# An Empirical Comparison of Benchmarking Methods for Economic Stock Time Series

Irene Brown

U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

**Abstract**

Economic Programs at the U.S. Census Bureau use the Causey-Trager method for benchmarking monthly and quarterly series to annual series and to the Economic Census every five years. This procedure uses an iterative, nonlinear constrained optimization technique to obtain the benchmarked series. Although constrained optimization works well for flow variables, Statistics Canada suggests using a different method for stock variables based on inconsistencies they found in their final benchmarked series. They use a simple method of interpolating the ratios of benchmark value to the original series value with a natural cubic spline and then apply the interpolated ratios to the original series to obtain the benchmarked series. Using historical economic time series from the Census Bureau, this paper investigates the effects of the two different methods in terms of level estimates of the benchmarked series, ratios of benchmarked series to original series, and preservation of the original series' period-to-period changes. Lastly, we examine stock to flow ratios (e.g. inventory to sales).

**Keywords:** benchmarking, time series, natural cubic splines, interpolation

## 1. Introduction

The U.S. Census Bureau's Economic Programs provide statistics about U.S. businesses and government organizations. ([www.census.gov/econ/progoverview.html](www.census.gov/econ/progoverview.html)) These statistics are obtained from an Economic Census and Census of Governments every five years and from over 100 separate surveys taken monthly, quarterly, and annually, including twelve principal economic indicators. Many of these economic surveys have the same target population as another survey but data are gathered at different frequencies (e.g. Monthly Retail Trade Survey and Annual Retail Trade Survey).

Generally, the different frequency of series balances the need for comprehensive, reliable, detailed statistics versus the need for timeliness. The Economic Census and annual surveys provide the reliable detailed level estimates, where the monthly and quarterly surveys track the higher frequency period-to-period changes. Benchmarking combines the information from both surveys into a consistent higher frequency time series by preserving some characteristic of the original higher frequency series (e.g. the period-to-period changes) while attaining the levels of the less frequent series. The majority of benchmarking methods use constrained minimization techniques to preserve the movement of the original series.

Dagum and Cholette (2006) suggest that these constrained minimization methods can actually alter the movement of the original series instead of preserving it when the

benchmarks are not temporally contiguous. "Benchmarks are temporally contiguous when the starting date of each benchmark is equal to the ending date of the previous benchmark plus one day" (Dagum 2006). Flow series are generally temporally contiguous based on their definition. They measure a variable over a time interval (e.g. sales) and therefore the benchmarks are the sum of the values over the coverage or benchmark period (e.g. January to December). Stock series measure the level of a variable at a point in time (e.g. inventories) and therefore the benchmarks are not the sum but a value at a specific time, so the span of the benchmark period is one time period. (e.g. December to December). In addition, when benchmarking annual figures to the quinquennial census figures, we consider all series types as "stock" series. For example, the sales value for the 2007 Census is equal to the 2007 annual value not the sum of sales from 2003 to 2007. By definition, stock series will never be temporally contiguous. Therefore, Statistics Canada suggests using natural cubic spline interpolation to benchmark stock series.

This paper investigates the claim that for stock series the constrained optimization methods violate the movement preservation principle. The paper also presents results from an empirical study using current economic stock time series from the Census Bureau that compares Statistics Canada's natural cubic spline method and a modified Denton's (1971) proportional first difference method to the Census Bureau's Causey-Trager method for benchmarking stock series. Lastly, the effect of using different benchmarking methods for each component of the stock to flow ratios (i.e. inventories to sales) is examined.

## 2. Causey-Trager Benchmarking Method

Causey and Trager provided a numerical algorithm to solve for the non-linear constrained minimization benchmarking problem (Bozik 1988), which has been predominantly used at the Census Bureau since the early 1980s. Define the higher frequency series as the original series and denote as:

$$x_t, \ t = 1, 2, \cdots, n.$$ 
(1)

Define the less frequent series as the benchmarks and denote as:

$$T_k, \ k = 1, 2, \cdots, m.$$ 
(2)

The indices $t$ and $k$ represent time periods that map to specific dates. Each benchmark covers a span of time periods from the original series, where $b_k$ is the beginning time period for the $k$th benchmark and $e_k$ is the ending time period for the $k$th benchmark such that

$$t = 1 \leq b_1 \leq e_1 < b_2 \leq e_2 < \cdots < b_m \leq e_m \leq n$$

for $k = 1, 2, ..., m$ non-overlapping benchmark periods. For stock series, $b_k$ always equals $e_k$. This means that there are time periods between $e_k$ to $b_{k+1}$, therefore making the series not temporally contiguous. Benchmarking finds a new series that is defined as the benchmarked/revised series and denote as:

$$y_t, \ t = 1, 2, \ldots, n$$ 
(3)

Such that:

$$\sum_{t=b_k}^{e_k} y_t = T_k, k = 1, \ldots, m.$$ 
(4)

The Causey-Trager (CTM) method attempts to preserve the period-to-period changes by employing the iterative technique of steepest feasible descent to minimize the non-linear function

$$F_{CTM} = \sum_{t=b_1}^{e_m-1} \left( \frac{y_{t+1}}{y_t} - \frac{x_{t+1}}{x_t} \right)^2 . \tag{5}$$

The starting point for the iterative procedure comes from the solution to a modified proportionate first differences method proposed by Denton (1971), which minimizes the function

$$F_{PFD} = \sum_{t=b_1}^{e_m-1} \left( \frac{y_{t+1}}{x_{t+1}} - \frac{y_t}{x_t} \right)^2 . \tag{6}$$

For those time periods before and after the benchmarks (i.e. $t < b_1$ and $t > e_m$), carry backward and carry forward factors are used. The carry backward factor is calculated and applied as

$$y_t = x_t * \frac{y_{b_1}}{x_{b_1}}, \text{ for } t = 1, ..., b_1 - 1 \tag{7}$$

and the carry forward factor is calculated and applied as

$$y_t = x_t * \frac{y_{e_m}}{x_{e_m}}, \text{ for } t = e_m + 1, ..., n \tag{8}$$

Figure 1 shows a simulated quarterly stock series and the Causey-Trager (C-T) revised series given five annual benchmarks. The graph shows a very seasonal pattern present in the original series, but any distortion to the period-to-period changes in the revised series is hard to determine. Dagum and Cholette (2006) say that the presence of seasonality often obscures when a violation of movement preservation occurs.
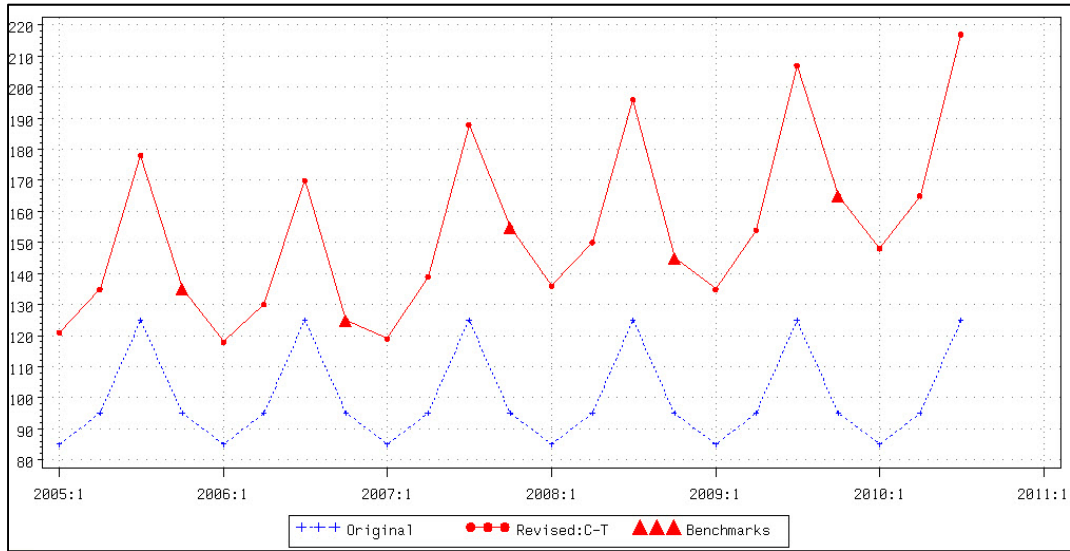


**Figure 1:** Original and Revised Stock Series Benchmarked using the Causey-Trager Method (CTM)

After removing the seasonal component from the original simulated quarterly series in Figure 1, a constant stationary series emerges. When benchmarking this constant series to the same annual benchmarks, a graph of the revised series shows a very similar pattern to a graph of the ratio of revised series to original series ($\frac{y_t}{x_t}, t = 1, \ldots, n$) across time periods for the original seasonal series. Therefore, Statistics Canada suggests reviewing the graphs of the ratio of revised to original for series when seasonality is present to determine if any movement violation has occurred.
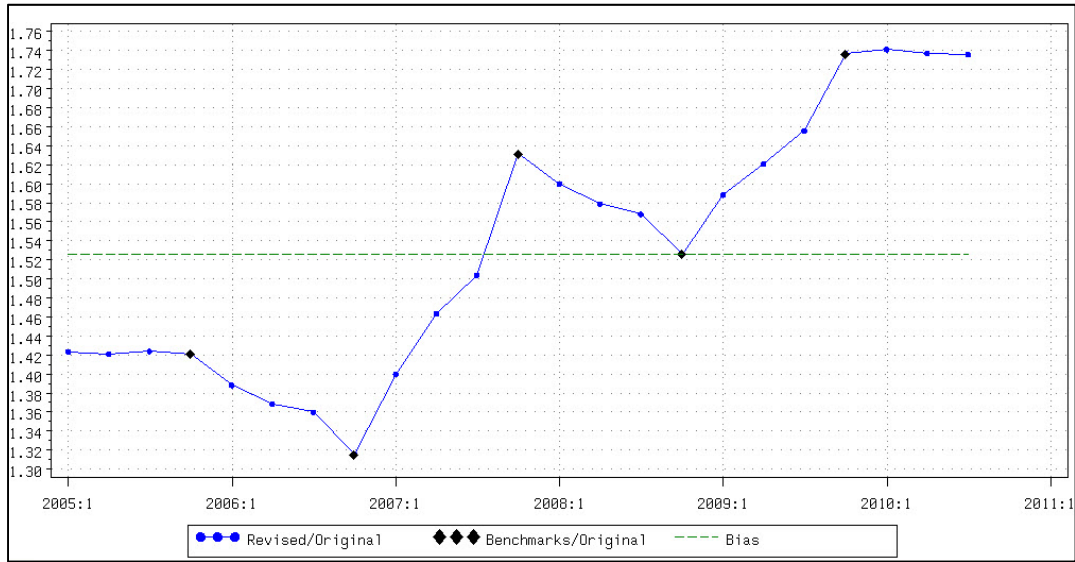


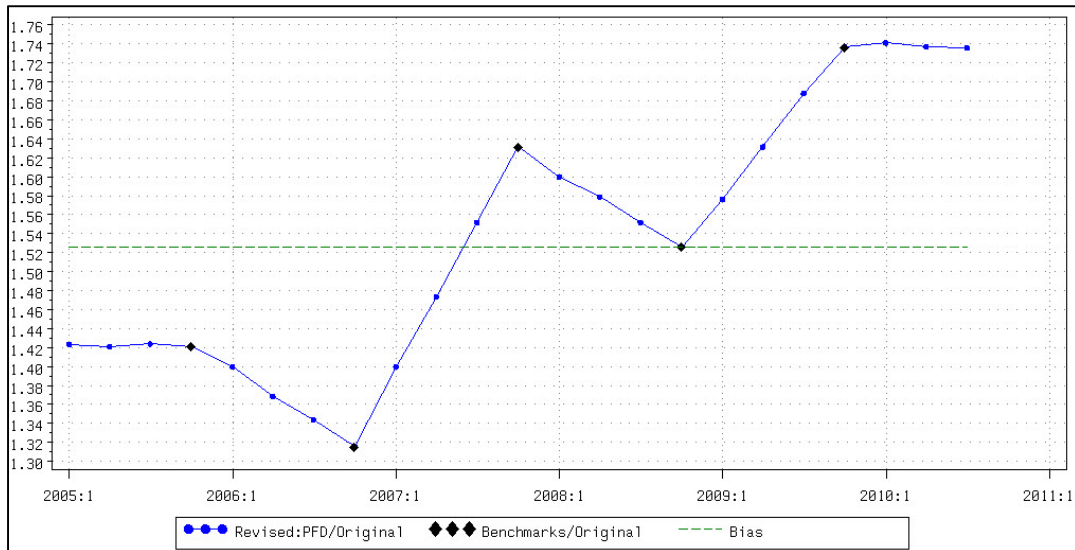**Figure 2:** Revised to Original Ratio Graph for CTM Method



**Figure 3:** Revised to Original Ratio Graph for PFD Method

The revised to original graph for the Causey-Trager (CTM) method (Figure 2) shows abrupt changes at the benchmarks and "kinks" in the movement with the largest jump being from 3rd quarter 2007 to 4th quarter 2007. In general, a large jump between points in this graph means there was a larger revision to the period-to-period change between those

time periods. The revised to original graph for the modified Denton (PFD) method (Figure 3) has less "kinks" but still has abrupt changes at the dates of the benchmarks.

## 3. Natural Cubic Splines Benchmarking Method

Statistics Canada implements benchmarking with a user-defined SAS® procedure: PROC BENCHMARKING (Latendresse 2007). The main method, based on Dagum and Cholette (2006), obtains the revised series as the solution of a constrained minimization problem. They have found that their constrained optimization methods can produce unlikely "kinks" in the benchmarked stock variable. Review of the revised to original ratio graphs of their benchmarked series show these strange movements. Statistics Canada believes a smooth curve passing through the benchmark to original discrepancies makes more economic sense for benchmarked stock series. (See Figure 4)
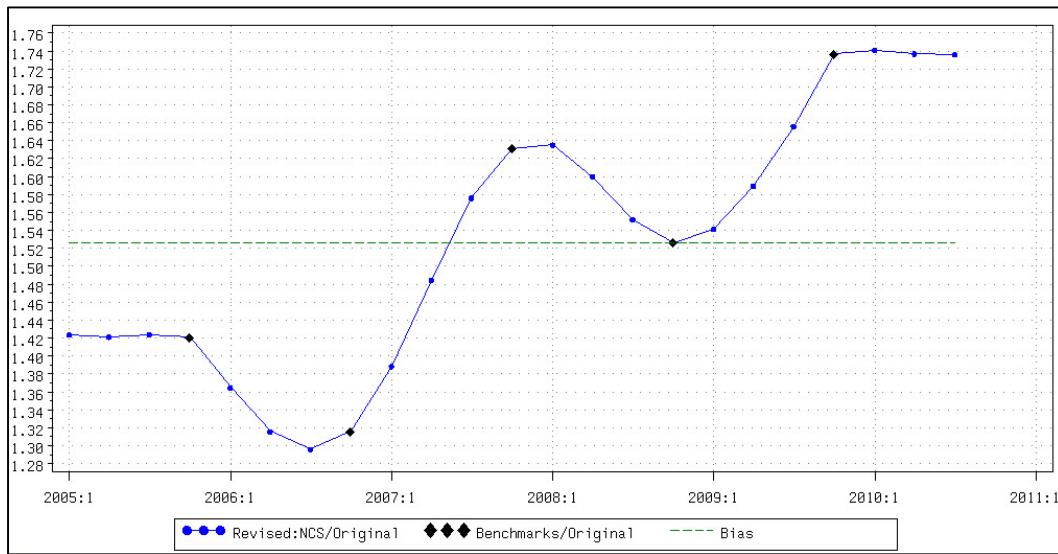


**Figure 4:** Revised to Original Ratio Graph for NCS Method

A smooth curve can be obtained with natural cubic splines and they suggest it as an alternate method for benchmarking, interpolation, and temporal distribution (Quenneville 2010). For benchmarking stock series, they use a simple method of interpolating the ratios of benchmark value to the original series value with a natural cubic spline and then apply the interpolated ratios to the original series to obtain the revised series. This is done in three general steps:

1. Calculate the annual/benchmark discrepancies:
$$z_k = \frac{T_k}{x_{b_k=e_k}}, \text{ for } k = 1, \ldots, m$$

2. Interpolate the missing ratios with a natural cubic spline using the Expand procedure in SAS:
$$\text{Output is } dr_t, \text{ for } t = b_1 = e_1, \ldots, b_m = e_m$$

3. Convert to original scale:
$$y_t = x_t * dr_t, \text{ for } t = b_1 = e_1, \ldots, b_m = e_m$$

The EXPAND procedure in SAS converts times series from one frequency to another and interpolates missing values in the time series. For example, quarterly estimates can be interpolated from an annual series. By default, the procedure fits cubic spline curves and the output series are generated from the spline approximations. "A cubic spline is a segmented function consisting of third-degree (cubic) polynomial functions joined together so that the whole curve and its first and second derivatives are continuous." (SAS 2004) In the context of benchmarking above, a function $S : [z_1, z_m] \rightarrow \Re$ is a natural cubic spline if $S$ is constrained to pass through the $m$ given data points such that $z_1 < z_2 < \ldots < z_m$, $S$ is a cubic polynomial between each pair of points, $S$ is continuously twice differentiable, and $S''(z_1) = S''(z_m) = 0$. (Quenneville 2010)

As Figure 4 shows, the curve is smoother in the revised to original graph for the natural cubic spline (NCS) benchmarking method. Growth rate preservation diagnostics challenge the assumption that the curve with the "kinks" violates movement preservation. For example, boxplots, by year and method, of the absolute revisions to the period-to-period changes ($U_t^{Method} = \left| \dfrac{y_t^{Method}}{y_{t-1}^{Method}} - \dfrac{x_t}{x_{t-1}} \right|, \text{for } t = 2, \ldots, n$) suggests that the smoother curve (NCS method) has larger and more variable absolute revisions than the other methods (CTM and PFD). (See Figure 5)
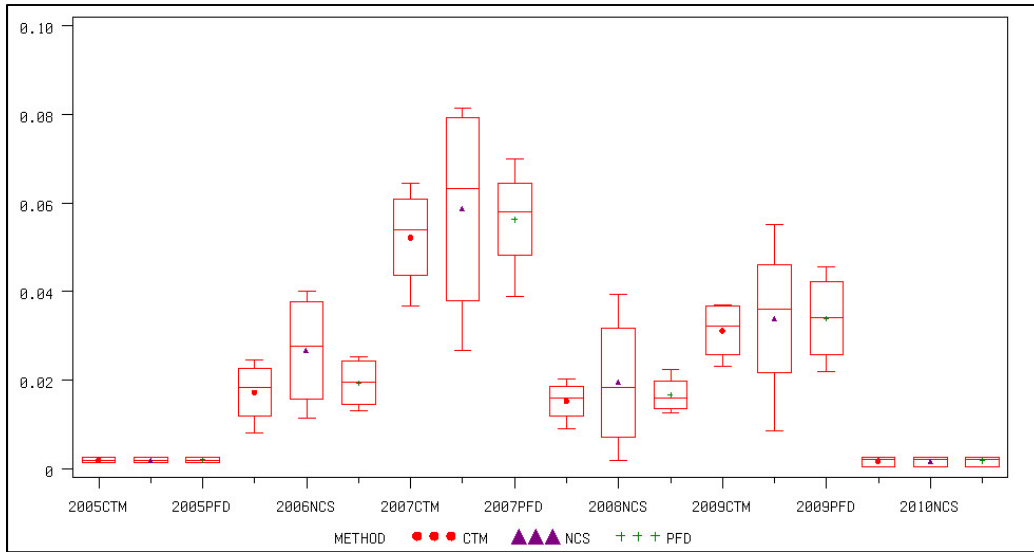


**Figure 5:** Comparison Boxplots by Year and Method of the Absolute Revisions

In addition, the absolute revisions can be compared at each time period by the difference between methods calculated as

$$D_t^{Method} = U_t^{Method} - U_t^{CTM}, t = 2, \ldots, n .$$

Negative differences show that the CTM method has the larger absolute revision. Table 1 shows the average, median, minimum, and maximum of the differences in absolute revisions across the series between the Causey-Trager method and the other two methods. Overall, on average the CTM method has smaller absolute revisions. The differences in the absolute revisions of the NCS method from the CTM method appear to be larger and more variable than the CTM method from the PFD method.

**Table 1:** Difference in Absolute Revisions from the Causey-Trager Method

|  | NCS | PFD |
|---|---|---|
| **Average** | 0.0041 | 0.0018 |
| **Median** | 0 | 0 |
| **Minimum** | -0.0376 | -0.0254 |
| **Maximum** | 0.0446 | 0.0332 |

## 4. Empirical Study

To compare the two types of benchmarking methods (constrained optimization or natural cubic splines) for stock series, 860 series from the following three annual surveys and three monthly surveys were used.

- Annual Retail Trade Survey (ARTS)
  http://www.census.gov/retail/index.html#arts
- Annual Wholesale Trade Survey (AWTS)
  http://www.census.gov/wholesale/index.html
- Service Annual Survey (SAS) http://www.census.gov/services/index.html
- Monthly Retail Trade Survey (MRTS)
  http://www.census.gov/retail/index.html#mrts
- Monthly Wholesale Trade Survey (MWTS)
  http://www.census.gov/wholesale/index.html
- Manufacturers' Shipments, Inventories, and Orders (M3) Survey
  http://www.census.gov/manufacturing/m3/

For each original series, three revised series were found using the following benchmarking methods: the Causey-Trager (CTM), the Natural Cubic Splines (NCS), and the modified Denton (PFD). To make the series more comparable, all methods used carry backward and carry forward factors to calculate the revised series values for those time periods not covered by benchmarks (see equations (7) and (8)).

## 4.1 Smoothness of Revised to Original Ratio Graphs
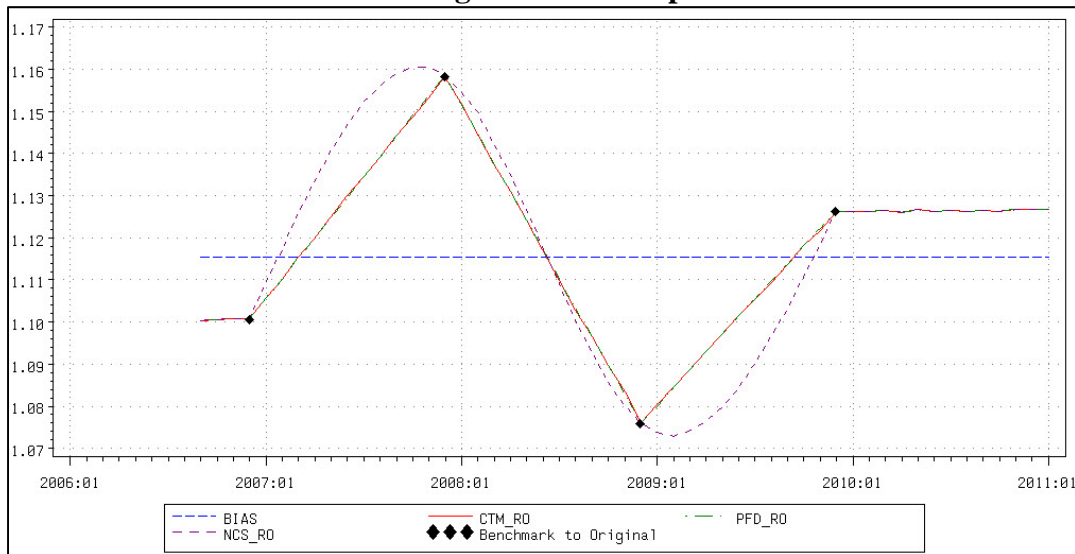


**Figure 6:** Revised to Original Graph (All Methods) for a MRTS Series

The smoothness of the ratio plot cannot be summarized in a qualitative measure, so graphs for each series contianing all three methods were created and reviewed. Overall, for the monthly series the natural cubic spline method (NCS) of benchmarking stock series did achieve a smoother curve for the revised to original graphs. (See Figure 6) Although, when the benchmark to original discrepancies follow a linear pattern (consistently going up or down) the resulting interploation of ratios between them is linear and produces revised series that are very similar for all methods. (see Figure 7) Therefore, the revsied to original ratios graphs are all similar also.
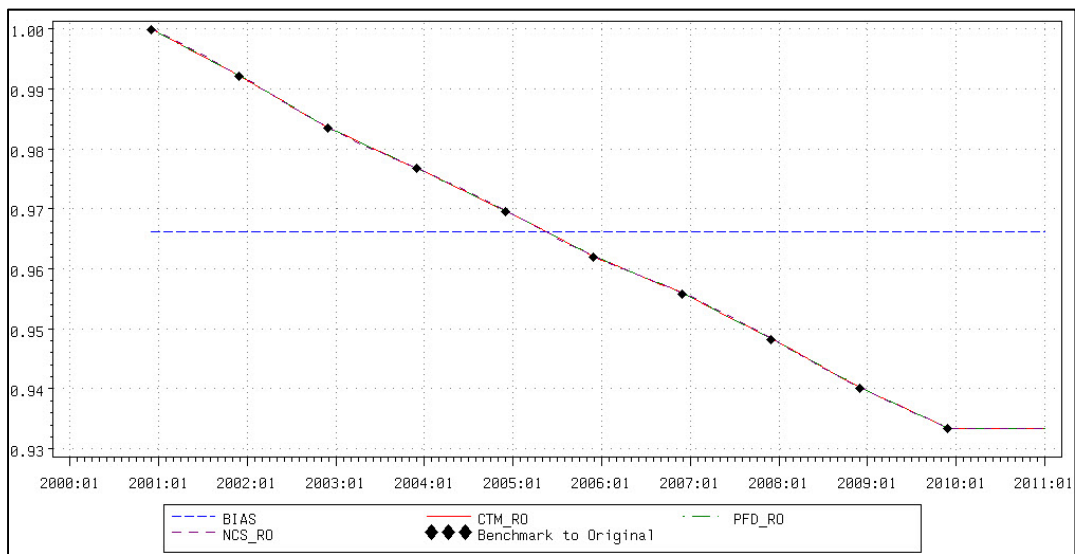


**Figure 7:** Revised to Original Graph (All Methods) for a M3 Series
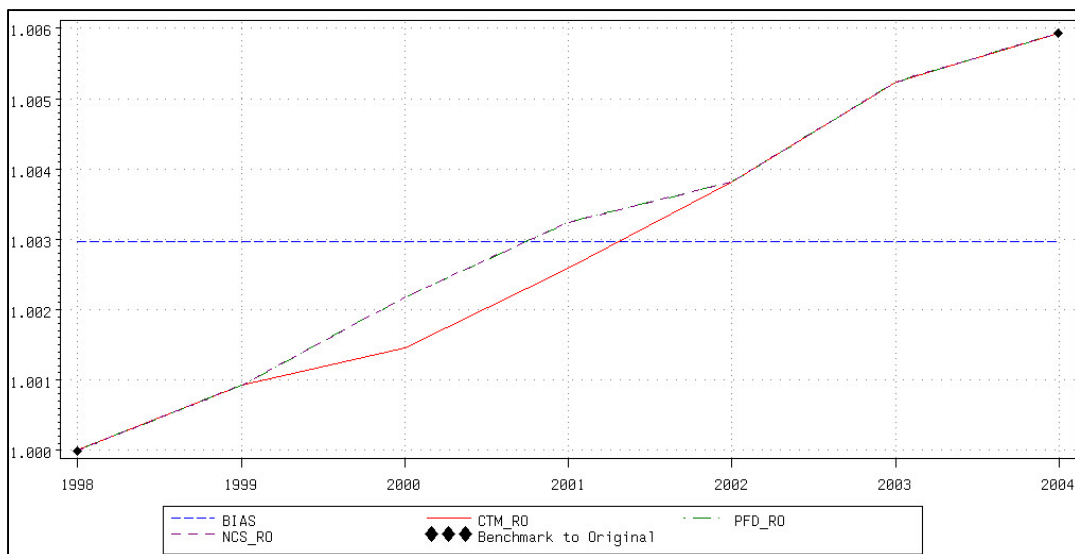


**Figure 8:** Revised to Original Graph (All Methods) for a ARTS Series

For the annual series, the NCS and PFD methods were almost always the same with only a few differences from the CTM graphs. (See Figure 8) The nature of fitting cubic splines and the number of data points explains the differences between the annual and monthly surveys. Generally, when benchmarking annuals to the Economic Census there are only two benchmark periods thus limiting the number of points through which to fit the natural

cubic spline. SAS (2004) documentation warns, "The accuracy of the results produced by PROC EXPAND may be somewhat less at the ends of the output series than at time periods for which there are several input values at both earlier and later times." Therefore, PROC EXPAND is more likely to find linear interpolations between two points instead of a curve and thus obtain results very close to the PFD and CTM methods.

To quantify the similarity of the methods and their resulting revised series, the absolute percent difference was calculated as

$$pct\_diff_t^{method1,method2} = 100 * \frac{\left| y_t^{method1} - y_t^{method2} \right|}{0.5 * \left( y_t^{method1} + y_t^{method2} \right)}, t = 1,...,n.$$

The maximum from each series was used to calculate the mean and maximum absolute percent difference for each survey. (See Table 2) For the annual surveys, the series appear to be very similar for all the methods with the NCS and PFD being the closest to each other. A different pattern emerges for the monthly surveys, where the CTM and PFD are closer to each other than to the NCS revised series. This confirms quantitatively what I observed in looking at the revised to original graphs.

**Table 2:** Maximum Absolute Percent Difference between Methods

| Survey | Number of Series | NCS and CTM | | NCS and PFD | | PFD and CTM | |
|--------|------------------|-------------|-----|-------------|-----|-------------|-----|
| | | Mean | Max | Mean | Max | Mean | Max |
| ARTS | 121 | 0.21 | 4.65 | 0.01 | 1.04 | 0.21 | 4.65 |
| AWTS | 74 | 0.56 | 4.69 | 0.01 | 0.06 | 0.56 | 4.68 |
| SAS | 188 | 0.97 | 24.24 | 0.01 | 0.09 | 0.96 | 24.24 |
| MRTS | 69 | 1.22 | 18.18 | 1.22 | 18.18 | 0.50 | 18.18 |
| MWTS | 19 | 0.89 | 2.72 | 0.89 | 3.01 | 0.14 | 1.41 |
| M3 | 199 | 2.84 | 24.36 | 2.79 | 24.30 | 0.74 | 17.32 |

## 4.2 Growth Rate Preservation

To evaluate which method achieves the smallest value of the Causey-Trager objective function (equation (5)), an index purposed by Di Fonzo and Marini (2010) was used. Both the NCS and PFD methods were compared to the CTM method by calculating:

$$r_2^{Method} = \left( \frac{\sum_{t=1}^{n-1} \left| \frac{y_{t+1}^{Method}}{y_t^{Method}} - \frac{x_{t+1}}{x_t} \right|^2}{\sum_{t=1}^{n-1} \left| \frac{y_{t+1}^{CTM}}{y_t^{CTM}} - \frac{x_{t+1}}{x_t} \right|^2} \right)^{\frac{1}{2}}$$

The index was calculated from the rounded revised series, the value used for publications. When the index is less than one, it indicates the PFD or NCS method preserves the growth rate of the series better. As expected, a large percentage of series show the CTM method getting smaller revisions to the period-to-period changes. Although, a small handful of series did have an index $r_2$ less than one suggesting that when rounded the PFD and NCS method can have smaller revisions to the growth rates. This mainly occurred in the ARTS and SAS surveys. None of the monthly series for the NCS method had a value of less than one.

Comparison boxplots, one for each of the six surveys and two methods (NCS and PFD), of the $r_2$ index were produced to determine the overall pattern of differences in the methods. For the annual surveys, the patterns for both the PFD and NCS method were right skewed with the indices being very close to one (See Figure 9). This reconfirms that the NCS method does not produce a "smoother" curve when only two benchmark periods are used in a stock series. It also suggests that for all but a few series the revised series and growth rate preservation are very close to the CTM method. These few series generally had more up and down movement in the time periods between benchmarks.
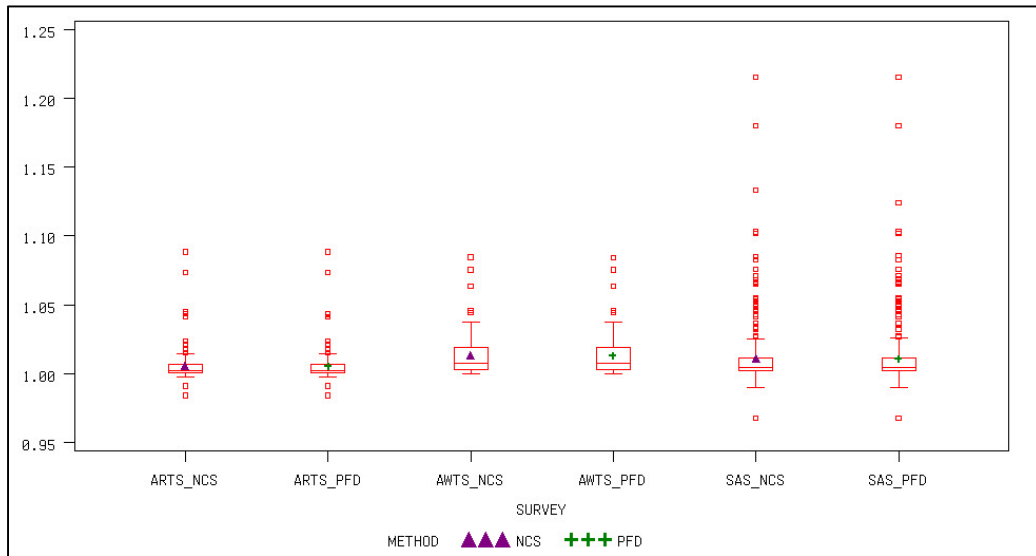


**Figure 9:** Boxplots of Index $r_2$ for Annual Surveys by Method

However, a distinctly different pattern emerged for the monthly surveys when using the NCS method. The boxplots appear normally distributed with many more series having larger values of the objective functions for the NCS method than the CTM method. (See Figure 10) The right skewed pattern holds for the monthly surveys and the PFD method.
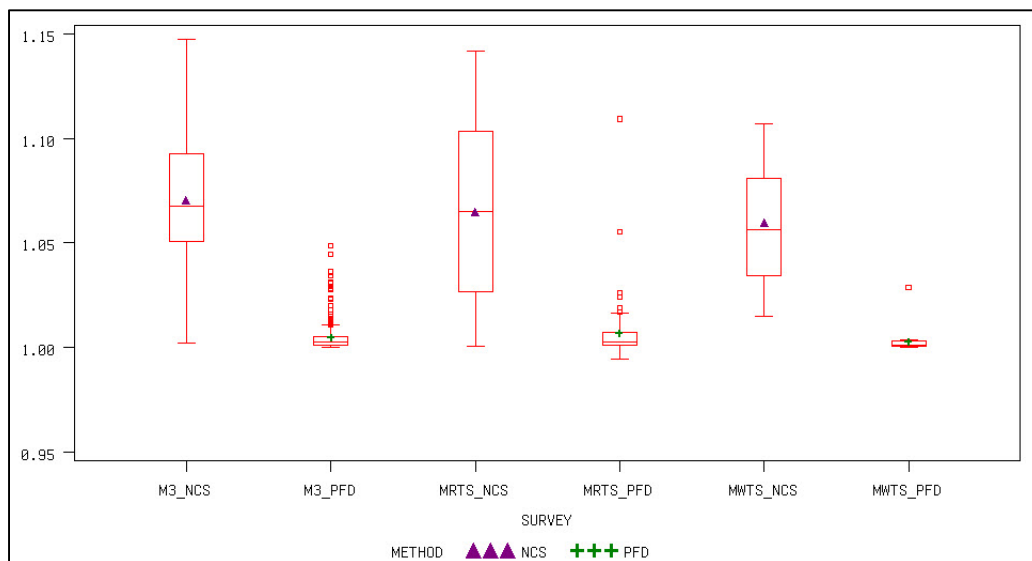


**Figure 10:** Boxplots of Index $r_2$ for Monthly Surveys by Method

An interesting pattern emerged when reviewing the absolute revisions in growth rates calculated as

$$U_t^{Method} = \left| \frac{y_t^{Method}}{y_{t-1}^{Method}} - \frac{x_t}{x_{t-1}} \right|, \text{ for } t = 2,...,n .$$

Figure 11 shows the absolute revisions for the three methods plotted across time for a monthly wholesale series. Within each year, the revisions appear consistent and similar for both the CTM and PFD methods but the NCS method shows large variability. Figure 5 in Section 3 shows this same pattern with boxplots by year and method. Also of note is that the sizes of the absolute revisions do not match the size of the growth rate. For example, a large absolute revision does not necessarily occur for the larger growth rates. This cyclical pattern for the NCS absolute revisions appeared in a large percentage of the monthly series. As the revised series between the three methods becomes more alike, this pattern disappears. Therefore, it was not prevalent in the annual series.
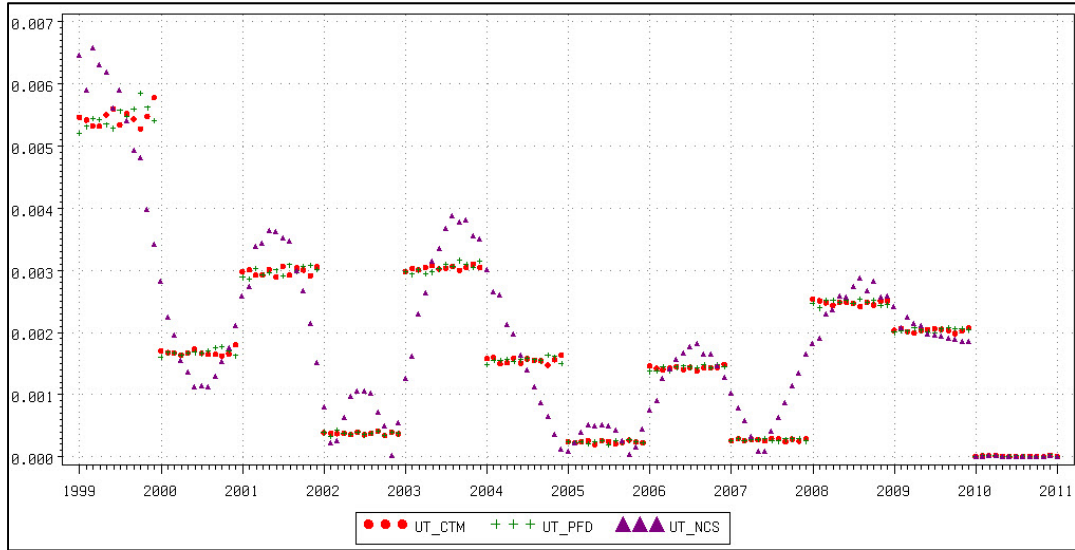


**Figure 11:** Graph of Absolute Revisions to Period-to-Period Changes by All Methods for a MWTS Series

## 4.3 Stock to Flow Ratios
A common economic ratio published is inventories to sales, which is the most common example of a stock to flow ratio. The original stock to flow ratio is calculated as

$$sfr_t^{orig} = \frac{x_t^{stock}}{x_t^{flow}}, \text{ for } t = 1,...,n .$$

Generally, the benchmarked ratio is calculated as

$$sfr_t^{CTM} = \frac{y_t^{stock,CTM}}{y_t^{flow,CTM}}, \text{ for } t = 1,...,n ,$$

where both the flow and stock benchmarked values come from the same benchmarking method. To compare the effect on the ratio of different benchmarking methods on the stock series, the ratio will be calculated using the same benchmarked flow series and the three stock benchmarking methods as

$$sfr_t^{method} = \frac{y_t^{stock,method}}{y_t^{flow,CTM}}, \text{ for } t = 1,...,n,$$

where the index *method* is one of three methods: *CTM, PFD,* and *NCS*.

The absolute differences between benchmarked ratios from the original ratio will be calculated as

$$diff_t^{method} = |sfr_t^{orig} - sfr_t^{method}|, \text{ for } t = 1,...,n.$$

To find which method is closest to the original ratio, the maximum difference was found for each series. Table 3 shows the percentage of series by method with the minimum maximum absolute difference. The CTM method clearly achieves the minimum for the majority of series in all three surveys. Thus suggesting that the Causey-Trager method achieves the closer original stock to flow ratios most often.

**Table 3:** Percentage of Series with Minimum Maximum Difference by Survey and Method

| Survey | Number of Series | CTM | PFD | NCS |
|--------|------------------|-------|-------|-------|
| M3 | 199 | 63.97 | 20.59 | 15.44 |
| MRTS | 69 | 80.60 | 5.97 | 13.43 |
| MWTS | 19 | 47.37 | 31.58 | 21.05 |

## 5. Conclusion

The natural cubic spline (NCS) method for benchmarking times series more often achieved a "smoother" curve through the benchmark to original discrepancies for the monthly series. Although, for the annual series there was little difference from the PFD method because of only having two benchmark periods (not enough knots for natural cubic splines). Based on this criterion only, the NCS method would be better for benchmarking a number of the monthly stock series but no improvement can be found for the annual series.

For growth rate preservation criteria, the Causey-Trager method (CTM) did a better job overall. This is especially apparent in the graphs of the absolute revisions, where the NCS method had greater variability within each year. In addition, all but a small handful of the $r_2$ indices comparing the values of the Causey-Trager objective functions by method were greater than one. For the monthly series, this index was more often further away from one for the NCS method than close to one like the PFD method. When defining movement preservation as growth rate preservation, the constrained optimization methods did not appear to violate the principle. The study also shows that using the same benchmarking method (CTM) for both components of the stock to flow ratios preserves the original ratio more often. More research is needed to determine which method to use for stock series. It depends greatly on which philosophy your organization prefers, the proportional movement preservation or growth rate preservation.

## Acknowledgements

Quenneville from Statistics Canada for example SAS code and information on benchmarking stock series using natural cubic splines.

All graphs and analysis were produced and preformed using SAS® 9.1.3.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

## References

Bozik, J.E. and M.C. Otto (1988), "Benchmarking: Evaluating Methods that Preserve Month-to-Month Changes." Statistical Research Report Series 88/07, U.S. Census Bureau: Washington, DC. Available online at www.census.gov/srd/papers/pdf/rr88-07.pdf.

Dagum, E. B. and Cholette, P. A. (2006), *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series,* New York: Springer, Lecture Notes in Statistics 186.

Denton, F.T. (1971), "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization." *Journal of the American Statistical Association,* 82, 99-102.

Di Fonzo, T., Marini, M. (2010), "*Benchmarking and movement preservation. Evidences from real-life and simulated series.*" Department of Statistical Sciences, University of Padua, Working Paper Series, n. 14. http://www.stat.unipd.it/ricerca/fulltext?wp=422

Latendresse, E., Djona, M. and Fortier, S. (2007), "Benchmarking Sub-Annual Series to annual Totals – From concepts to SAS® Procedure and SAS® Enterprise Guide® Custom Task", Proceedings of the *2007 SAS Global Forum.*

Quenneville B., Picard, F., Fortier, S., "Interpolation, Benchmarking and Temporal Distribution with Natural Splines." In JSM proceedings, Business and Economic Statistics Section. Alexandria, VA: American Statistical Association, 2010.

SAS Institute Inc. 2004. *SAS/ETS® 9.1 User's Guide.* Cary, NC: SAS Institute Inc.