i are shown in Figure 9, which indicate relatively small ranges for both distributions. The p-values for t-tests of H_0: $E(\delta) = 0$ and H_0: $E(\lambda) = 0$ regarded as descriptive statistics, were 0.86 and 0.78, respectively. Thus, the assumptions underlying the Census Bureau estimators are not untenable, making the estimators competitive with the more statistically based alternatives.

4. Conclusion

The carryover problem in foreign trade statistics and the operationally oriented Census Bureau method of revision of the aggregate level series have been described. Two alternative estimators based on the proportional change in the one month late data have been developed. Neither of these alternatives is markedly better when applied to data on the dollar value of total imports. Examination of the data indicates that the assumptions underlying the Census Bureau estimator are not unreasonable, thereby explaining its relatively good performance.

Acknowledgments

Nash Monsour made several helpful comments during the early stages of this research. Bruce Walter provided information on the data described in Section 3. Diane Bailie provided computer programming assistance and Kevin Thompson assisted in constructing the figures.

References

Gardner, E. S. (1985). Exponential smoothing: The state of the art. **J. of Forecasting**, 4, 1-38.

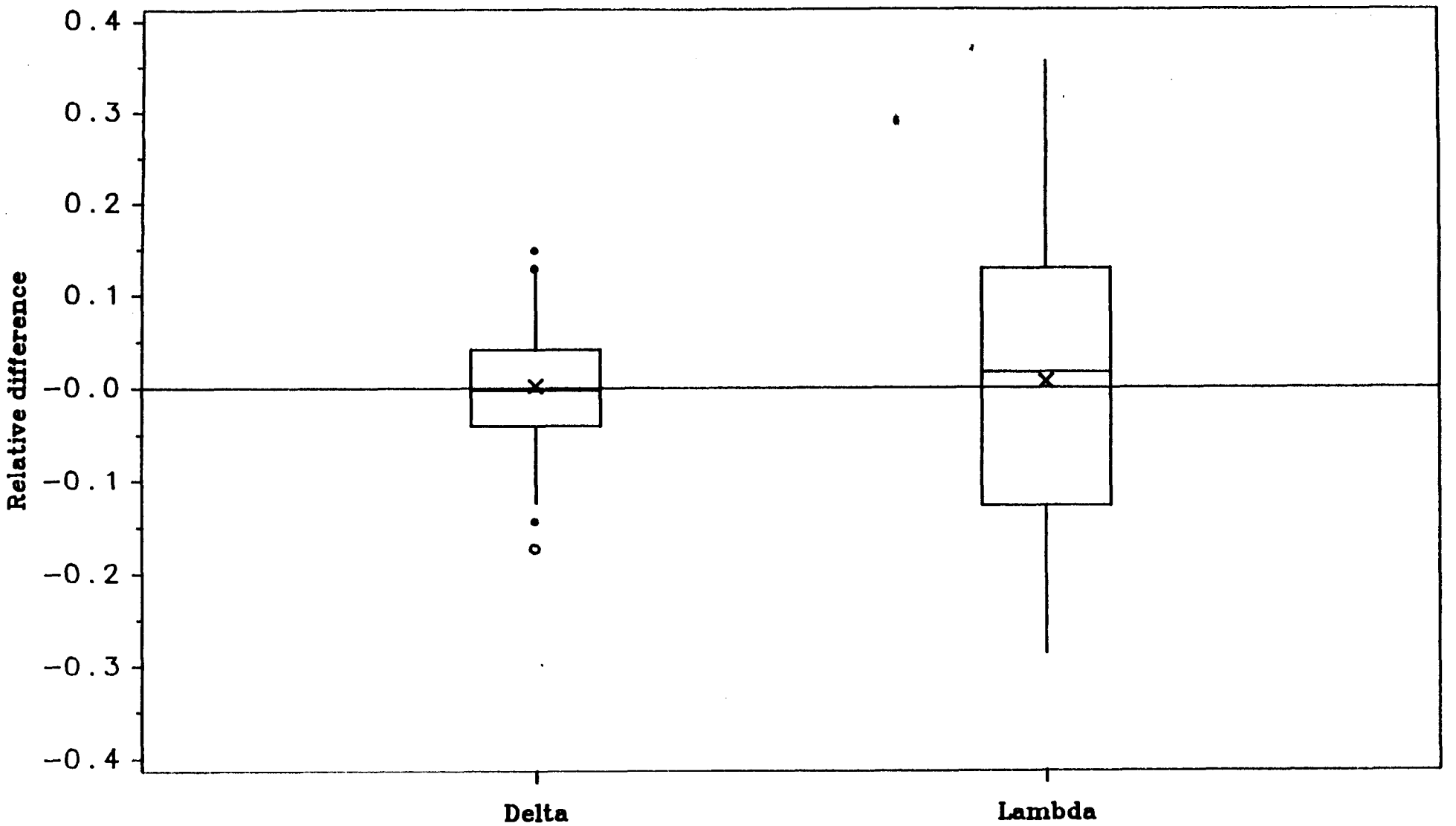

Figure 9. Relative differences between carryover and late reported data (Delta) and between late reported data for consecutive months (Lambda).

Harvey, A. C. , C. R. McKenzie, D. P. C. Blake, and M. J. Desai (1983). Irregular data revisions. In **Applied Time Series Analysis of Economic Data**, ed. A. Zellner. U.S. Department of Commerce, Bureau of the Census, Economic Research Report, ER-5, 329-347.

Harvey, A. C. and P. H. J. Todd (1983). Forecasting economic time series with structural and Box-Jenkins models: A case study. **J. of Business and Econ. Stat.**, 1, 299-315.

Howrey, E. P. (1978). The use of preliminary data in economic forecasting. **Rev. of Econ. and Stat.**, 60, 193-200.

Newbold, P. and C. W. J. Granger (1974). Experience with forecasting univariate time series and the combination of forecasts. **J. Royal Stat. Soc. A**, 137, 131-164.

Rao, J. N. K., K. P. Srinath, and B. Quenneville (1986). Optimal estimation of level and change using current preliminary data. **Proc. of International Symposium on Panel Surveys**, to appear.

SAS/ETS User's Guide (1984). SAS Institute Inc., Cary, NC.

U.S. Bureau of the Census (1985). **Summary of the U.S. Export and Import Merchandise Trade , Series FT900 ADV** (85-08).

U.S. Bureau of the Census (1986). **Summary of the U.S. Export and Import Merchandise Trade, Series FT900 ADV** (86-01).