A comparison of new and established benchmarking methods

Jennifer Davies, Duncan Elliott (ONS)
John Aston, Homesh Sayal (University of Cambridge)

August 6, 2015

Abstract

Benchmarking is a well established method used widely in official statistics to ensure consistency between time series of differing frequencies. For example, in the production of UK National Accounts a number of areas constrain quarterly time series to sum over the year to a related annual time series from a different source. The UK Office for National Statistics (ONS) currently use the Cholette-Dagum regression based method for benchmarking.

This paper compares two well established benchmarking methods (Denton and Cholette-Dagum) to a novel wavelet based approach to benchmarking. Wavelet benchmarking can be useful for handling more complicated benchmarking problems due to the time and frequency localisation properties of wavelets. These different benchmarking methods are compared using both real ONS time series and simulated data through a range of quality measures relevant to both users and producers of official statistics.

Key words: wavelets, benchmarking, national accounts

1 Introduction

Benchmarking is a well established method used widely in official statistics to ensure consistency between time series of differing frequencies. One example is in the production of UK National Accounts where a number of areas constrain quarterly time series to sum over the year to a related annual time series from a different source. Typically the quarterly series provide good estimates of growth rates, but not necessarily levels, which are better captured by the annual time series. Another example of benchmarking at ONS are the Workforce Jobs series which are benchmarked on the third quarter of each year to employment data from the Business Register and Employment Survey (BRES), a survey with a larger sample size (ONS (2013)). Currently ONS use the Cholette-Dagum (Cholette and Dagum (1994)) regression based method for this type of benchmarking.

Benchmarking is also used in the context of seasonal adjustment. For certain statistics it is desirable that the annual totals of a seasonally adjusted series equal the annual total of the unadjusted series. In these cases the method of Denton (1971) is also used within the X-13ARIMA-SEATS software (U.S. Census Bureau (2015)) to constrain annual totals of seasonally adjusted series to the annual total in the non-seasonally adjusted data. The aim of the Denton method is to preserve growth rates when benchmarking which is appropriate in this context.

Wavelet methods have been applied in statistics since the late 1980s (Nason (2010)). These methods have been used to develop a method for wavelet benchmarking detailed in the paper of Sayal et al. (2014). This is a novel approach to benchmarking with the potential benefit of the time and frequency localisation properties of wavelets being able to handle more complicated benchmarking problems.

This paper compares the established benchmarking methods of Denton, Cholette-Dagum with the wavelet based approach using both real ONS time series and simulated data. The methods are compared through a range of quality measures relevant to both users and producers of official statistics.

The paper continues as follows. Section 2 introduces some terminology that will be used throughout the paper. Section 3 outlines the benchmarking methods of Denton (Denton (1971)), Cholette-Dagum (Cholette and Dagum (1994)) and wavelet benchmarking (Sayal et al. (2014)). Section 4 details the results of the comparison of the benchmarking methods. Section 5 discusses the results and possible future work.

2 Terminology

The following terminology will be used throughout.

- Path series - high frequency series to be constrained.
- Benchmark series - low frequency series of constraints.
- Benchmarked series - high frequency series after being benchmarked which now meets the constraints of the benchmark series.
- Aggregation structure - relationship between the path and benchmark series. Typically the path will sum to or average to benchmark series, or be equal to the benchmark series at a point in time for example the last quarter of the year.

3 Methods

3.1 Denton

The original method of Denton (Denton (1971)) with the subsequent modification of Cholette (Cholette (1979)) aims to produce a benchmarked series which preserves movements in the path series.

The Denton method is used indirectly for benchmarking in practice in ONS to constrain seasonally adjusted annual totals to the corresponding total of the non-seasonally adjusted series.

3.2 Cholette-Dagum

The method of Cholette-Dagum (Cholette and Dagum (1994)) puts the problem of benchmarking into a regression framework. The observed path series is assumed to be the true path with a bias and error term. The benchmark series is assumed to be an aggregation of the true path plus an error term. The errors in the path and benchmark series are assumed to be uncorrelated.

The Cholette-Dagum method had additive and proportional variants. While the error process can be specified, typically an AR(1) process is used for flow series and an AR(2) process for stock series. At ONS the autoregressive parameters are typically set to 0.8 for monthly series and 0.512 for quarterly.

The method of Cholette-Dagum is used in practice in ONS in the compilation of the UK National Accounts where accounting constraints need to be met. An extension of the method of Cholette-Dagum has also been implemented which restricts benchmarked series to be positive if they should not take negative values.

3.3 Wavelet benchmarking

Using wavelets for benchmarking is a new, novel approach to the benchmarking problem presented in Sayal et al. (2014). The method uses the wavelet decomposition of the path and benchmark series then replaces the low frequency wavelet coefficients in the path with those in the benchmark series to meet the benchmark constraints.

Two methods of wavelet benchmarking have been developed and are tested in this paper: elementary wavelet benchmarking and wavelet thresholding. Thresholding builds on the elementary method by removing wavelet coefficients that are determined to be noise.

The localised nature of wavelets means that this method could have advantages in benchmarking path series containing anomalies such as outliers and level shifts.

4 Comparison of methods

4.1 Simulations

A simulation study was run to compare the results of benchmarking using the following four methods with the benchmark constraint that the high frequency (monthly) series sums to the low frequency (annual) series.

- Proportional modified Denton with one order of differencing
- Proportional Cholette-Dagum with and AR(1) error process with autoregressive parameter of 0.8.
- Elementary wavelet benchmarking
- Wavelet benchmarking with thresholding

The simulations were run as follows. 79 seasonally adjusted, monthly time series from the Index of Production (IoP) were selected as the true path. The benchmark series are the annual totals of the monthly series. Noise was generated by taking the irregular component of the IoP series from a multiplicative time series decomposition including a trend and irregular component. The true paths were multiplied by the noise series to produce observed paths. Observed paths were then benchmarked to benchmark series.

The resultant benchmarked series were compared using the following metrics:

- Did the benchmarking method produce a result?
- Did the benchmarked series meet the constraints?
- Average mean squared difference over time and series between the benchmarked series and observed path
- Average mean squared difference over time and series between the benchmarked series and true path
- Percentage of month-on-month growth rates in benchmarked series that are in the same direction as the corresponding growth rate in the observed path
- Percentage of month-on-month growth rates in benchmarked series that are in the same direction as the corresponding growth rate in the true path
- Percentage of month-on-same-month-previous-year growth rates in benchmarked series that are in the same direction as the corresponding growth rate in the observed path
- Percentage of month-on-same-month-previous-year growth rates in benchmarked series that are in the same direction as the corresponding growth rate in the true path

The results of the simulations are presented in table 1.

In the results of the simulations presented in table 1, all methods produced benchmarked series which ran and met the benchmark constraints. Elementary wavelet benchmarking produces benchmarked series that, on average, have lower mean squared differences compared with both the true and observed paths than the established methods of Denton and Cholette-Dagum.

When the wavelet method is extended to include thresholding, the resulting benchmarked series have lower mean squared differences compared with the true path however they are larger compared with the observed path. This is to be expected as thresholding removes noise and smoothes the series, so movements in the observed path deemed to be noise are removed, hence the benchmarked series will not be so close to the observed path. Similarly with the growth rate comparison, wavelet thresholding has a lower percentage of growth rates with the same sign in the benchmarked series as the original or true paths (except for the month-on-same-month-previous-year growth in the true path). Where the differences in sign are seen it is typically the benchmarked series displaying zero growth because thresholding has removed noisy wavelet coefficients.

4.2 Effect of outliers and level shifts

Ideally anomalies in path series would be adjusted for before benchmarking, however in a large-scale production environment this is not always possible. This section looks at the effect on the benchmarked series of outliers and level shifts that are present in the path series but not the benchmark series, on the benchmarked series.

The 79 IoP series used for the simulations in section 4.1 are taken as the true path. The annual totals of the path series were used as the benchmark series. Two sets of tests were run; one for outliers and one for level shifts.

1. For outliers the value of the true path in January 2005 was multiplied by 0.95 to create the observed path. The observed path is equal to the true path for all months except January 2005. The observed path meets the benchmark constraint in all years except for 2005.
2. For level shifts the values of the true path were multiplied by 0.95 before May 2005 to create the observed path. The observed path is equal to the true path from May 2005 onwards. The observed path meets the benchmark constraints in years from 2006 onwards.

Metric	Denton	Cholette-Dagum	Elementary wavelets	Wavelet thresholding
Percentage of series producing a benchmarked series	100%	100%	100%	100%
Percentage of benchmarked series meeting constraints	100%	100%	100%	100%
Mean squared difference observed path	102	97	86	894
Mean squared difference true path	2143	2135	2089	1544
Percentage of month-on-month growth rates with the same sign as the observed path	99.5%	99.5%	99.7%	59.7%
Percentage of month-on-month growth rates with the same sign as the true path	76.5%	76.5%	76.5%	52.1%
Percentage of month-on-same-month-previous-year growth rates with the same sign as the observed path	96.8%	96.9%	97.1%	90.4%
Percentage of month-on-same-month-previous-year growth rates with the same sign as the true path	86.2%	86.2%	86.5%	86.6%

Table 1: Results of benchmarking simulations.

Metric	Denton	Cholette-Dagum	Elementary wavelets	Wavelet thresholding
Mean squared difference true path - outlier	0.13	0.13	0.13	1.79
Mean squared difference true path - level shift	0.13	0.16	0.24	2.00

Table 2: Results of testing effects of outliers and level shifts.

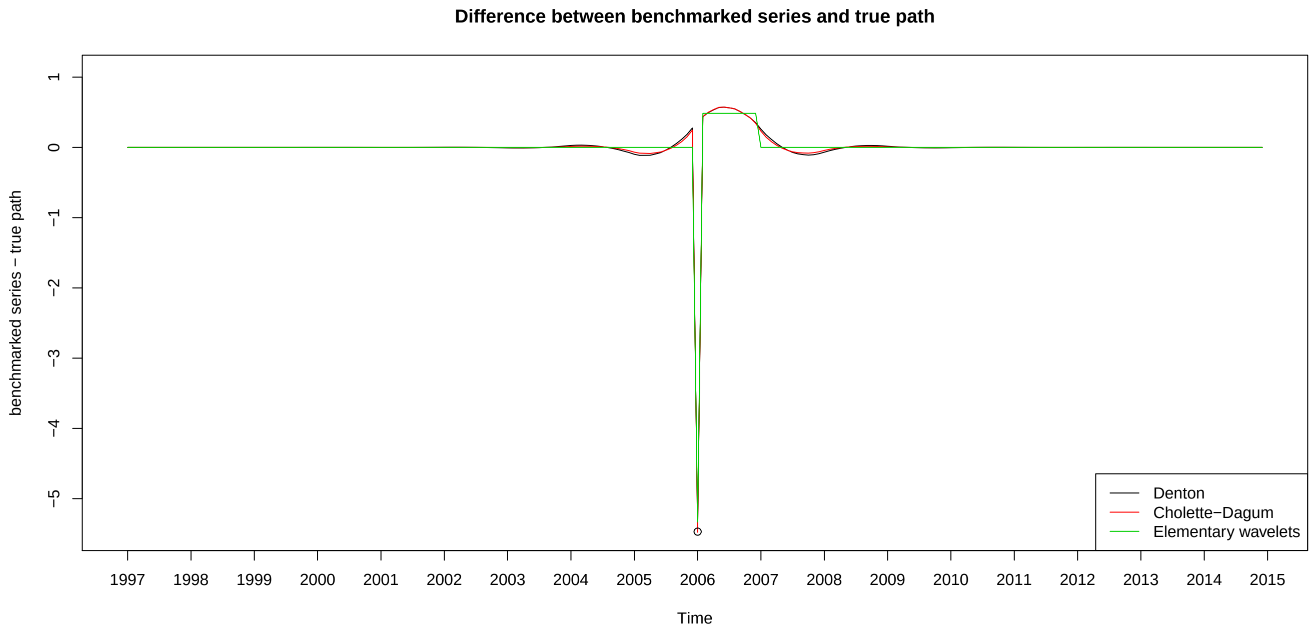

Figure 1: Example of difference between benchmarked series and true path in a series with a simulated outlier in the path but not the benchmark series.

The mean squared difference between the true path and benchmarked series was used to assess the performance of the benchmarking methods. The results can be seen in table 2.

Table 2 shows that elementary wavelets performs as well as the Denton an Cholette-Dagum methods in the presence of an outlier, however performs worst of the three for level shifts. Wavelet thresholding has the worst performance. This is to be expected as when thresholding is included in the wavelet method it smoothes the benchmarked series, hence removing noise from the path series. When compared back to the true path there will obviously be differences due to the smoothing as well as the outlier or level shift.

To understand more of the effect that outliers and level shifts are having on the benchmarking processes charts have been plotted showing the differences between the true path and the benchmarked series for one time series. Figure 1 illustrates the effect of an outlier and figure 2 illustrates the effect of a level shift. The wavelet thresholding method has not been shown because of the additional smoothing as this example is assuming no noise is present in the observed path.

The example of the outlier in figure 1 shows that all benchmarked series are affected by the outlier around the time of its appearance in the path series. With wavelets the effect is present only in the year the outlier occurs. In all other years where the benchmark constraints are already met by the observed path the benchmarked series is unchanged from the path. The effect of the outlier is largest at the time point where the outlier is present but the rest of the effect is distributed evenly in the entire year. The methods of Cholette-Dagum and Denton are affected for additional year before and after the year of the outlier, however the effect decays with distance from the time point where the outlier is present.

The example of the level shift in figure 2 shows that all series are affected prior to the level shift (where the benchmarking constraints are not met) and during the year of the level shift. The methods of Denton and Cholette-Dagum are further affected in years after the level shift however the wavelets method is not. The Cholette-Dagum method is affected at the beginning of the time series and this is likely because of the exclusion of a bias term.

This example was re-run on one series where the level shift was simulated as an additive level

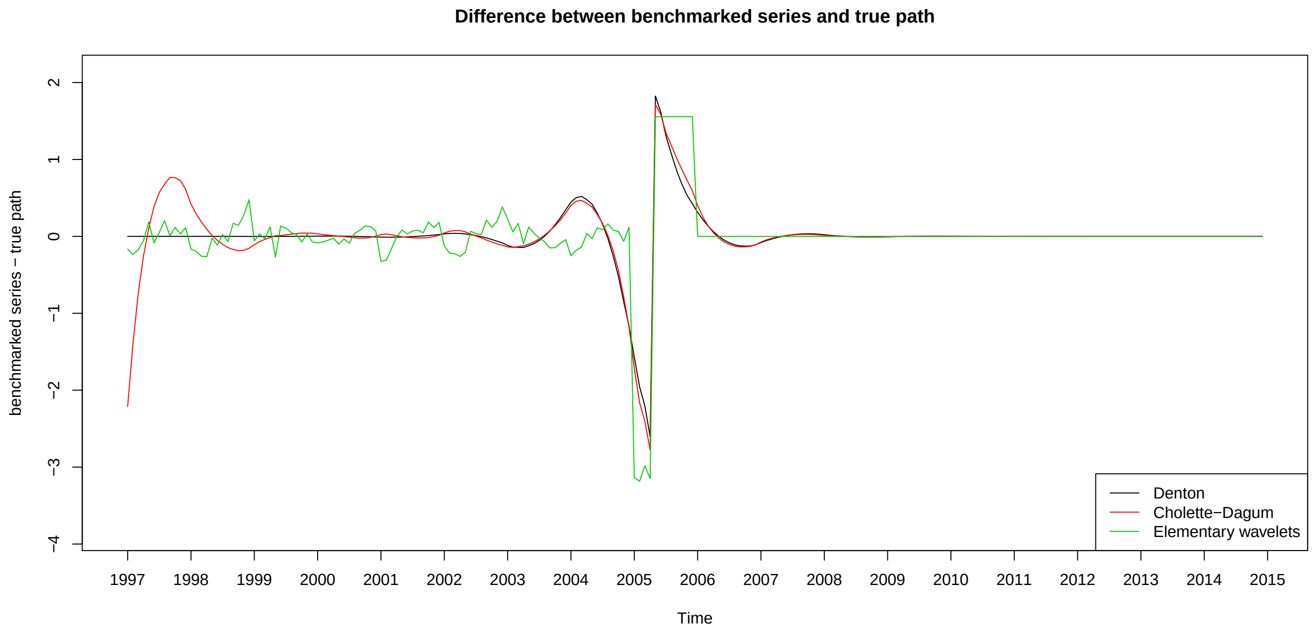

Figure 2: Example of difference between benchmarked series and true path in a series with a simulated level shift in the path but not the benchmark series.

shift where the observed path is the true path minus 40 prior to May 2005. Figure 3 shows the difference between the benchmarked series and the true path. This time the additive versions of Denton and Cholette-Dagum were used for a fair comparison.

Figure 3 shows that the wavelets method is not only affected in the year of the level shift and the effect is constant before and after the level shift. The Denton and Cholette-Dagum methods are still affected either side of the level shift and Cholette-Dagum is still affected dramatically at the beginning of the series. This illustrates the benefits of the localised approach of the wavelet method for additive level shifts.

5 Discussion

In the simulations presented in section 4.1, elementary wavelet benchmarking produces benchmarked series that have lower mean squared differences compared with both the true and observed paths than the established methods of Denton and Cholette-Dagum. When the method is extended to include thresholding, the resulting benchmarked series have lower mean squared differences compared to all other methods for the true paths. However, they are larger for the observed path. This is to be expected as thresholding removes noise and smoothes the series, so movements in the observed path deemed to be noise are removed, hence the benchmarked series will not be so close to the observed path.

A method that gets closer to the truth might be theoretically better, however, particularly with the motivation of the Denton method, there is currently importance on benchmarking preserving movements in the path series. Using wavelet benchmarking with thresholding would require an acceptance of a change in attitudes towards benchmarking in that there are errors in the growth rates in the path series.

The wavelet thresholding method performs well with outliers, only being affected in the year of the outlier and not changing any of the series where constrains are already met. This compares

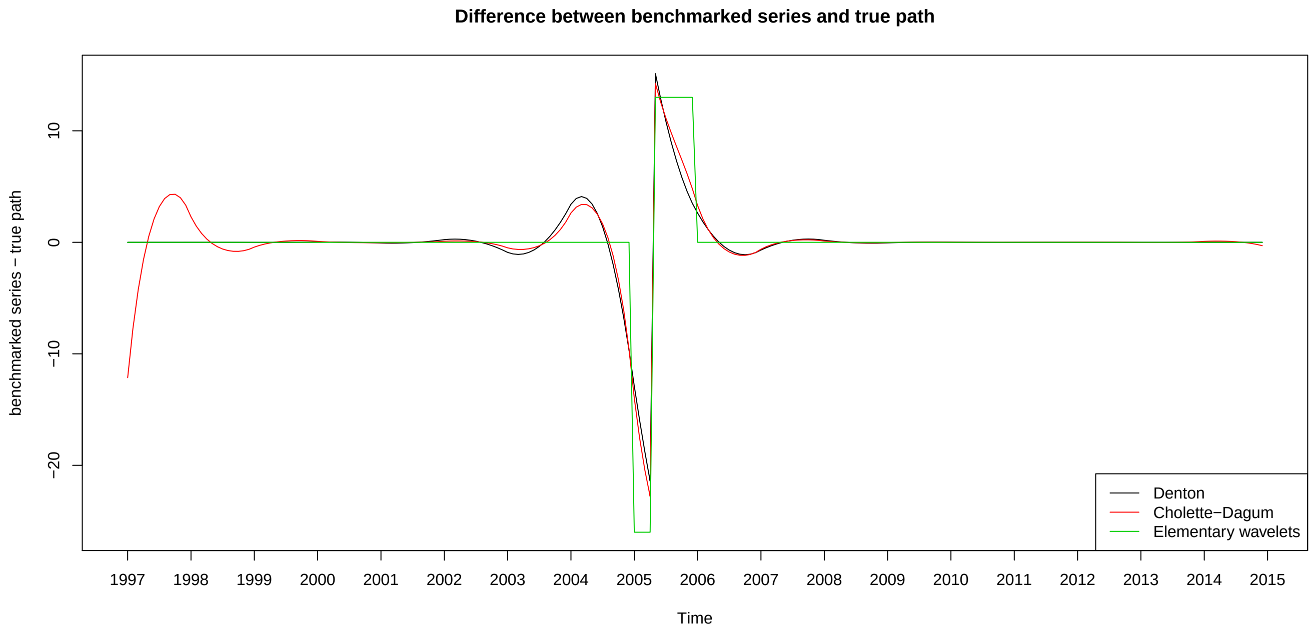

Figure 3: Example of difference between benchmarked series and true path in a series with a simulated additive level shift in the path but not the benchmark series.

with Denton and Cholette-Dagum which are affected for longer before and after the outlier and take longer to return to the true path.

In the level shift example the wavelet method is better after the level shift however prior to the level shift there are a lot of differences. In the example of an additive level shift the differences before the level shift in the benchmarked series compared with the true path were no longer present. This is to be expected as wavelets is a linear method and it deals well with additive errors. This shows that there are benefits of using wavelets in such situations. It also illustrates that there could be a need for a proportional variant of wavelet benchmarking similar to the variants of Denton and Cholette-Dagum.

6 Further work

Suggestions for further testing and development are detailed below.

- Incorporating wavelet forecasting in the benchmarking method to allow benchmarking of incomplete years, or when benchmarks are unavailable. Currently wavelet benchmarking requires the path and benchmark series to cover the same spans and to have complete years of the path series. In practice, benchmarking happens to the path series before benchmarks are available. Developing forecasting would allow the benchmarks to be forecast and the paths to be forecast to produce complete years.

- Analysing revisions to the benchmarked series as more data are added to both the path and benchmark series. Ideally a benchmarking method would minimise revisions at the current end of the time series, particularly when benchmark values become available.

- Refining the model calculating the variance for the wavelet thresholding, or adjusting the parameters for determining wavelet coefficients as noise. This would allow more control over the extent to which the benchmarked series are smoothed.

- Running the examples of outliers and level shifts with noise added on to the observed path series to allow a fairer evaluation of the wavelet thresholding method.

- Developing a method which allows a multiplicative relationship between errors in the path and benchmark series. Currently the methods of Denton and Cholette-Dagum have additive or proportional variants. The errors and anomalies simulated in this research assume a multiplicative relationship between the true path, error and the observed path. In the testing by Sayal et al. (2014) additive errors were used and more positive results were found for wavelet benchmarking. In practice in the production of official statistics, proportional relationships are often more appropriate.

Bibliography

Cholette, PA (1979). "Adjustment methods of sub-annual series to yearly benchmarks". In: *Proceeding of the Computer Science 12 th Annual Symposium on the Interface, ed. JF Gentleman*, pp. 358–360.

Cholette, Pierre A and Estela Bee Dagum (1994). "Benchmarking time series with autocorrelated survey errors". In: *International Statistical Review/Revue Internationale de Statistique*, pp. 365–377.

Denton, Frank T (1971). "Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization". In: *Journal of the American Statistical Association* 66.333, pp. 99–102.

Nason, Guy (2010). *Wavelet methods in statistics with R*. Springer Science & Business Media.

ONS (2013). *Summary Quality Report for Workforce Jobs*. URL: `http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/labour-market/summary-quality-report-for-workforce-jobs.pdf`.

Sayal, Homesh et al. (2014). "An Introduction to Applications of Wavelet Benchmarking with Seasonal Adjustment". In: *arXiv preprint arXiv:1410.7148*.

U.S. Census Bureau (2015). *X-13ARIMA-SEATS Reference Manual*. Version 1.1. URL: `https://www.census.gov/ts/x13as/docX13AS.pdf`.