SOCIETY
DATA | EVIDENCE | DECISIONS
Statistics in Society
Series A

J. R. Statist. Soc. A (2017)

An introduction to applications of wavelet benchmarking with seasonal adjustment

Homesh Sayal and John A. D. Aston,

University of Cambridge, UK

Duncan Elliott

Office for National Statistics, Newport, UK

and Hernando Ombao

University of California at Irvine, USA

[Received October 2014. Final revision July 2016]

Summary. Before adjustment, low and high frequency data sets from national accounts are frequently inconsistent. Benchmarking is the procedure used by economic agencies to make such data sets consistent. It typically involves adjusting the high frequency time series (e.g. quarterly data) so that they become consistent with the lower frequency version (e.g. annual data). Various methods have been developed to approach this problem of inconsistency between data sets. The paper introduces a new statistical procedure, namely wavelet benchmarking. Wavelet properties allow high and low frequency processes to be jointly analysed and we show that benchmarking can be formulated and approached succinctly in the wavelet domain. Furthermore the time and frequency localization properties of wavelets are ideal for handling more complicated benchmarking problems. The versatility of the procedure is demonstrated by using simulation studies where we provide evidence showing that it substantially outperforms currently used methods. Finally, we apply this novel method of wavelet benchmarking to official data from the UK's Office for National Statistics.

Keywords: Benchmarking; Seasonal adjustment; Structural time series; Thresholding; Wavelets

1. Introduction

National statistics institutes (NSIs) such as the UK's Office for National Statistics (ONS) are responsible for collecting and analysing economic data, e.g. national accounts data and labour data (Cholette and Dagum (2006), chapter 1). Data sets that are collected by such agencies are typically adjusted for a variety of reasons. Benchmarking (the focus of this paper) is an adjustment procedure that is used to make measurements from the same statistical process across different periodicities consistent. Since national accounts data must satisfy specific accounting conditions, benchmarking has important applications. It is well documented for example that unmodified quarterly gross domestic product (GDP) data are not consistent with their annual GDP version (i.e. the quarterly totals do not sum to the corresponding annual value). Since data sets of different periodicities (temporal resolutions) are often collected from different sample surveys and compiled differently, such discrepancies occur naturally as a result of survey errors.

Address for correspondance: John A. D. Aston, Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK.
E-mail: j.aston@statslab.cam.ac.uk

In many cases, for example, a larger sample is used for the less frequent survey; hence the lower frequency series is typically more reliable than its corresponding high frequency version. The aim of benchmarking is to adjust the high frequency series so that it becomes consistent with the lower frequency version while preserving short-term fluctuations. The low frequency and adjusted high frequency time series are referred to as the benchmark and benchmarked series respectively.

Benchmarking can be considered as a subclass of signal extraction problems. Current literature can be classified as providing either numerical or model-based solutions. Denton (1971) approached benchmarking by using a numerical method based on quadratic minimization. A penalty function defined by the user specifies this minimization procedure. Cholette and Dagum (1994) expressed benchmarking in terms of a stochastic regression model; hence a regression-type solution is provided. Owing to difficulties in estimating parameters, NSIs typically simplify the benchmarking method that was originally proposed by Cholette and Dagum. However, since it has a regression setting, confidence intervals can be obtained and so uncertainty about point estimates can be quantified. In practice, NSIs often implement methods which make simplifying assumptions to allow for easier estimation and greater transparency of the model.

In this paper, we present a new non-parametric methodology for benchmarking. It is based on the natural idea that the time series can be decomposed into different timescale components, and these components are subsequently used to constrain the high frequency series. Wavelets (Daubechies, 1992) provide a natural time–frequency decomposition and can adapt to local conditions in the time series. This is important in macroeconomic times series routinely analysed by the ONS. Wavelets extend the ideas of Fourier decompositions by removing the assumption of stationarity in the time series. By combining data sets from different wavelet decomposition levels, and making use of the unbalanced Haar (UH) decomposition (Fryzlewicz, 2007) to account for the non-dyadic nature of the analysis, our proposed method can reconstruct a benchmarked series with high frequency components that still satisfy the low frequency constraints.

Outliers and abrupt structural changes are commonplace in observed time series. Current methods provide global benchmarking solutions; hence volatile regions of the high frequency series have the potential to introduce artefacts in the benchmarked series. The time–frequency localization properties of wavelets (Percival and Walden (2000), page 59) provide a local solution to benchmarking and thus overcome such a problem.

In addition, NSIs frequently publish a seasonally adjusted version of the high frequency series. Seasonal adjustment is another procedure that is applied to data to remove unwanted effects (Findley, 2005), but care must be taken when combining seasonal adjustment and benchmarking. Along with adjustments for calendar effects (e.g. trading day effects) a version of benchmarking must be applied so that both the original and the seasonally adjusted high frequency series satisfy the benchmark constraint. We show that, by using a suitable seasonal model, wavelet benchmarking and seasonal adjustment can be combined within the same framework.

The paper proceeds as follows. Section 2 provides an introduction to current benchmarking methods and a short introduction to wavelets. Section 3 describes the process of benchmarking in the wavelet domain. Additional issues which require consideration such as thresholding and seasonal adjustment are also discussed. In Section 4 wavelet benchmarking is applied to a variety of simulated data and official ONS data. Section 5 concludes the paper. Details on the simulation implementations are given in Appendices A–C.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

```
http://wileyonlinelibrary.com/journal/rss-datasets
```

2. Background

The requirement of benchmarking is frequently demanded by the ONS. Currently a variety of benchmarking methods have been proposed in Denton (1971), Cholette and Dagum (1994), Durbin and Quenneville (1997), Quenneville *et al.* (2013) and Di Fonzo and Marini (2012) to name only a few. Currently the Cholette and Dagum method is the preferred method of benchmarking within the ONS (Brown *et al.*, 2012). However, this can also incorporate Denton benchmarking so we shall therefore consider these two approaches and provide a comparison of wavelet benchmarking with them.

Consider the following introductory example. A quarterly GDP time series needs to be benchmarked to an annual GDP time series; typically the annual series is less noisy than its quarterly version. To simplify the benchmarking procedure many NSIs assume that such an annual series is not contaminated with noise, and hence the high periodicity series must equate to the lower periodicity series when suitably aggregated (binding benchmarking). Throughout this paper the above example of quarterly–annual benchmarking is used to provide a concrete description; however, the methodology is applicable to general periodicity relationships. For completeness the following discussion expresses benchmarking in a general and more formal way.

Let $Y^{\mathrm{H}}_{\mathrm{T},t}$ and $Y^{\mathrm{L}}_{\mathrm{T},s}$ describe the true evolution of high (i.e. quarterly) and low (i.e. annual) frequency time series. The disturbance terms ϵ^{H}_t and ϵ^{L}_s contaminate these time series, with $Y^{\mathrm{H}}_{\mathrm{O},t}$ and $Y^{\mathrm{L}}_{\mathrm{O},s}$ denoting the observed noisy versions respectively. This is summarized as follows:

$$\begin{aligned} Y^{\mathrm{H}}_{\mathrm{O},t} &= Y^{\mathrm{H}}_{\mathrm{T},t} + \epsilon^{\mathrm{H}}_t, & t &= 1,\ldots,n, \\ Y^{\mathrm{L}}_{\mathrm{T},s} &= g_s(\{Y^{\mathrm{H}}_{\mathrm{T},t}\}^{fs}_{t=1}), & s &= 1,\ldots,m, \\ Y^{\mathrm{L}}_{\mathrm{O},s} &= Y^{\mathrm{L}}_{\mathrm{T},s} + \epsilon^{\mathrm{L}}_s, & s &= 1,\ldots,m. \end{aligned}$$

Here $g_s(\cdot)$ represents some function linking the unobserved true low frequency series to the true unobserved high frequency series. Often $g_s(\cdot)$ is a summation over a small range, $Y^{\mathrm{H}}_{\mathrm{T}} = (Y^{\mathrm{H}}_{\mathrm{T},1},\ldots,Y^{\mathrm{H}}_{\mathrm{T},n})$, $Y^{\mathrm{L}}_{\mathrm{T}} = (Y^{\mathrm{H}}_{\mathrm{L},1},\ldots,Y^{\mathrm{L}}_{\mathrm{T},m})$ and $f = n/m$ denotes the aggregation order between the two series. In the setting of quarterly to annual binding benchmarking (binding benchmarking assumes that the low frequency series is non-noisy, i.e. $\epsilon^{\mathrm{L}}_s = 0$), $Y^{\mathrm{L}}_{\mathrm{O},s} = Y^{\mathrm{L}}_{\mathrm{T},s} = \Sigma^{4s}_{t=4s-3} Y^{\mathrm{H}}_{\mathrm{T},t}$, and $f = 4$ $(n = 4m)$. Although subsequent methods rely on various statistical techniques they have a fundamental similarity in how benchmarking may be interpreted. The estimated series $\hat{Y}^{\mathrm{H}}_{\mathrm{T}}$ can typically be expressed as a linear combination of the observed high ($Y^{\mathrm{H}}_{\mathrm{O}}$) and low ($Y^{\mathrm{L}}_{\mathrm{O}}$) frequency processes. This results in the estimator

$$\hat{Y}^{\mathrm{H}}_{\mathrm{T}} = A \begin{pmatrix} Y^{\mathrm{H}}_{\mathrm{O}} \\ Y^{\mathrm{L}}_{\mathrm{O}} \end{pmatrix}, \tag{2.1}$$

where $\hat{Y}^{\mathrm{H}}_{\mathrm{T}} = (\hat{Y}^{\mathrm{H}}_{\mathrm{T},1},\ldots,\hat{Y}^{\mathrm{H}}_{\mathrm{T},n})'$, $Y^{\mathrm{H}}_{\mathrm{O}} = (Y^{\mathrm{H}}_{\mathrm{O},1},\ldots,Y^{\mathrm{H}}_{\mathrm{O},n})'$ and $Y^{\mathrm{L}}_{\mathrm{O}} = (Y^{\mathrm{L}}_{\mathrm{O},1},\ldots,Y^{\mathrm{L}}_{\mathrm{O},m})'$.

Embedded within matrix A is information describing the relationship between the high ($Y^{\mathrm{H}}_{\mathrm{O}}$) and low ($Y^{\mathrm{L}}_{\mathrm{O}}$) frequency series. Conditionally on the benchmarking procedure that is implemented, additional information summarizing statistical features, such as the time series correlation structure or estimates of model parameters, may be present. In particular for the parametric or non-parametric approach, the matrix A is respectively explicitly or implicitly data dependent.

This paper considers a particular type of benchmarking, namely binding benchmarking of flow variables; in the example of quarterly to annual benchmarking the sum of the four quarterly values must equal the corresponding value from the annual series. Other types of benchmarking exist such as ensuring that the beginnings of time period values are equal between two series. Implementation of these benchmarking methods simply requires a different specification of the benchmarking matrix A.

2.1. Denton method

Denton (1971) benchmarking, which was the first widely used benchmarking procedure, is based on the principle of movement preservation. This ensures that the benchmarked high frequency series $\hat{Y}^{\mathrm{H}}_{\mathrm{T}}$ evolves similarly to the observed series $Y^{\mathrm{H}}_{\mathrm{O}}$ (i.e. $\hat{Y}^{\mathrm{H}}_{\mathrm{T}}$ is approximately a level shift or proportionate to $Y^{\mathrm{H}}_{\mathrm{O}}$ depending on which Denton method is implemented). As described in Cholette and Dagum (2006), chapter 6, the Denton method has the following underlying model for discrete data:

$$Y^{\mathrm{H}}_{\mathrm{O},t} = Y^{\mathrm{H}}_{\mathrm{T},t} + \epsilon_t, \tag{2.2}$$

$$Y^{\mathrm{L}}_{\mathrm{O},s} = \sum_{t=p_{s,1}}^{p_{s,f}} j_{s,t} Y^{\mathrm{H}}_{\mathrm{T},t}, \tag{2.3}$$

with equation (2.3) giving the binding benchmarking constraint. In quarterly to annual binding benchmarking $j_{s,t} = 1$, with $p_{s,1}$ and $p_{s,4}$ representing the beginning and end quarters corresponding to year s respectively.

Two primary variants of Denton benchmarking are additive and proportional differencing with each best suited for additive and multiplicative time series respectively. (Although simulations generated in Section 4 have an additive form, the proportional Denton variant is also considered since it is the most commonly used version of Denton benchmarking.) Additive first differencing keeps the discrepancy between the benchmarked and original series $\hat{Y}^{\mathrm{H}}_{\mathrm{T},t} - Y^{\mathrm{H}}_{\mathrm{O},t}$ as close as possible to a constant by minimizing the following objective function (equation (2.4)) subject to the benchmark constraint (equation (2.5)) being satisfied:

$$\sum_{t=2}^{n} \{Y^{\mathrm{H}}_{\mathrm{T},t} - Y^{\mathrm{H}}_{\mathrm{O},t} - (Y^{\mathrm{H}}_{\mathrm{T},t-1} - Y^{\mathrm{H}}_{\mathrm{O},t-1})\}^2, \tag{2.4}$$

subject to

$$Y^{\mathrm{L}}_{\mathrm{O},s} = \sum_{t=p_{s,1}}^{p_{s,f}} j_{s,t} Y^{\mathrm{H}}_{\mathrm{T},t}, \qquad \forall\, s = 2, \ldots, m. \tag{2.5}$$

(Equation (2.4) expresses the Denton method in its modified form. This prevents transient spurious movements from being introduced in the benchmarked series which may occur by using the exact form (Cholette, 1984) of Denton benchmarking. Hence the modified version is used in Section 4.) The benchmarked series is approximately a vertical shift of the original series, i.e. $\hat{Y}^{\mathrm{H}}_{\mathrm{T},t} \approx Y^{\mathrm{H}}_{\mathrm{O},t} + c$, $c \in \mathbb{R}, \forall\, t$.

Denton (1971) devised the following solution based on Lagrangian optimization:

$$\begin{pmatrix} \hat{Y}^{\mathrm{H}}_{\mathrm{T}} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} A & B \\ B' & \mathbf{0}_{m\times m} \end{pmatrix}^{-1} \begin{pmatrix} A & \mathbf{0}_{n\times m} \\ B' & I_m \end{pmatrix} \begin{pmatrix} Y^{\mathrm{H}}_{\mathrm{O}} \\ Y^{\mathrm{L}}_{\mathrm{O}} - B' Y^{\mathrm{H}}_{\mathrm{O}} \end{pmatrix}. \tag{2.6}$$

In equation (2.6), $\hat{\lambda}$ corresponds to the Lagrangian multiplier. The matrices $\mathbf{0}_{m\times m}$, $\mathbf{0}_{n\times m}$ and I_m correspond to the null matrices of dimensions $m \times m$ and $n \times m$ and the identity matrix of dimension $m \times m$ respectively. Finally the matrices A and B take the form

$$B = \begin{pmatrix} \mathbf{j} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{j} & \ldots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{j} \end{pmatrix}_{n\times m}, \qquad A = D'D \quad \text{with} \quad D = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}_{n-1\times n}, \tag{2.7}$$

where $\mathbf{j}$ and $\mathbf{0}$ are $f = n/m$-dimensional column vectors taking values 1 and 0 respectively. In the aforementioned example, $\mathbf{j} = (1111)'$, B' annualizes the quarterly series and D calculates the differences between the error terms ϵ_t. Defining the matrix G as

$$G = \begin{pmatrix} A & B \\ B' & \mathbf{0}_{m\times m} \end{pmatrix}^{-1} \begin{pmatrix} A & \mathbf{0}_{n\times m} \\ B' & I_m \end{pmatrix} \tag{2.8}$$

enables equation (2.6) to be expressed in the form

$$\begin{pmatrix} \hat{Y}_{\mathrm{T}}^{\mathrm{H}} \\ \hat{\lambda} \end{pmatrix} = G \begin{pmatrix} Y_{\mathrm{O}}^{\mathrm{H}} \\ Y_{\mathrm{O}}^{\mathrm{L}} - B' Y_{\mathrm{O}}^{\mathrm{H}} \end{pmatrix}. \tag{2.9}$$

Matrix G can be decomposed into the following constituent components (Cholette and Dagum, 2006):

$$G = \begin{pmatrix} I_n & G_{Y_{\mathrm{T}}^{\mathrm{H}}} \\ \mathbf{0}_{m\times n} & G_{\lambda} \end{pmatrix}, \tag{2.10}$$

where I_n and $\mathbf{0}_{m\times n}$ correspond to the $n \times n$ identity matrix and $m \times n$ null matrix respectively. $G_{Y_{\mathrm{T}}^{\mathrm{H}}}$ and G_{λ} affect only the estimation of the unobserved high frequency and Lagrangian multiplier components respectively. As a result of equation (2.10), the following equation expresses the benchmarking solution in the form of equation (2.1):

$$\hat{Y}_{\mathrm{T}}^{\mathrm{H}} = (I - G_{Y_{\mathrm{T}}^{\mathrm{H}}} B' \quad G_{Y_{\mathrm{T}}^{\mathrm{H}}}) \begin{pmatrix} Y_{\mathrm{O}}^{\mathrm{H}} \\ Y_{\mathrm{O}}^{\mathrm{L}} \end{pmatrix}. \tag{2.11}$$

It is important to note that working with matrix D from equation (2.7) means that the modified Denton benchmarking solution as described in Cholette and Dagum (2006) is considered, which avoids the initial condition of the original Denton method and often results in a more satisfactory solution.

It is possible to specify equation (2.4) in terms of higher order additive differences between the original and adjusted series. For example $\Sigma_{i=h+1}^{n}(\Delta^h \hat{Y}_{\mathrm{T},t}^{\mathrm{H}} - \Delta^h Y_{\mathrm{O},t}^{\mathrm{H}})^2$ corresponds to the hth-order additive model with Δ^h being the hth difference operator and values outside the adjustment range being defined as $Y_{\mathrm{O},t}^{\mathrm{H}} = Y_{\mathrm{T},t}^{\mathrm{H}}, t = 0, -1, \ldots, 1-h$. Section 4 implements additive Denton benchmarking with values of $h = 1$ (we also considered $h = 2$ but the results were not useful in practice so $h = 2$ was not further considered) and the proportional Denton variant (matrix D is replaced by $D \times \mathrm{diag}(Y_{\mathrm{O}}^{\mathrm{H}})$) with $h = 1$. Hereafter such benchmarked series are referred to as the $\mathrm{Denton}_{\mathrm{a},1}$ and $\mathrm{Denton}_{\mathrm{p},1}$ series.

Although it is not computationally demanding and requires only basic assumptions on the structural form of the time series being analysed, the Denton method occasionally performs poorly. This is evident in time series which evolve unconventionally; for example consider a time series containing a small number of extreme data points. In this case, the Denton method would adjust a disproportionate number of data points. As mentioned in Section 1, one motivation for considering wavelets is that their time–frequency localization properties can help to overcome this problem.

2.2. Cholette and Dagum method

The Cholette and Dagum (1994) benchmarking method is based on stochastic regression. However, its original structure is typically simplified in NSI applications. As such, the description that was provided in Quenneville *et al.* (2003) is discussed in what follows. For completeness the methodology that was proposed by Cholette and Dagum can be found in Appendix B.

The regression benchmarking model consists of the two equations

$$Y_{\mathrm{O}}^{\mathrm{H}} = Y_{\mathrm{T}}^{\mathrm{H}} + C\epsilon_{\mathrm{H}}, \qquad \mathbb{E}(\epsilon_{\mathrm{H}}) = 0, \quad \mathrm{cov}(\epsilon_{\mathrm{H}}) = \sigma_{\epsilon_{\mathrm{H}}}^2 \Omega_{\epsilon_{\mathrm{H}}}, \tag{2.12}$$

$$Y_{\mathrm{O}}^{\mathrm{L}} = J Y_{\mathrm{T}}^{\mathrm{H}}, \tag{2.13}$$

with equation (2.13) giving the benchmarking constraint. In equation (2.12) the observed noisy high frequency time series is a linear combination of the true unobserved low frequency time series and a distortion term $C\epsilon_{\mathrm{H}}$. Here ϵ_{H} is a zero-mean Gaussian process with auto-correlation matrix $\sigma_{\epsilon_{\mathrm{H}}}^2 \Omega_{\epsilon_{\mathrm{H}}}$. $\Omega_{\epsilon_{\mathrm{H}}}$ is the autocovariance matrix of an auto-regressive AR(1) process with parameter ρ, and $\sigma_{\epsilon_{\mathrm{H}}}^2$ is a nuisance parameter which does not require estimation. The matrix C is an $n \times n$ matrix with weights $c_t \propto |Y_{\mathrm{O},t}^{\mathrm{H}}|^{\lambda}$ on the main diagonal and 0 elsewhere. Typically the parameter value for λ is set to 0, $\frac{1}{2}$ or 1. The Cholette and Dagum approach used by the ONS sets $\lambda = 0$, resulting in the further simplification that $C = I$. In equation (2.13) J is an annualizing matrix equivalent to matrix B' from the Denton method. Equation (2.13) has the interpretation that the sum of the subannual time periods from the high frequency time series $Y_{\mathrm{T}}^{\mathrm{H}}$ must equal the corresponding annual value from the low frequency time series $Y_{\mathrm{O}}^{\mathrm{L}}$ (this enforces the benchmarking constraint).

When we assume that the benchmark solution is characterized by equations (2.12)–(2.13), we obtain the following equations:

$$\hat{Y}_{\mathrm{T}}^{\mathrm{H}} = Y_{\mathrm{O}}^{\mathrm{H}} + C\Omega_{\epsilon_{\mathrm{H}}} C J' (J C \Omega_{\epsilon_{\mathrm{H}}} C J')^{-1} (Y_{\mathrm{O}}^{\mathrm{L}} - J Y_{\mathrm{O}}^{\mathrm{H}}), \tag{2.14}$$

$$\hat{Y}_{\mathrm{T}}^{\mathrm{H}} = Y_{\mathrm{O}}^{\mathrm{H}} + \Omega_{\epsilon_{\mathrm{H}}} J' (J \Omega_{\epsilon_{\mathrm{H}}} J')^{-1} (Y_{\mathrm{O}}^{\mathrm{L}} - J Y_{\mathrm{O}}^{\mathrm{H}}), \qquad \text{when } C = I, \tag{2.15}$$

$$\hat{Y}_{\mathrm{T}}^{\mathrm{H}} = Y_{\mathrm{O}}^{\mathrm{H}} + K (Y_{\mathrm{O}}^{\mathrm{L}} - J Y_{\mathrm{O}}^{\mathrm{H}}), \qquad K = \Omega_{\epsilon_{\mathrm{H}}} J' (J \Omega_{\epsilon_{\mathrm{H}}} J')^{-1}. \tag{2.16}$$

Equation (2.16) has the interpretation that the estimated high frequency time series $\hat{Y}_{\mathrm{T}}^{\mathrm{H}}$ is a linear combination of the observed noisy high frequency time series $Y_{\mathrm{O}}^{\mathrm{H}}$ and a scaled difference of the observed non-noisy low frequency time series and annualized version of the noisy observed high frequency time series. The matrix K determines the scaling effect. This solution can be expressed in a form that is consistent with equation (2.1) as follows:

$$\hat{Y}_{\mathrm{T}}^{\mathrm{H}} = (I - K \quad K) \begin{pmatrix} Y_{\mathrm{O}}^{\mathrm{H}} \\ Y_{\mathrm{O}}^{\mathrm{L}} \end{pmatrix}. \tag{2.17}$$

As discussed in Appendix B, the ONS uses values of ρ 0.8 and 0.8^3 for monthly and quarterly to annual benchmarking respectively.

2.3. Wavelets

Stationarity underpins many time series methods; this assumption is often unreasonable. Wavelets' time or frequency localization enables segmentation of data over various frequency or time levels, thus providing a framework to analyse high (i.e. quarterly data) and low (i.e. annual data) frequency series jointly. Intuitively, wavelets can be seen to perform frequency analysis over localized time segments, producing a joint time–frequency analysis, in a related (but not identical) vein to windowed Fourier analysis. A more formal definition will be given in the next section.

Although wavelets have facilitated recent advances in time series, i.e. alternative modelling of non-stationary processes (Nason *et al.*, 2000), their primary use lies in non-parametric regression and involves removing noise from a statistical process in a non-parametric setting (Donoho and

Johnstone, 1994). Subsequent sections show that the combination of a strict benchmarking and thresholding (denoising) step produces a benchmarking procedure which can outperform those currently used.

2.3.1. *Unbalanced Haar wavelets*

The remainder of this section discusses UH wavelets (Fryzlewicz, 2007). Data sets observed are typically non-dyadic in length (i.e. $n \neq 2^J$, $J \in \mathbb{N}$). UH wavelets are a generalization of Haar wavelets (Daubechies, 1992) and enable the transformation of such non-dyadic data sets into the wavelet domain. Whereas discontinuities in Haar basis functions occur in the middle of their support (Fig. 1), UH basis functions have discontinuities at arbitrary locations (Fig. 2). Consequently high and low frequency data sets with arbitrary lengths or factor differences can be jointly analysed.

Consider the temporal support set $\{1, \ldots, n\}$. The elementary father wavelet $\varphi^{-1,1}(t)$ is defined as

$$\varphi^{-1,1}(t) = \frac{1}{\sqrt{n}} \mathbb{1}(1 \leqslant t \leqslant n). \tag{2.18}$$

Let $s^{j,k} < b^{j,k} < e^{j,k}$ denote the start point, break point and end point of a mother wavelet at scale level j and translation level k. The mother wavelet $\varphi_{s^{j,k},b^{j,k},e^{j,k}}(t)$ is defined as follows (see Fig. 2):

$$\begin{aligned}\varphi_{s^{j,k},b^{j,k},e^{j,k}}(t) = &\left(\frac{1}{b^{j,k} - s^{j,k} + 1} - \frac{1}{e^{j,k} - s^{j,k} + 1}\right)^{1/2} \mathbb{1}(s^{j,k} \leqslant t \leqslant b^{j,k}) \\ &- \left(\frac{1}{e^{j,k} - b^{j,k}} - \frac{1}{e^{j,k} - s^{j,k} + 1}\right)^{1/2} \mathbb{1}(b^{j,k} + 1 \leqslant t \leqslant e^{j,k}).\end{aligned} \tag{2.19}$$

Given $\varphi^{j,k}(t) := \varphi_{s^{j,k},b^{j,k},e^{j,k}}(t)$, its two daughter wavelets $\varphi^{j+1,2k-1}(t)$ and $\varphi^{j+1,2k}(t)$ (mother wavelets existing on higher frequency levels) with arbitrary break points $b^{j+1,2k-1}$ (where $s^{j,k} < b^{j+1,2k-1} < b^{j,k}$) and $b^{j+1,2k}$ (where $b^{j,k} < b^{j+1,2k} < e^{j,k}$) are obtained as follows:

$$\varphi^{j+1,2k-1}(t) = \varphi_{s^{j,k},b^{j+1,2k-1},b^{j,k}}(t), \tag{2.20}$$

$$\varphi^{j+1,2k}(t) = \varphi_{b^{j,k},b^{j+1,2k},e^{j,k}}(t). \tag{2.21}$$

This recursive process continues until an orthonormal wavelet basis is formed. Selecting appropriate break points $\{b^{0,1}, b^{1,1}, b^{1,2}, \ldots\}$ to ensure that benchmarking can be performed is discussed in subsequent sections. In the setting of Haar wavelets ($n = 2^J$), for a particular frequency level j it has a dyadic number of translation levels, i.e. $k = 1, \ldots, 2^j$. Furthermore break points also occur at dyadic intervals; the first break point $b^{0,1} = n/2$ and those at higher frequency levels are given by $b^{j+1,2k-1} = b^{j,k}/2$ and $b^{j+1,2k} = b^{j,k} + b^{j,k}/2$. At all frequency levels the first start point $s^{j,1} = 1$ and the last end point $e^{j,2^j} = n$.

For a given set of start points $\{s^{0,1}, s^{1,1}, s^{1,2}, \ldots\}$, break points $\{b^{0,1}, b^{1,1}, b^{1,2}, \ldots\}$ and end points $\{e^{0,1}, e^{1,1}, e^{1,2}, \ldots\}$ the discrete UH transform of a series $Y = \{Y_t\}_{t=1}^n$ is defined as

$$w(j,k) := \langle Y, \varphi^{j,k} \rangle = \sum_{t=1}^{n} Y_t \varphi^{j,k}(t), \qquad j = -1, \ldots, J, \quad k = 1, \ldots, k_j. \tag{2.22}$$

To shorten the notation $w(j,k)$'s and $\varphi^{j,k}$'s dependence on $\{s^{j,k}, b^{j,k}, e^{j,k}\}$ is implicit. In particular $w(-1,1) = (1/\sqrt{n})\Sigma_{t=1}^n Y_t$ denotes the elementary father wavelet coefficient; it summarizes

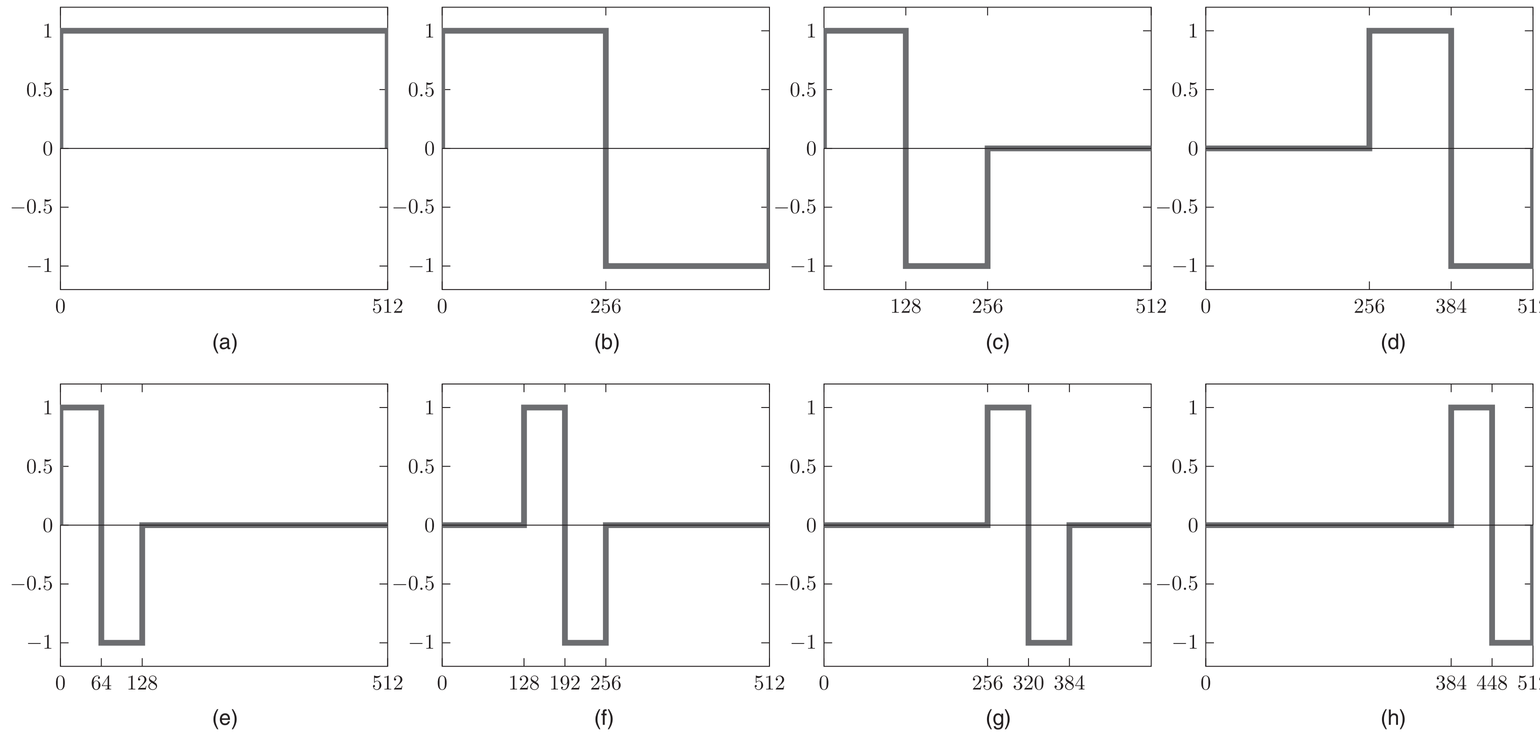

Fig. 1. Haar wavelet segmentation for a time series of length $t = 512$ for frequency scales $-1, 0, 1$ and 2 (for illustration the vertical axes have been rescaled to 1 and -1; the exact scales can be determined by equations (2.18) and (2.20)): (a) $\varphi^{-1,1}(\cdot)$; (b) $\varphi^{0,1}(\cdot)$; (c) $\varphi^{1,1}(\cdot)$; (d) $\varphi^{1,2}(\cdot)$; (e) $\varphi^{2,1}(\cdot)$; (f) $\varphi^{2,2}(\cdot)$; (g) $\varphi^{2,3}(\cdot)$; (h) $\varphi^{2,4}(\cdot)$

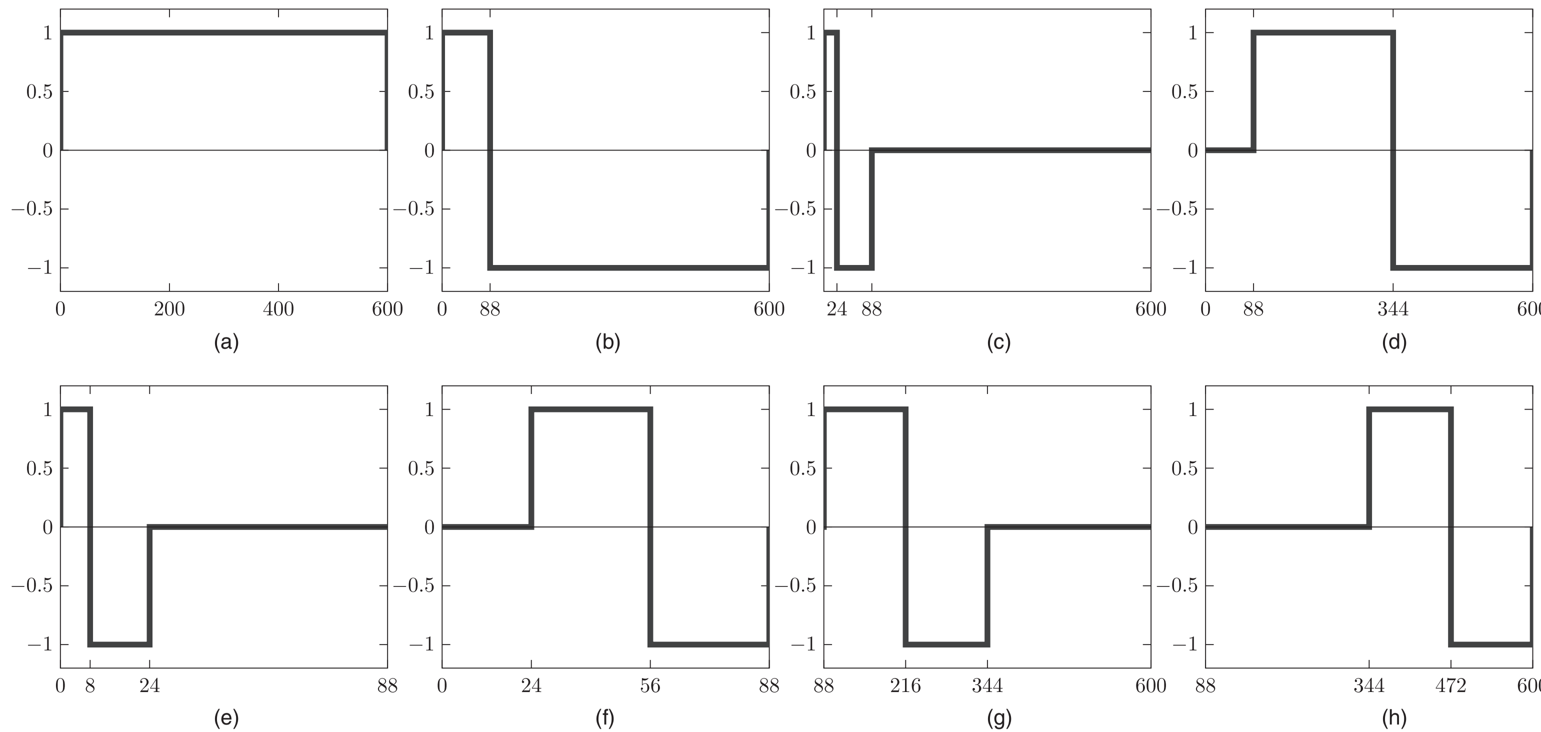

Fig. 2. Example of a UH wavelet segmentation for a time series of length $t = 600$ for frequency scales $-1, 0, 1$ and 2 (for illustration the vertical axes have been rescaled to 1 and -1; the exact scales can be determined by equations (2.18) and (2.20)): (a) $\varphi^{-1,1}(\cdot)$; (b) $\varphi^{0,1}(\cdot)$; (c) $\varphi^{1,1}(\cdot)$; (d) $\varphi^{1,2}(\cdot)$; (e) $\varphi^{2,1}(\cdot)$; (f) $\varphi^{2,2}(\cdot)$; (g) $\varphi^{2,3}(\cdot)$; (h) $\varphi^{2,4}(\cdot)$

the average behaviour of the time series. Mother wavelet coefficients $w(j,k)$, $j \geqslant 0$, provide information describing local features. Larger values of j make the region of the time series considered narrower whereas k determines the position considered on the timescale. The following equation resynthesizes the original series Y from the set of wavelet coefficients:

$$Y_t = w(-1,1)\varphi^{-1,1}(t) + \sum_{j=0}^{J} \sum_{k=1}^{k_j} w(j,k)\varphi^{j,k}(t), \qquad t = 1,\ldots,n. \tag{2.23}$$

Equation (2.23) expresses Y as a weighted linear combination of the elementary father wavelet and mother wavelets across various frequency and translation levels. Weights are given by their corresponding wavelet coefficients. This allows reconstruction of the benchmarked series after wavelet analysis.

3. Methodology

In this section we discuss the selection of wavelet bases that are used to facilitate benchmarking. Elementary wavelet benchmarking is introduced along with an application to simulated data. Finally the additional issue of thresholding and its integration with seasonal adjustment is considered.

3.1. Wavelet basis selection

3.1.1. Forming a basis for non-dyadic data by using unbalanced Haar wavelets

This paper segments data by using a basis that is similar to the traditional Haar basis. Only a limited number of UH wavelets are used to construct such a basis. The formation of such a basis is outlined as follows. At each iteration the positive region of the mother wavelet being considered is segmented into a daughter wavelet with the largest possible dyadic region and non-dyadic positive region. Its negative region is segmented into a daughter wavelet with positive and negative regions of equal length (Haar segmentation).

More formally, consider the support of $\varphi^{j,k}$ along with the support of its positive and negative regions. Denote their cardinality by $n_{j,k}$, $n^+_{j,k}$ and $n^-_{j,k}$ respectively. For the father wavelet $\varphi^{-1,1}$, $|\text{supp}(\varphi^{-1,1})| = n_{-1,1} = n^+_{-1,1}$ ($\text{supp}(f) := \{x; f(x) \neq 0\}$), with $n_{-1,1}$ being the length of the time series. $\varphi^{-1,1}$ is decomposed, forming the mother wavelet $\varphi^{0,1}$, with $|\text{supp}(\varphi^{0,1})| = n_{0,1} = n_{-1,1}$. Setting $n^-_{0,1} = 2^{\lfloor \log_2(n_{0,1}) \rfloor}$ ($\lfloor\cdot\rfloor$ denotes the greatest integer function) ensures that the negative region of $\varphi^{0,1}$ has the largest possible dyadic support. Consequently $n^+_{0,1} = n_{0,1} - n^-_{0,1}$; typically $n^+_{0,1}$ is non-dyadic in length. Its corresponding region is segmented in a similar manner to $\varphi^{0,1}$ whereas regions of dyadic support (regions of $\varphi^{0,1}$ corresponding to $n^-_{0,1}$) are segmented by using the Haar transform. This iterative process continues until a basis is formed.

Fig. 2 illustrates an example of the above segmentation using UH wavelets with the frequency levels $-1, 0, 1$ and 2 considered. Table 1 records the support of these wavelets. To provide a comparison Fig. 1 illustrates the Haar segmentation on the same frequency levels.

3.1.2. Creating a benchmarking basis

The set of break points $\{b^{0,1}, b^{1,1}, b^{1,2}, \ldots\}$ determines the UH wavelet basis. Benchmarking requires the bases for low and high frequency processes to be comparable.

The low ($\{Y^L_t\}^m_{t=1}$) and high ($\{Y^H_t\}^n_{t=1}$) frequency series are observed, with $n = fm$ and f being the factor difference. Let $L_{\text{BP}} = \{b^{0,1}_{\text{L}}, b^{1,1}_{\text{L}}, b^{1,2}_{\text{L}}, \ldots, b^{J_{\text{L}},k_{J_{\text{L}}}}_{\text{L}}\}$ and $H_{\text{BP}} = \{b^{0,1}_{\text{H}}, \ldots, b^{J_{\text{L}},k_{J_{\text{L}}}}_{\text{H}}, b^{J_{\text{L}}+1,1}_{\text{H}}, \ldots b^{J_{\text{H}},k_{J_{\text{H}}}}_{\text{H}}\}$ represent the low and high frequency series set of break points respectively.

Table 1. Support of vectors used to perform the wavelet transform for non-dyadic data of length 600

UH wavelet length	*Value*
$n_{0,1}$	600
$n_{1,1}$	600
$n_{1,1}^{+}$	88
$n_{1,1}^{-}$	512
$n_{2,1}$	88
$n_{2,1}^{+}$	24
$n_{2,1}^{-}$	64
$n_{2,2}$	512
$n_{2,2}^{+}$	256
$n_{2,2}^{-}$	256
$n_{3,1}$	24
$n_{3,1}^{+}$	8
$n_{3,1}^{-}$	16
$n_{3,2}$	64
$n_{3,2}^{+}$	32
$n_{3,2}^{-}$	32
$n_{3,3}=n_{3,4}$	256
$n_{3,3}^{+}=n_{3,4}^{+}$	128
$n_{3,3}^{-}=n_{3,4}^{-}$	128

The set of break points L_{BP} is selected by the method that was described in Section 3.1.1. Break points for H_{BP} with overlapping frequency levels with L_{BP} are defined as

$$b_{\mathrm{H}}^{j,k}=fb_{\mathrm{L}}^{j,k}, \qquad j=1,\ldots,J_{\mathrm{L}}, \quad f=1,\ldots,k_j. \tag{3.1}$$

Remaining break points $\{b_{\mathrm{H}}^{J_{\mathrm{L}}+1,1},\ldots,b_{\mathrm{H}}^{J_{\mathrm{H}},k_{J_{\mathrm{H}}}}\}$ can be chosen arbitrarily as they exist on frequency levels that are not affected by elementary benchmarking. To maintain consistency the procedure in Section 3.1.1 is used. The sets L_{BP} and H_{BP} provide the foundation that is required to perform elementary wavelet benchmarking.

3.2. Elementary wavelet benchmarking

Consider quarterly to annual GDP binding benchmarking. The quarterly $\{Y_{\mathrm{O},t}^{\mathrm{Q}}\}_{t=1}^{n}$ and annual $\{Y_{\mathrm{O},t}^{\mathrm{A}}\}_{t=1}^{m}$ GDP series are observed ($m=n/4$). Both series are expressed in the form described by equation (2.23):

$$Y_{\mathrm{O},t}^{\mathrm{Q}}=w_{\mathrm{O}}^{\mathrm{Q}}(-1,0)\varphi_{-1,0}^{\mathrm{Q}}(t)+\sum_{j=1}^{J+2}\sum_{k=0}^{k_j}w_{\mathrm{O}}^{\mathrm{Q}}(j,k)\varphi_{j,k}^{\mathrm{Q}}(t), \qquad t=1,\ldots,n,$$

$$Y_{\mathrm{O},t}^{\mathrm{A}}=w_{\mathrm{T}}^{\mathrm{A}}(-1,0)\varphi_{-1,0}^{\mathrm{A}}(t)+\sum_{j=1}^{J}\sum_{k=0}^{k_j}w_{\mathrm{T}}^{\mathrm{A}}(j,k)\varphi_{j,k}^{\mathrm{A}}(t), \qquad t=1,\ldots,m.$$

The construction of wavelet functions $\varphi_{j,k}^{\mathrm{Q}}(\cdot)$ and $\varphi_{j,k}^{\mathrm{A}}(\cdot)$ defined on the sets $\{1,\ldots,n\}$ and $\{1,\ldots,m\}$ respectively is discussed in Section 2.3.1. $w_{\mathrm{O}}^{\mathrm{Q}}(-1,0)$, $w_{\mathrm{O}}^{\mathrm{Q}}(j,k)$ and $w_{\mathrm{T}}^{\mathrm{A}}(-1,0)$,

$w_{\mathrm{T}}^{\mathrm{A}}(j,k)$ denote noisy and non-noisy wavelet coefficients from the quarterly and annual time series respectively. J is the highest frequency level of the annual time series; the quarterly time series has two additional frequency levels.

Quarterly wavelet coefficients existing on lower frequency levels of the wavelet domain have corresponding annual wavelet coefficients with similar interpretations, i.e. coefficients existing on frequency levels $j=-1,\ldots,J$. A comparison of elementary quarterly and annual father wavelet coefficients illustrates this:

$$w_{\mathrm{O}}^{\mathrm{Q}}(-1,1)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}Y_{\mathrm{O},t}^{\mathrm{Q}}, \tag{3.2}$$

$$\begin{aligned} w_{\mathrm{T}}^{\mathrm{A}}(-1,1)&=\frac{1}{\sqrt{m}}\sum_{t=1}^{m}Y_{\mathrm{O},t}^{\mathrm{A}}=\frac{1}{\sqrt{m}}\sum_{t=1}^{m}Y_{\mathrm{T},t}^{\mathrm{A}} \qquad \text{(since the annual GDP series is non-noisy)},\\ &=\frac{1}{\sqrt{m}}\sum_{t=1}^{m}\sum_{j=4t-3}^{4t}Y_{\mathrm{T},j}^{\mathrm{Q}}=\frac{1}{\sqrt{m}}\sum_{t=1}^{n}Y_{\mathrm{T},t}^{\mathrm{Q}}=\frac{2}{\sqrt{n}}\sum_{t=1}^{n}Y_{\mathrm{T},t}^{\mathrm{Q}}=2w_{\mathrm{T}}^{\mathrm{Q}}(-1,1) \qquad \text{(since } n=4m\text{)}. \end{aligned} \tag{3.3}$$

This illustrates the key idea of elementary wavelet benchmarking; replacing $w_{\mathrm{O}}^{\mathrm{Q}}(j,k)$ with $\frac{1}{2}w_{\mathrm{T}}^{\mathrm{A}}(j,k)$ for wavelet coefficients on frequency levels $j=-1,\ldots,J$ produces the benchmarked series $\{\hat{Y}_{\mathrm{T},t}^{\mathrm{Q}}\}_{t=1}^{n}$:

$$\hat{Y}_{\mathrm{T},t}^{\mathrm{Q}}=\underbrace{\frac{w_{\mathrm{T}}^{\mathrm{A}}(-1,0)}{2}\varphi_{-1,0}^{\mathrm{Q}}(t)+\sum_{j=0}^{J}\sum_{k=0}^{k_j}\frac{w_{\mathrm{T}}^{\mathrm{A}}(j,k)}{2}\varphi_{j,k}^{\mathrm{Q}}(t)}_{Y_{\mathrm{T},t}^{\mathrm{A,Q}}}+\underbrace{\sum_{j=J+1}^{J+2}\sum_{k=0}^{k_j}w_{\mathrm{O}}^{\mathrm{Q}}(j,k)\varphi_{j,k}(t)}_{R_{Y_{\mathrm{O},t}^{\mathrm{Q}}}^{J+1,J+2}}. \tag{3.4}$$

Equation (3.4) decomposes the benchmarked series into the two components $Y_{\mathrm{T},t}^{\mathrm{A,Q}}$ and

$$R_{Y_{\mathrm{O},t}^{\mathrm{Q}}}^{J+1,J+2}.$$

$Y_{\mathrm{T},t}^{\mathrm{A,Q}}$ expresses the non-noisy annual series on a quarterly timescale with no intra-annual fluctuations (i.e. quarterly values in a given year take the same value). The second component isolates fluctuations that are unique to the quarterly time series. Since it exists on the frequency levels $J+1$ and $J+2$ it has no effect on the annualized version of $\hat{Y}_{T,t}^{\mathrm{Q}}$. Therefore the benchmark constraint is satisfied, i.e.

$$\hat{Y}_{\mathrm{T},t}^{\mathrm{A}}=\sum_{j=4t-3}^{4t}\hat{Y}_{\mathrm{T},t}^{\mathrm{Q}}=Y_{\mathrm{T},t}^{\mathrm{A}}.$$

Thresholding $\{R_{Y_{\mathrm{O},t}^{\mathrm{Q}}}^{J+1,J+2}\}_{t=1}^{n}$ can further improve the estimation of $\{\hat{Y}_{\mathrm{T}}^{\mathrm{Q}}\}_{t=1}^{n}$, as discussed in later sections.

Elementary wavelet benchmarking is expressed in a form that is consistent with equation (2.1) as follows. Although computationally inefficient, the wavelet transform can be expressed as an orthogonal matrix; see Nason (2008), chapter 2, for details. Suppose that W^{Q} and W^{A} transform the quarterly and annual series into the wavelet domain respectively:

$$\begin{aligned} y_{\mathrm{O}}^{\mathrm{Q}}&=W^{\mathrm{Q}}Y_{\mathrm{O}}^{\mathrm{Q}},\\ y_{\mathrm{T}}^{\mathrm{A}}&=W^{\mathrm{A}}Y_{\mathrm{O}}^{\mathrm{A}}. \end{aligned} \tag{3.5}$$

Decomposing W^{Q} into low ($W^{\mathrm{Q,A}}$) and high ($W^{\mathrm{Q,Q}}$) frequency components enables the low ($y_{\mathrm{O}}^{\mathrm{Q,A}}$) and high ($y_{\mathrm{O}}^{\mathrm{Q,Q}}$) frequency wavelet coefficients to be obtained:

$$y_{\mathrm{O}}^{\mathrm{Q}}=\begin{pmatrix} y_{\mathrm{O}}^{\mathrm{Q,A}} \\ y_{\mathrm{O}}^{\mathrm{Q,Q}} \end{pmatrix}=\begin{pmatrix} W^{\mathrm{Q,A}} \\ W^{\mathrm{Q,Q}} \end{pmatrix} Y_{\mathrm{O}}^{\mathrm{Q}}.$$

The non-noisy annual wavelet coefficients $y_{\mathrm{T}}^{\mathrm{A}}$ can be incorporated in the noisy quarterly wavelet coefficients $y_{\mathrm{O}}^{\mathrm{Q}}$ as follows:

$$\hat{y}_{\mathrm{T}}^{\mathrm{Q}}=\begin{pmatrix} \frac{1}{c} y_{\mathrm{T}}^{\mathrm{A}} \\ y_{\mathrm{O}}^{\mathrm{Q,Q}} \end{pmatrix}=\begin{pmatrix} 0 & 0 & \frac{1}{c} I \\ 0 & I & 0 \end{pmatrix}\begin{pmatrix} y_{\mathrm{O}}^{\mathrm{Q,A}} \\ y_{\mathrm{O}}^{\mathrm{Q,Q}} \\ y_{\mathrm{T}}^{\mathrm{A}} \end{pmatrix}.$$

As seen from equation (3.3), in the example of quarterly to annual benchmarking $c=2$. The benchmarked series $\{\hat{Y}_{\mathrm{T},t}^{\mathrm{H}}\}_{t=1}^{n}$ is then calculated as follows:

$$\hat{Y}_{\mathrm{T}}^{\mathrm{H}}=(W^{\mathrm{Q}})^{\mathrm{T}}\hat{y}_{\mathrm{T}}^{\mathrm{Q}}=(W^{\mathrm{Q}})^{\mathrm{T}}\begin{pmatrix} \frac{1}{c} y^{\mathrm{A}} \\ \tilde{y}^{\mathrm{Q,Q}} \end{pmatrix}=\underbrace{(W^{\mathrm{Q}})^{\mathrm{T}}\begin{pmatrix} 0 & 0 & \frac{1}{c} I \\ 0 & I & 0 \end{pmatrix}\begin{pmatrix} W^{Q} & 0 \\ 0 & W^{\mathrm{A}} \end{pmatrix}}_{\text{elementary wavelet benchmarking matrix}}\begin{pmatrix} Y_{\mathrm{O}}^{\mathrm{Q}} \\ Y_{\mathrm{O}}^{\mathrm{A}} \end{pmatrix}.$$

3.3. Thresholding

In the previous example the benchmarked series $\{\hat{Y}_{\mathrm{T},t}^{\mathrm{Q}}\}_{t=1}^{n}$ was decomposed into a low frequency non-noisy component $\{Y_{\mathrm{T},t}^{\mathrm{A,Q}}\}_{t=1}^{n}$ and a high frequency noisy component

$$\{R_{Y_{\mathrm{O},t}^{\mathrm{Q}}}^{J+1,J+2}\}_{t=1}^{n}.$$

Thresholding wavelet coefficients corresponding to $\{R_{Y_{\mathrm{O},t}^{\mathrm{Q}}}^{J+1,J+2}\}_{t=1}^{n}$ produces a more reliable benchmarked series, as it removes spurious noise.

Technical details of thresholding are available in Vidakovic (1999), chapter 6; however, two features of the error term are important. Firstly its structure: in real data sets, random components typically exhibit some form of auto-correlation. Hence independently and identically distributed Gaussian noise is not appropriate. Thus simulations in this paper use an auto-regressive moving average ARMA(1,1) process to generate disturbance terms. Consequently, wavelet coefficients on a given frequency level are correlated; hence thresholding based on Stein's unbiased risk estimator SURE (Percival and Walden (2000), chapter 10) is used. Secondly, we use the method in Percival (1995) for estimating the error variance across different frequency levels.

3.3.1. Thresholding framework

Suppose that $Y_{\mathrm{O},t}=Y_{\mathrm{T},t}+\epsilon_t$, $t=1,\ldots,n$, is observed, with ϵ_t being an error term. Transforming $Y_{\mathrm{O}}=(Y_{\mathrm{O},1},\ldots,Y_{\mathrm{O},n})'$ into the wavelet domain by using the orthonormal matrix W, we have

$$w=WY_{\mathrm{O}}, \tag{3.6}$$

$w=(w_1,\ldots,w_n)'$. Typically either hard or soft thresholding (Donoho and Johnstone, 1994) is used. In this paper, we use soft thresholding with estimates obtained as follows:

$$\hat{w}_i=\mathrm{sgn}(w_i)(|w_i|-\lambda)\mathbb{1}(|w_i|\geqslant\lambda), \tag{3.7}$$

$$=\left(1-\frac{\lambda}{|w_i|}\right)\mathbb{1}\,(|w_i|\geqslant\lambda)\,w_i, \tag{3.8}$$

where λ denotes the threshold value (a parameter depending on the noise variance), $\mathrm{sgn}(\cdot)$ is the sign function and $\mathbb{1}(\cdot)$ is the indicator function.

If the magnitude of an observed wavelet coefficient is greater than λ it is shrunk in magnitude by λ. Otherwise it is set to 0. As mentioned above λ is estimated based on SURE; such an estimator depends on both the series length and the variance of the noise term. In particular Percival's estimator based on the maximal overlap discrete wavelet transform is used to estimate the variance of wavelet coefficients; Percival (1995) discussed this in more detail. Consequently λ is a data-dependent parameter, i.e. $\lambda = \lambda(Y_{\mathrm{O}})$. Using equation (3.7), estimates of the true wavelet coefficients are obtained as follows:

$$\hat{w} = \underbrace{\mathrm{diag}(z_1, \ldots, z_n)}_{Z_w} w. \tag{3.9}$$

The diagonal elements of Z_w are $z_i = (1 - \hat{\lambda}/|w_i|)\mathbb{1}(|w_i| \geqslant \hat{\lambda})$, with $\hat{\lambda}$ being a threshold estimate. An estimate of the unobserved true series Y_T is now obtained as

$$\hat{Y}_{\mathrm{T}} = W' Z_w W Y_{\mathrm{O}}. \tag{3.10}$$

3.4. Alternative seasonal model

As seen in Section 4, in many cases the noisy high frequency series is seasonally adjusted (Durbin and Koopman (2001), chapter 3) before benchmarking or thresholding and is reintroduced afterwards. The seasonal component is unknown and hence must be estimated. The time series data that are studied in this paper are represented in state space form (Durbin and Koopman (2001), chapter 3) as this allows most generic models that are used in seasonal adjustment to be represented in one form. To estimate the seasonal component we apply the Kalman smoother (Durbin and Koopman (2001), chapter 4). A stochastic seasonal model that takes the following form is often used:

$$\gamma_{t+1} = -\sum_{j=1}^{f-1} \gamma_{t+1-j} + \omega_t, \qquad \omega_t \overset{\mathrm{IID}}{\sim} N(0, \sigma_\omega^2). \tag{3.11}$$

However, the zero-sum constraint of the seasonal component is violated (i.e. $\Sigma_{j=1}^{f} \gamma_{t+1-j} \neq 0$). Consequently the benchmark constraint will no longer be satisfied once the seasonal estimate has been reintroduced in the series. Therefore the following representation, which allows the seasonal process to vary stochastically while ensuring that the zero-sum constraint is satisfied, is considered:

$$\begin{pmatrix} \gamma_{1,t+1} \\ \vdots \\ \gamma_{f,t+1} \end{pmatrix} = \begin{pmatrix} \gamma_{1,t} \\ \vdots \\ \gamma_{f,t} \end{pmatrix} + \begin{pmatrix} \omega_{1,t} \\ \vdots \\ \omega_{f,t} \end{pmatrix}, \tag{3.12}$$

or, equivalently,

$$\boldsymbol{\gamma}_{\mathbf{t+1}} = \boldsymbol{\gamma}_{\mathbf{t}} + \boldsymbol{\omega}_{\mathbf{t}}. \tag{3.13}$$

In equation (3.12) any season j within a given year $t+1$ takes the value $\gamma_{j,t+1}$ and is equal to its value from the previous year $\gamma_{j,t}$ plus a disturbance term. One way to ensure that the seasonally adjusted series satisfies the benchmark constraint is to define an appropriate correlation structure $(\mathrm{var}(\omega_t) = \sigma_\omega^2\{\mathbf{I_f} - (1/f)\mathbf{I_{f\times 1}}\mathbf{I}'_{\mathbf{f\times 1}}\})$ between the components of $\boldsymbol{\omega}_{\mathbf{t}}$. Therefore the sum of each year's f seasons is constant, i.e.

$$\sum_{j=1}^{f} \gamma_{j,t+1} = \sum_{j=1}^{f} \gamma_{j,t} = \ldots = \sum_{j=1}^{f} \gamma_{j,0} \tag{3.14}$$

holds. Imposing the above correlation structure results in $\mathbb{E}(\mathbf{I}'_{\mathbf{f}\times\mathbf{1}}\boldsymbol{\omega}_{\mathbf{t}}) = 0 = \mathrm{var}(\mathbf{I}'_{\mathbf{f}\times\mathbf{1}}\boldsymbol{\omega}_{\mathbf{t}})$. This, along with the initialization condition $\Sigma_{j=1}^{f}\gamma_{j,0} = 0$, forces the benchmark constraint to hold.

To maintain consistency, other components from structural time series models (i.e. trend, slope and error components) are represented in a similar form to equation (3.12). Such models are known as periodic structural time series; more information is provided in Tripodis and Penzer (2004).

4. Data analysis

We now consider the application of wavelet benchmarking to simulated data and ONS data. The advantages of a wavelet approach to benchmarking discussed in previous sections are supported by diagnostic measures of performance. Since simulated time series are additive, additive methods of benchmarking have been used. However, analogous results hold for multiplicative time series. An algorithmic summary of wavelet benchmarking can be found in Appendix A to aid interpretation of the steps. Initialization and parameter values that were used to generate subsequent simulations are available from `http://wileyonlinelibrary.com/journal/rss-datasets`, along with the code to reproduce the simulations.

4.1. Revision metric for benchmarking

Subsequent sections assess benchmarking methods by using a number of metrics; mean-squared error (MSE), a revision metric and a growth rate metric.

The MSE metric assesses the performance of simulations but real data sets require an alternative metric since the true high frequency series is unobserved. As mentioned earlier, since published economic data impacts decisions that are made by policy makers, producing a stable benchmarked series is important. Therefore, when current data sets are revised or new data become available, adjustments to a benchmarked series should be minor. In particular the effect on latter regions of the benchmarked series is most important since these points describe most recent economic conditions. The following metric measures the sensitivity of the latter regions of a benchmarked series when the observed high and low frequency series are adjusted.

Consider quarterly to annual GDP benchmarking; the series $\{Y^{\mathrm{Q}}_{\mathrm{O},t}\}_{t=1,\ldots,n}$ and $\{Y^{\mathrm{A}}_{\mathrm{O},s}\}_{s=1,\ldots,m}$ are observed with corresponding benchmarked series $\{\hat{Y}^{\mathrm{Q}}_{\mathrm{T},t}\}_{t=1,\ldots,n}$. When new data become available the new benchmarked series $\{\tilde{Y}^{\mathrm{Q}}_{\mathrm{T},t}\}_{t=1,\ldots,l}$ is observed with $l \geqslant n$. A metric focusing on the last year of common benchmarked data is used. It measures the discrepancy between the last four quarters of overlapping time points:

$$\mathrm{metric} = 100 \times \frac{1}{4} \sum_{t=n-3}^{n} \left(\left| 1 - \frac{\tilde{Y}^{\mathrm{Q}}_{t}}{\hat{Y}^{\mathrm{Q}}_{t}} \right| \right). \tag{4.1}$$

(Equation (4.1) provides larger metric readings for upward movements of the benchmarked series compared with downward movements. However, since changes in benchmarked series are relatively small such differences are negligible.) Suppose that p years of additional data become available, so $\{Y^{\mathrm{Q}}_{\mathrm{O},t}\}_{t=1,\ldots,n+4p}$ and $\{Y^{\mathrm{A}}_{\mathrm{O},s}\}_{s=1,\ldots,m+p}$ are now observed. Consequently, we construct p new benchmarked series $\{\tilde{Y}^{\mathrm{Q}}_{\mathrm{O},t}\}_{t=1,\ldots,n+4}, \ldots, \{\tilde{Y}^{\mathrm{Q}}_{\mathrm{O},t}\}_{t=1,\ldots,n+4p}$. We compare each of these new benchmarked series and the original benchmarked series $\{\hat{Y}^{\mathrm{Q}}_{\mathrm{T},t}\}_{t=1,\ldots,n}$. In subsequent sections, the mean of these differences for a suitably chosen p, which will depend on the length of the data, will be referred to as the revision metric. In particular, this metric will be 0 for both the original series and elementary wavelet benchmarking. In this case, additional data have no effect on the estimated high frequency series at earlier time points.

4.2. Growth rate metric for benchmarking

Ideally benchmarked series published by NSIs would additionally preserve the movements of the true original high frequency series. Hence the growth rate metric that is now described is used to measure the ability of a particular benchmarking method to preserve the movements of a series. Denote $\{Y^{\mathrm{H}}_{\mathrm{T},t}\}^n_{t=1}$ and $\{\hat{Y}^{\mathrm{H}}_{\mathrm{T},t}\}^n_{t=1}$ as the true observed high frequency series and benchmarked series respectively. The following metric is used:

$$\text{metric} = \frac{1}{n-1}\sum_{t=1}^{n-1}\left(\left|\frac{Y^{\mathrm{H}}_{\mathrm{T},t+1}-Y^{\mathrm{H}}_{\mathrm{T},t}}{Y^{\mathrm{H}}_{\mathrm{T},t}} - \frac{\hat{Y}^{\mathrm{H}}_{\mathrm{T},t+1}-\hat{Y}^{\mathrm{H}}_{\mathrm{T},t}}{\hat{Y}^{\mathrm{H}}_{\mathrm{T},t}}\right|\right). \tag{4.2}$$

When considering simulated data this growth rate can be calculated exactly. However, when analysing real data sets $Y^{\mathrm{H}}_{\mathrm{T},t}$ is unobserved and so replaced by $Y^{\mathrm{H}}_{\mathrm{O},t}$.

As shown in the simulations, wavelet benchmarking outperforms currently used methods when the growth rate metric compares the true unobserved data and benchmarked data. However, when noisy observed data are compared with benchmarked data, the proportionate Denton method minimizes the growth rate metric. This occurs as a result of the proportionate Denton method's definition; it minimizes the sum of squares between the observed and benchmarked series. Since structural time series model simulations provide a good representation of economic time series being analysed comparisons are made with the true 'unobserved' series. However, if NSIs believe that noisy observations are the best indicator in terms of explaining fluctuations, then the Denton proportionate method provides the optimal solution in terms of this metric.

4.3. Dyadic quarterly and annual data

Wavelet benchmarking is applied to simulated quarterly–annual GDP time series. For simplicity, data are dyadic allowing the Haar transform to be applied. Since univariate structural time series models (Durbin and Koopman (2001), chapter 3) adequately describe many economic processes they are used to generate simulations and in particular do not conform to any of the chosen methodologies providing a valid comparison, not biased to any of the underlying benchmarking methods. 500 simulated data sets were generated by the structural time series model in Appendix C. (Parameter and initialization values used to generate simulations in this section can be found at `http://wileyonlinelibrary.com/journal/rss-datasets`.) Since in certain circumstances performing elementary wavelet benchmarking is sufficient (i.e. small survey error), it is included in the data analysis.

Equation (3.4) decomposed the benchmarked quarterly series into a signal ($Y^{\mathrm{A,Q}}_{\mathrm{T},t}$) and noisy ($R^{J+1,J+2}_{Y^{\mathrm{Q}}_{\mathrm{O},t}}$) component:

$$\hat{Y}^{\mathrm{Q}}_{\mathrm{T},t} = \underbrace{\frac{w^{\mathrm{A}}_{\mathrm{T}}(-1,0)}{2}\varphi^{\mathrm{Q}}_{-1,1}(t) + \sum_{j=0}^{J}\sum_{k=0}^{k_j}\frac{w^{\mathrm{A}}_{\mathrm{T}}(j,k)}{2}\varphi^{\mathrm{Q}}_{j,k}(t)}_{Y^{\mathrm{A,Q}}_{\mathrm{T},t}} + \underbrace{\sum_{j=J+1}^{J+2}\sum_{k=0}^{k_j} w^{\mathrm{Q}}_{\mathrm{O}}(j,k)\varphi_{j,k}(t)}_{R^{J+1,J+2}_{Y^{\mathrm{Q}}_{\mathrm{O},t}}}.$$

As mentioned previously, thresholding the noisy component should produce a more reliable series. However, the structural form of the quarterly time series needs to be considered. Its seasonal component exists primarily on the high frequency regions of the wavelet domain. Thresholding has a tendency to interpret such subtle and localized features as noise; consequently thresholding inadvertently removes the seasonal component.

Removing the seasonal component before benchmarking or thresholding and reintroducing it

Table 2. MSE, revision metric and growth rate metric values of various benchmarking methods averaged over 500 dyadic quarterly and annual simulated series

Series type	*MSE*	*Revision metric*	*Growth rate metric*
Original	13744.63	—	14.42
$\text{Denton}_{a,1}$	8638.15	1.58	13.62
$\text{Denton}_{p,1}$	9113.68	1.52	13.84
Cholette and Dagum	8574.88	1.09	13.59
Elementary wavelet benchmarking	8540.80	0.00†	13.66
Wavelet benchmarking	4728.56	0.83	8.78

†When elementary wavelet benchmarking is implemented, no revisions are made. This is a natural consequence of replacing contaminated wavelet coefficients from the high frequency series with their non-contaminated versions from the low frequency series.

afterwards offers one solution, as seen in Section 3.4. Therefore we used wavelet benchmarking with seasonal adjustment for analysis of the simulations.

MSE, revision metric values (with $p=3$, corresponding to three additional years and approximately 5% of data being available) and growth rate metric values (to calculate the average revision and growth rate metric values the median across all 500 simulations is recorded; when the mean is used a small number of simulations can disproportionately affect the average) for the different benchmarking methods are summarized in Table 2. (When additional data are introduced, it should be noted that data sets are no longer dyadic. Hence a traditional Haar basis is no longer appropriate to transform the data from the time to wavelet domain. Therefore a UH Haar basis was used.) Clearly wavelet benchmarking outperforms all previous methods discussed so far; this is illustrated by its MSE values being lower than the other benchmarking methods' corresponding values. In terms of revisions, elementary wavelet benchmarking produces a benchmarked series which is not revised when new data become available. The revision metric value also implies that wavelet benchmarking outperforms currently used methods in terms of producing a stable benchmarked series. (The same results were also qualitatively found for other values of p; the data are not shown.) The growth rate metric also indicates that wavelet benchmarking outperforms other benchmarking methods in terms of preserving movements in the true unobserved series.

4.4. Comparison with current methods

In this section and Section 4.5, four additional sets of simulations were generated by using different parameter values. The results of these different simulations along with parameter values can be found at `http://wileyonlinelibrary.com/journal/rss-datasets`. These results support the conclusion that wavelet benchmarking outperforms currently used methods.

Simulated data from Section 4.3 relied on the unrealistic assumption that both data sets have dyadic length. This assumption can be relaxed and now non-dyadic monthly and quarterly data sets are analysed. Furthermore, the monthly series has a periodicity of 3, resulting in a non-dyadic relationship between these two data sets. As in Section 4.3 the model that is specified by equations (C.1)–(C.7) in Appendix C is used to generate the high frequency monthly data.

Table 3. MSE, revision metric and growth rate metric values of various benchmarking methods averaged over 500 simulations from non-dyadic monthly and quarterly data sets

Series type	*MSE*	*Revision metric*	*Growth rate metric*
Original (noisy)	6731.98	—	13.11
$\text{Denton}_{a,1}$	3382.58	1.58	11.48
$\text{Denton}_{p,1}$	3657.26	1.40	12.15
Cholette and Dagum	3378.83	1.51	11.48
Elementary wavelet benchmarking	3436.33	0.00	11.71
Wavelet benchmarking	1856.62	0.81	8.26

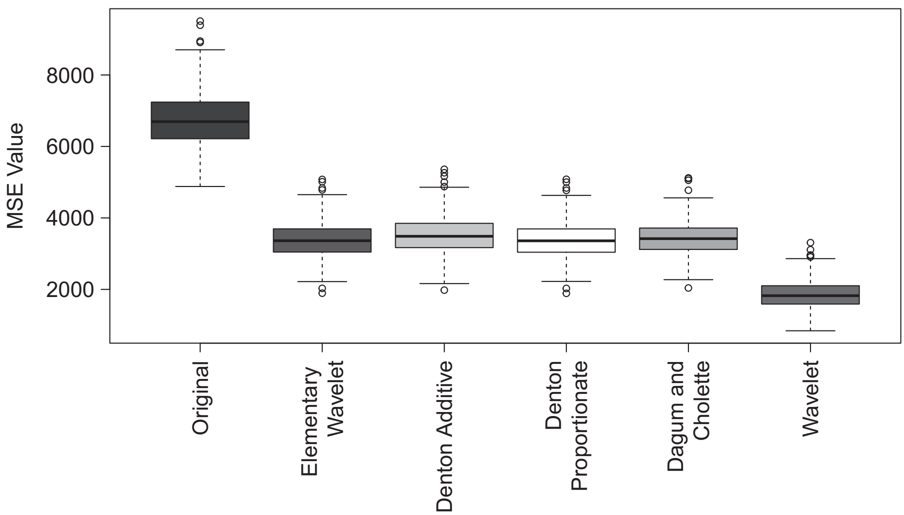

Fig. 3. Boxplot comparing MSE values of various benchmarking methods for 500 simulations generated from non-dyadic monthly and quarterly data sets

Once again 500 simulations were generated. The MSE, revision metric ($p=3$) and growth rate metric values for various benchmarking methods are recorded in Table 3. Fig. 3 shows a boxplot comparing MSE values of the observed series with the benchmarked series.

Results from Table 3 and Fig. 3 are consistent with results from Section 4.3. Elementary wavelet benchmarking performs similarly to currently used benchmarking methods with improvements being offered with wavelet benchmarking. As would be expected wavelet benchmarking outperforms elementary wavelet benchmarking in terms of the MSE for each of the 500 simulations. The revision metric once again implies that wavelet benchmarking produces a more stable benchmarked series in terms of revisions compared with currently used methods. The growth rate metric shows that wavelet benchmarking is the best method in terms of preserving movements in the true series. Such evidence suggests that wavelet benchmarking significantly outperforms currently used methods implemented by NSIs.

4.5. Comparison with current methods (shorter series)

Previous examples used simulated time series with long lengths which are not typically seen in

Table 4. MSE, revision metric and growth rate metric values of various benchmarking methods averaged over 500 simulations from non-dyadic monthly and quarterly simulated series

Series type	*MSE*	*Revision metric*	*Growth rate metric*
Original (noisy)	2392.07	—	20.39
$\text{Denton}_{a,1}$	926.92	2.94	17.68
$\text{Denton}_{p,1}$	1069.29	2.87	18.66
Cholette and Dagum	922.37	2.75	17.71
Elementary wavelet	996.02	0.00	18.75
Wavelet benchmarking	566.46	0.95	13.18

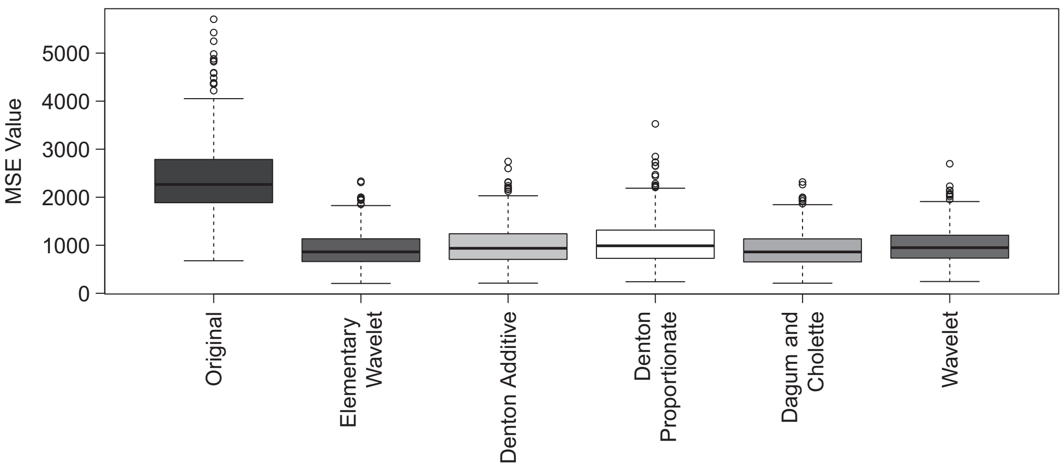

Fig. 4. Boxplot comparing Denton, Cholette and Dagum and wavelet benchmarking: 500 simulations were generated from non-dyadic monthly and quarterly series

time series that are published by NSIs. In reality time series being analysed have smaller lengths. Hence the performance of wavelet benchmarking in this setting is of interest. The same structural time series model as defined by equations (C.1)–(C.7) in Appendix C was used to generate data. The quarterly and monthly time series considered have respective lengths of 10 and 30.

Once again 500 simulations were generated with results summarized in Table 4 and Fig. 4.

As expected, wavelet benchmarking outperforms both Denton and Cholette and Dagum benchmarking; however, improvements from wavelet benchmarking are reduced. This is reflected by comparing MSE values recorded in Table 3 and Table 4. For shorter time series the revision metric (here with $p=1$, given the short length of the series) shows that wavelet benchmarking produces more stable benchmarked series compared with currently used methods. In the shorter series, wavelet benchmarking is the best method for preserving movements in the true unobserved series as seen by the growth rate metric.

4.6. Benchmarking a time series with outliers

As mentioned in Section 1 the time–frequency localization properties of wavelets provide a local solution of benchmarking and thus reduce the potential for artefacts to be introduced

Table 5. MSE, revision metric and growth rate metric values of various benchmarking methods averaged over 500 simulations from non-dyadic monthly and quarterly simulated series with an additive outlier and level shift in the noisy observed high frequency series

Series type	*MSE*	*Revision metric*	*Growth rate metric*
Original (noisy)	6731.98	—	10.73
$\text{Denton}_{a,1}$	3382.58	0.92	9.31
$\text{Denton}_{p,1}$	3635.00	0.85	9.90
Cholette and Dagum	3378.83	0.81	9.31
Elementary wavelet	3436.33	0.00	9.60
Wavelet benchmarking	1121.83	0.62	5.28

when dealing with series with outliers and abrupt structural changes. The 500 simulations from Section 4.4 are reconsidered, but within these series an outlier and level shift were introduced in the observed high frequency series (but not in the true unobserved high frequency series). Initialization and parameter values (including outlier values and times of occurrence) can be found at `http://wileyonlinelibrary.com/journal/rss-datasets`.

The results of these simulations are recorded in Table 5.

Table 5 shows that when outliers are present wavelet benchmarking outperforms currently used methods in terms of MSE, revision metric and growth rate metric readings. This results from the ability of wavelets to isolate local features of a time series.

4.7. *Official Office for National Statistics data*

As mentioned previously benchmarking is frequently applied to time series originating from national accounts. The set of national accounts includes GDP, which is a measure of the value of goods and services produced in a geographic area for a particular period (Office for National Statistics, 2012). Data that are used to construct estimates of GDP come from a range of surveys and administrative data sources measuring the value of goods and services produced by different areas of the economy. There are three different measures of GDP (Office for National Statistics, 2012), and each of these can be broken down into different components measuring different areas of the economy. Benchmarking methods are typically applied early in the process of estimating GDP at a level of detail where the series may be disclosive and are therefore not published because of concerns over confidentiality. However, benchmarking is not only applied in national accounts and could be used in other areas of official statistics. For example, the ONS published data on total turnover from the Annual Business Survey (ABS) (Office for National Statistics, 2010a), and also a higher frequency estimate of total turnover from the Monthly Business Survey (Office for National Statistics, 2010b). Estimates from the monthly survey could be used as an indicator variable and benchmarked to ABS data to give a high frequency estimate that is consistent with the ABS. In the following sections we assess the performance of benchmarking methods by using a set of quarterly and annual series from a component of GDP (Section 4.7.1) and a set of monthly and annual series from the Monthly Business Survey and ABS (Section 4.7.2).

4.7.1. *Component of UK gross domestic product data*

The following section investigates the application of various benchmarking methods to data

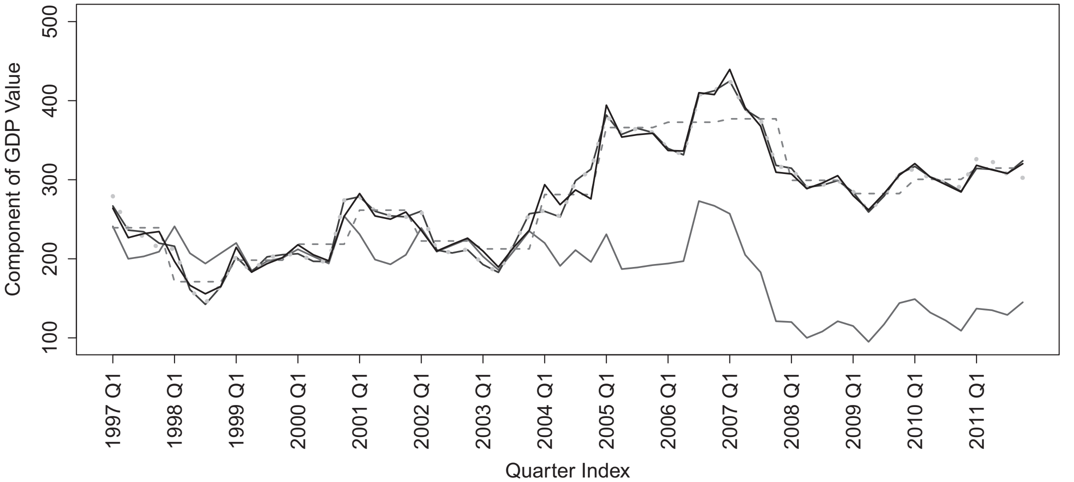

Fig. 5. Application of various benchmarking methods to components of the UK GDP data: ———, quarterly component of GDP series; ———, wavelet benchmarked series; ———, Denton benchmarked series; · · · · ·, Cholette and Dagum benchmarked series; – – – –, quarterly form of annual component of GDP series

Table 6. Metric values for various benchmarked series corresponding to official ONS data

Series type	*Revision metric*	*Growth rate metric*
$\text{Denton}_{a,1}$	0.61	7.27
$\text{Denton}_{p,1}$	1.16	5.33
Cholette and Dagum	1.95	7.19
Elementary wavelet	0.00	6.12
Wavelet benchmarking	0.87	6.61

from UK national accounts. In particular one component of GDP data is considered. For confidentiality (this is a low level component of GDP), the component cannot be named but the data themselves are available from `http://wileyonlinelibrary.com/journal/rss-datasets`. Fig. 5 shows the results of applying quarterly to annual benchmarking to this one component of GDP data.

In Fig. 5 the Denton and Cholette and Dagum versions of the benchmarked series perform similarly. The output of wavelet benchmarking is similar to currently used methods; however, in some time periods wavelet benchmarking performs better at preserving movements in the observed quarterly series. One such time period is from 2004, quarter 1, to 2005, quarter 1. This is due to the localized nature of a wavelet benchmarking solution.

In the time period from 2007, quarter 1, to 2008, quarter 1, the observed quarterly time series seems to exhibit a structural break. This structural break most likely is a result of the economic recession, which began in 2007. By creating a wavelet basis that considers the structure of the observed time series, wavelet benchmarking has the ability to offer further improvements in terms of ensuring that movements in the original quarterly time series are maintained in the formation of the benchmarked quarterly time series. In this paper wavelet bases are determined solely by the length of observed time series. Future work could incorporate the structure of these time series during the selection of wavelet bases.

Table 7. Average metric values for various benchmarked series corresponding to 39 official ONS time series

Series type	*Revision metric*	*Growth rate metric*
$\text{Denton}_{a,1}$	0.35	0.92
$\text{Denton}_{p,1}$	0.34	0.51
Cholette and Dagum	0.08	0.81
Elementary wavelet	0.00	0.92
Wavelet benchmarking	0.46	1.50

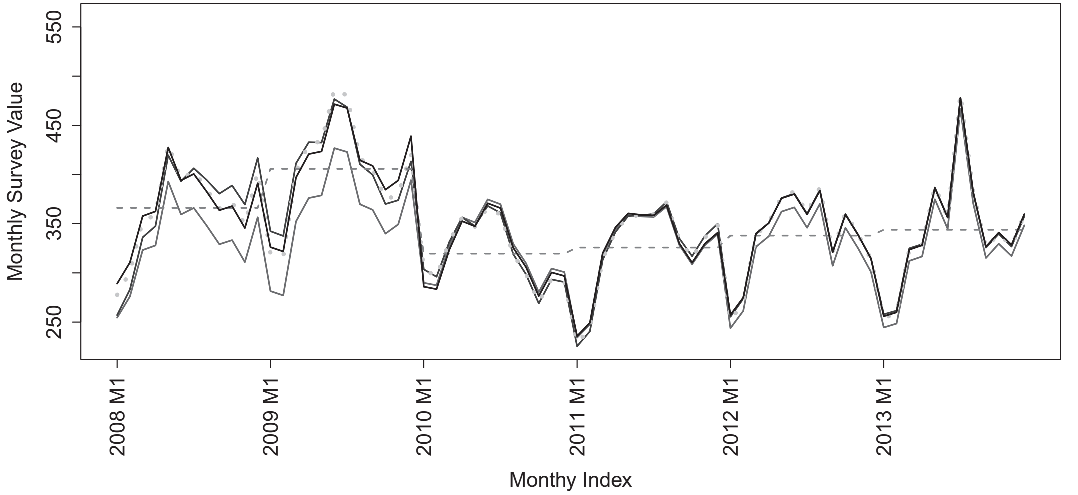

Fig. 6. Application of various benchmarking methods to business survey data on the turnover of companies manufacturing soft drinks and bottled water: ———, monthly component; ———, wavelet benchmarked series; ———, Denton benchmarked series; · · · · ·, Cholette and Dagum benchmarked series; – – – –, monthly form of annual component

Table 6 records the revision and growth rate metric values for the various benchmarking methods. Wavelet benchmarking produces a stable benchmarked series and on the whole performs similarly to currently used methods. For this example the maximum lag length $p=3$ was used, since this corresponds to 20% of the length of the observed series. However, results were not qualitatively different for smaller values of p.

4.7.2. Summary of benchmarking 39 Office for National Statistics quarterly to annual time series

This section considers the mass application of various ONS time series. Monthly to annual benchmarking is considered. However, these time series originate from business surveys as opposed to being components of GDP. For these time series a lag value of $p=1$ is used. The length of the annual time series was 6; this restricted the length of the lag considered.

From Table 7 it can be seen that, across the 39 different business survey time series, Cholette and Dagum and proportionate Denton benchmarking have the lowest revision (excluding elementary wavelet benchmarking) and growth rate metrics respectively. However, the difference between these various benchmarking methods is relatively small. Although wavelet benchmarking is no longer conclusively outperforming currently used methods, it still produces stable benchmarked estimates. Fig. 6 illustrates the application of benchmarking of one of the 39 time series that were considered. In particular, it contains information on the turnover of companies manufacturing soft drinks and bottled water. As can been seen all the benchmarking methods perform relatively similarly.

5. Discussion

Benchmarking is a problem that is frequently encountered by NSIs; this paper has provided an introduction to wavelet-based solutions. Wavelet-based benchmarking consists of a non-parametric and a parametric step. The first step involved introducing true information from the benchmark series into the noisy observed high frequency series via the wavelet domain. Afterwards high frequency wavelet coefficients were thresholded to remove any remaining noise. However, the structural form of time series being analysed had to be considered; in particular the seasonal component is often incorrectly identified as noise and inadvertently removed. Consequently periodic structural time series models were used to adjust the high frequency series seasonally while ensuring that the benchmark constraint was satisfied. After thresholding the seasonally adjusted high frequency series the estimated seasonal component was reintroduced to form the final benchmarked series.

To illustrate wavelet benchmarking both simulated and real data sets were analysed. Simulation studies showed that wavelet benchmarking outperformed currently used methods, whereas the real data showed that elementary wavelet benchmarking has some useful properties in practice.

There are several areas which could be considered to extend work on wavelet benchmarking. Seasonal adjustment could be performed in the wavelet domain, thus allowing the entire benchmarking problem to be considered in the wavelet domain. Secondly, by forcing the benchmarked series to be consistent with the benchmark series there is an implicit and unrealistic assumption that the benchmark series is not contaminated with noise. This assumption can be relaxed; both high and low frequency processes can be treated as noisy. Benchmarking can now be described as optimally combining both high and low frequency processes to create a benchmarked series. The weights would depend on the variances of the distortion terms ϵ_t^{H} and ϵ_t^{L}. This results in the additional complexities of estimating the distortion terms' variances along with determining an optimal estimator. Although NSIs rarely consider non-binding benchmarking, it does have applications to areas outside national accounts. Furthermore since national accounts data sets originate from sample surveys an argument could be presented that non-binding benchmarking is more appropriate. However, this would require a substantial change in the manner in which NSIs process such data.

Wavelet benchmarking can also be extended to situations where multiple constraints must be satisfied. One such example occurs when a time series is classified according to periodicity and geographical location. Benchmark constraints need to be satisfied on both individual and aggregate levels; wavelet benchmarking could facilitate this also.

The selection of wavelet bases needs to be considered in greater detail. This paper constructed such bases on the basis of the length of observed time series. Although a reasonable starting point for an introduction to wavelet benchmarking, bases which incorporate the structure of observed time series could be used in future work. It should also be noted that although Denton and Cholette and Dagum benchmarking naturally provide a framework to consider extrapolation this is currently not considered in the case of benchmarking via wavelets and would be of considerable interest to explore in future research. Finally the ONS performs benchmarking on a large number of time series and therefore would require a method of wavelet benchmarking which can be used in a mass production setting.

Acknowledgement

JA gratefully acknowledges support from the Engineering and Physical Sciences Research Council (EP/K021672/2).

Appendix A: Wavelet benchmarking algorithm

The following summarizes wavelet benchmarking: *input*, a high and low frequency series denoted $Y_{\mathrm{O}}^{\mathrm{H}}$ (length n) and $Y_{\mathrm{O}}^{\mathrm{L}}$ (length m) respectively; *output*, a benchmarked series $\hat{Y}_{\mathrm{T}}^{\mathrm{H}}$.

If seasonality is present *then* seasonally adjust the high frequency series,

$$\check{Y}_{\mathrm{O}}^{\mathrm{H}} = Y_{\mathrm{O}}^{\mathrm{H}} - \hat{\gamma},$$

where $\hat{\gamma}$ denotes the estimated seasonal component; otherwice do not perform seasonal adjustment:

$$\check{Y}_{\mathrm{O}}^{\mathrm{H}} = Y_{\mathrm{O}}^{\mathrm{H}}.$$

(a) *Transform $\check{Y}_{\mathrm{O}}^{\mathrm{H}}$ and $Y_{\mathrm{O}}^{\mathrm{L}}$ from the time to wavelet domain (Section 2.3)*: represent the wavelet transform for $\check{Y}_{\mathrm{O}}^{\mathrm{H}}$ and $Y_{\mathrm{O}}^{\mathrm{L}}$ by the orthogonal matrices W^{H} and W^{L} respectively. This produces the following vector of wavelet coefficients:

$$\begin{pmatrix} y_{\mathrm{O}}^{\mathrm{H}} \\ y_{\mathrm{T}}^{\mathrm{L}} \end{pmatrix} = \begin{pmatrix} W^{\mathrm{H}} & 0 \\ 0 & W^{\mathrm{L}} \end{pmatrix} \begin{pmatrix} \check{Y}_{\mathrm{O}}^{\mathrm{H}} \\ Y_{\mathrm{O}}^{\mathrm{L}} \end{pmatrix}.$$

(b) *Apply elementary benchmarking and thresholding (Section 3.2 and Section 3.3)*: $y_{\mathrm{O}}^{\mathrm{H}}$ is decomposed into a noisy low frequency ($y_{\mathrm{O}}^{\mathrm{H,L}}$) and high frequency ($y_{\mathrm{O}}^{\mathrm{H,H}}$) component:

$$\begin{pmatrix} y_{\mathrm{O}}^{\mathrm{H,L}} \\ y_{\mathrm{O}}^{\mathrm{H,H}} \\ y_{\mathrm{T}}^{\mathrm{L}} \end{pmatrix} = \begin{pmatrix} y_{\mathrm{O}}^{\mathrm{H}} \\ y_{\mathrm{T}}^{\mathrm{L}} \end{pmatrix}.$$

Applying elementary benchmarking results in the following set of high frequency wavelet coefficients:

$$\begin{pmatrix} \frac{1}{c} y_{\mathrm{T}}^{\mathrm{L}} \\ y_{\mathrm{O}}^{\mathrm{H,H}} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \frac{1}{c} I \\ 0 & I & 0 \end{pmatrix} \begin{pmatrix} y_{\mathrm{O}}^{\mathrm{H,L}} \\ y_{\mathrm{O}}^{\mathrm{H,H}} \\ y_{\mathrm{T}}^{\mathrm{L}} \end{pmatrix}.$$

c represents the constant taking the scale difference between the high and low frequency series into account. Thresholding is applied to coefficients existing on high frequency regions, i.e. the coefficients $y_{\mathrm{O}}^{\mathrm{H,H}}$:

$$\hat{y}_{\mathrm{T}}^{\mathrm{H}} = \begin{pmatrix} \frac{1}{c} y_{\mathrm{T}}^{\mathrm{L}} \\ \hat{y}^{\mathrm{H,H}} \end{pmatrix} = \begin{pmatrix} I_m & \mathbf{0}_{m,n-m,} \\ \mathbf{0}_{n-m,m} & Z_{n-m,n-m}(y_{\mathrm{T}}^{\mathrm{H,H}}) \end{pmatrix} \begin{pmatrix} \frac{1}{c} y^{\mathrm{L}} \\ y^{\mathrm{H,H}} \end{pmatrix}.$$

$Z_{n-m,n-m}(y_{\mathrm{T}}^{\mathrm{H,H}})$ is a data-dependent matrix performing the thresholding operation.

(c) *Transform the estimated high frequency wavelet coefficients to the time domain:* this results in the benchmarked series $\tilde{Y}_{\mathrm{T}}^{\mathrm{H}}$:

$$\begin{aligned} \tilde{Y}_{\mathrm{T}}^{\mathrm{H}} &= (W^{\mathrm{H}})' \hat{y}_{\mathrm{T}}^{\mathrm{H}} \\ &= \underbrace{(W^{\mathrm{H}})' \begin{pmatrix} I_m & \mathbf{0}_{m,n-m} \\ \mathbf{0}_{n-m,m} & Z_{n-m,n-m}(y_{\mathrm{T}}^{\mathrm{H,H}}) \end{pmatrix} \begin{pmatrix} 0 & 0 & \frac{1}{c} I \\ 0 & I & 0 \end{pmatrix} \begin{pmatrix} W^{\mathrm{H}} & 0 \\ 0 & W^{\mathrm{L}} \end{pmatrix}}_{A \equiv \text{elementary wavelet benchmarking and thresholding matrix}} \begin{pmatrix} \check{Y}_{\mathrm{O}}^{\mathrm{H}} \\ Y_{\mathrm{O}}^{\mathrm{L}} \end{pmatrix}. \end{aligned}$$

Matrix A expresses the overall benchmarking process in a form that is consistent with equation (2.1). *If* seasonality is present *then* reintroduce the seasonal component $\hat{\gamma}$

$$\hat{Y}_{\mathrm{T}}^{\mathrm{H}}=\tilde{Y}_{\mathrm{T}}^{\mathrm{H}}+\hat{\gamma};$$

otherwise set

$$\hat{Y}_{\mathrm{T}}^{\mathrm{H}}=\tilde{Y}_{\mathrm{T}}^{\mathrm{H}}.$$

Appendix B: Cholette and Dagum benchmarking method

This section provides a comprehensive description of Cholette and Dagum (1994) benchmarking. It is based on the following three stochastic equations:

$$Y_{\mathrm{O}}^{\mathrm{H}}=Hb+Z\delta+\theta+\epsilon_{\mathrm{H}}, \qquad \mathbb{E}(\epsilon_{\mathrm{H}})=0, \quad \mathbb{E}(\epsilon_{\mathrm{H}}\epsilon_{\mathrm{H}}')=V_{\epsilon_{\mathrm{H}}}, \tag{B.1}$$

$$Y_{\mathrm{O}}^{\mathrm{L}}=JZ\delta+J\theta+\epsilon_{\mathrm{L}}, \qquad \mathbb{E}(\epsilon_{\mathrm{L}})=0, \quad \mathbb{E}(\epsilon_{\mathrm{L}}\epsilon_{\mathrm{L}}')=V_{\epsilon_{\mathrm{L}}}, \tag{B.2}$$

$$S\theta=\eta, \qquad \mathbb{E}(\eta)=0, \quad \mathbb{E}(\eta\eta')=V_{\eta}. \tag{B.3}$$

These equations are now discussed in the setting of quarterly to annual GDP benchmarking. The high ($Y_{\mathrm{O}}^{\mathrm{H}}$) and low ($Y_{\mathrm{O}}^{\mathrm{L}}$) frequency series in equations (B.1)–(B.3) are replaced by their quarterly ($Y_{\mathrm{O}}^{\mathrm{Q}}$) and annual ($Y_{\mathrm{O}}^{\mathrm{A}}$) forms respectively.

Equation (B.1) decomposes the observed quarterly process into its true unobserved quarterly process ($Y_{\mathrm{T}}^{\mathrm{Q}}=Z\delta+\theta$) and deterministic ($Hb$) and stochastic ($\epsilon_{\mathrm{H}}$) disturbance terms. Typically H is a vector of 1s and b is a constant column vector forming a bias term capturing the average difference between the observed quarterly ($Y_{\mathrm{O}}^{\mathrm{Q}}$) and annual ($Y_{\mathrm{O}}^{\mathrm{A}}$) series. Z is an $n\times p$ matrix of known regressors and δ is a $p\times 1$ vector of unknown coefficients modelling calendar effects. Although θ may be modelled by a variety of statistical processes, in Cholette and Dagum benchmarking it typically has an auto-regressive integrated moving average structure; this is discussed below.

Equation (B.2) decomposes the observed annual series $Y_{\mathrm{O}}^{\mathrm{A}}$ into its true unobserved annual series $Y_{\mathrm{T}}^{\mathrm{A}}=JZ\delta+J\theta$ and a disturbance term ϵ_{L}. J is an annualizing matrix equivalent to matrix B' from the Denton method. The disturbance component ϵ_{L} is assumed to be Gaussian noise.

Matrix S in equation (B.3) transforms the stochastic component θ into a stationary time series; this requires estimating the order of the seasonal and non-seasonal differencing operators. Set $\theta_t=\upsilon_t+\gamma_t+\epsilon_t$, with υ_t being approximately linear, i.e. $\upsilon_t\approx a+bt$, γ_t capturing quarterly seasonality and ϵ_t being Gaussian random noise. In this scenario, to make θ stationary, the following matrix is required:

$$S=\begin{pmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & \dots \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}_{n-5\times n}. \tag{B.4}$$

This is equivalent to applying the differencing operators $1-L$ and $1-L^4$ to θ. They remove linear and seasonal components from the series respectively, with L denoting the lag operator, i.e. $L\theta_t=\theta_{t-1}$.

Model (B.1)–(B.3) can be written more concisely as

$$\begin{pmatrix} Y_{\mathrm{O}}^{\mathrm{H}} \\ Y_{\mathrm{O}}^{\mathrm{L}} \\ 0 \end{pmatrix}=\begin{pmatrix} H & Z & I_n \\ 0 & JZ & J \\ 0 & 0 & S \end{pmatrix}\begin{pmatrix} b \\ \delta \\ \theta \end{pmatrix}+\begin{pmatrix} \epsilon_{\mathrm{H}} \\ \epsilon_{\mathrm{L}} \\ -\eta \end{pmatrix}, \tag{B.5}$$

or equivalently

$$y=X\alpha+e, \qquad \mathbb{E}(e)=0, \quad V_e:=\mathbb{E}(ee')=\mathrm{block}(V_{\epsilon_{\mathrm{H}}}, V_{\epsilon_{\mathrm{L}}}, V_{\eta}); \tag{B.6}$$

$\mathrm{block}(\cdot,\cdot,\cdot)$ denotes a block diagonal matrix. Cholette and Dagum (1994) provided the following solution:

$$\hat{\alpha}=(X'V_e^{-1}X)^{-1}X'V_e^{-1}y. \tag{B.7}$$

The benchmarked estimate is given by $\hat{\beta} = X^* \hat{\alpha}$, where $X^* = (0\ Z\ I_n)$. The following equation expresses the solution in a form that is consistent with equation (2.1):

$$\hat{Y}_{\mathrm{T}}^{\mathrm{H}} = (O \quad Z \quad I_n) \begin{pmatrix} \hat{b} \\ \hat{\delta} \\ \hat{\theta} \end{pmatrix} = (O \quad Z \quad I_n)(X' V_\mu^{-1} X)^{-1} X' V_\mu^{-1} \begin{pmatrix} Y_{\mathrm{O}}^{\mathrm{H}} \\ Y_{\mathrm{O}}^{\mathrm{L}} \\ 0 \end{pmatrix}. \tag{B.8}$$

In Cholette and Dagum benchmarking the matrices S, $V_{\epsilon_{\mathrm{H}}}$, $V_{\epsilon_{\mathrm{L}}}$ and V_η need to be estimated. To circumvent these estimation difficulties NSIs usually simplify the above model. The behaviour describing the unobserved stochastic component θ is ignored, i.e. equation (B.3) is removed. Since NSIs usually implement binding benchmarking $\epsilon_{\mathrm{L}} = 0$. Finally ϵ_{H} is modelled as an AR(1) process. For practical implementation of Cholette and Dagum benchmarking, Cholette and Dagum (2006), chapter 3, recommended setting the AR(1) parameter value between 0.7 and 0.9 for monthly series and between 0.7^3 and 0.9^3 for quarterly series. For monthly and quarterly time series, the ONS uses parameter values of 0.8 and 0.8^3 (these parameter values can be varied if necessary) respectively (Brown *et al.*, 2012). Naturally such adjustments can in some cases have a negative effect on the accuracy of the benchmarking process. In particular the ONS uses the following form of the Cholette–Dagum model:

$$\begin{pmatrix} Y_{\mathrm{O}}^{\mathrm{H}} \\ Y_{\mathrm{O}}^{\mathrm{L}} \end{pmatrix} = \begin{pmatrix} I_n \\ J \end{pmatrix} (\theta) + \begin{pmatrix} \epsilon_{\mathrm{H}} \\ 0 \end{pmatrix}, \tag{B.9}$$

with ϵ_{H} being an AR(1) process with parameter value 0.8 or 0.8^3 for monthly and quarterly time series respectively.

Appendix C: Simulation methodology

The following section describes how simulated time series were generated. The model below generates the unobserved true high frequency data points:

$$Y_{\mathrm{T},t}^{\mathrm{H}} = \mu_t + \gamma_t, \tag{C.1}$$

$$\mu_t = \mu_{t-1} + \upsilon_t + \varrho_t, \qquad \varrho_t \sim N(0, \sigma_\varrho^2), \tag{C.2}$$

$$\upsilon_t = \upsilon_{t-1} + \zeta_t, \qquad \zeta_t \sim N(0, \sigma_\zeta^2), \tag{C.3}$$

$$\gamma_t = -\sum_{i=1}^{f-1} \gamma_{t-i} + \omega_t, \qquad \omega_t \sim N(0, \sigma_\omega^2). \tag{C.4}$$

The observed non-noisy low frequency time series is obtained by using

$$Y_{\mathrm{O},s}^{\mathrm{L}} = \sum_{t=f(s-1)+1}^{ft} Y_{\mathrm{T},t}^{\mathrm{H}}. \tag{C.5}$$

An ARMA(1,1) process is used throughout the paper to generate disturbance terms. This results in the following observed high frequency series:

$$Y_{\mathrm{O},t}^{\mathrm{H}} = Y_{\mathrm{T},t}^{\mathrm{H}} + \epsilon_t, \qquad \epsilon_t \sim \mathrm{ARMA}(1, 1), \tag{C.6}$$

$$\epsilon_t = \phi \epsilon_{t-1} + \theta \tau_t, \qquad \tau_t \sim N(0, \sigma_\tau^2), \quad |\phi|, |\theta| < 1. \tag{C.7}$$

Initialization values are required to begin the simulation. The values $\mu_1, \upsilon_1, \gamma_1, \ldots, \gamma_{f-1}$ are generated independently from a zero-mean Gaussian process with respective variances $\sigma_{\mu_1}^2, \sigma_{\upsilon_1}^2, \sigma_{\gamma_1}^2, \ldots, \sigma_{\gamma_{f-1}}^2$.

To ensure that simulations can be reproduced the `set.seed()` (R Development Core Team, 2008) function is used to generate pseudorandom numbers. For the slope, trend and seasonal components the following pseudorandom numbers are used respectively: `set.seed(simulation number ×time series number`; `set.seed(2×simulation number×time series number)`; `set.seed(3 ×simulation number×time series number)`. The term simulation number identifies the current simulation being generated, whereas the time series number corresponds to the time point in that current simulation. Details of generating these simulations are available from `http://wileyonlinelibrary.com/journal/rss-datasets`.

References

Brown, G., Parkin, N. and Stuttard, N. (2012) A review of benchmarking methods. In *Proc. 17th Government Statistical Service Conf. Methodology*. Newport: Office for National Statistics.

Cholette, P. (1984) Benchmarking, temporal distribution, and reconciliation methods for time series. Time Series Research and Analysis Division, Methodology Branch, Statistics Canada, Ottawa.

Cholette, P. and Dagum, E. (2006) *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series*. New York: Springer.

Cholette, P. and Dagum, E. B. (1994) Benchmarking time series with autocorrelated survey errors. *Int. Statist. Rev.*, **62**, 365–377.

Daubechies, I. (1992) *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics.

Denton, F. T. (1971) Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *J. Am. Statist. Ass.*, **66**, 99–102.

Di Fonzo, T. and Marini, M. (2012) Benchmarking time series according to a growth rates preservation principle. *J. Econ. Socl Measmnt*, **37**, 225–252.

Donoho, D. L. and Johnstone, I. M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Durbin, J. and Koopman, S. J. (2001) *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Durbin, J. and Quenneville, B. (1997) Benchmarking by state space models. *Int. Statist. Rev.*, **65**, 23–48.

Findley, D. F. (2005) Some recent developments and directions in seasonal adjustment. *J. Off. Statist.*, **21**, 343–365.

Fryzlewicz, P. (2007) Unbalanced Haar technique for nonparametric function estimation. *J. Am. Statist. Ass.*, **102**, 1318–1327.

Nason, G. (2008) *Wavelet Methods in Statistics with R*. Berlin: Springer.

Nason, G. P., von Sachs, R. and Kroisandt, G. (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc.* B, **62**, 271–292.

Office for National Statistics (2010a) Quality and methodology information for annual business survey. Office for National Statistics, Newport. (Available from `http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/quality-and-methodology-information-reports-by-theme/business-and-energy/quality-and-methodology-information-for-annual-business-survey--abs-.pdf/`.)

Office for National Statistics (2010b) Quality and methodology information for monthly business survey. Office for National Statistics, Newport. (Available from `http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/business-and-energy/monthly-business-survey.pdf/`.)

Office for National Statistics (2012) Quality and methodology information for gross domestic product. Office for National Statistics, Newport. (Available from `http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/economic-statistics/gross-domestic-product--gdp-.pdf/`.)

Percival, D. B. (1995) On estimation of the wavelet variance. *Biometrika*, **82**, 619–631.

Percival, D. B. and Walden, A. T. (2000) *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.

Quenneville, B., Cholette, P., Huot, G., Chiu, K. and Di Fonzo, T. (2003) Adjustment of seasonally adjusted series to annual totals. In *Proc. Symp. Challenges in Survey Taking for the Next Decade*, pp. 1–12.

Quenneville, B., Picard, F. and Fortier, S. (2013) Calendarization with interpolating splines and state space models. *Appl. Statist.*, **62**, 371–399.

R Development Core Team. (2008) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Tripodis, Y. and Penzer, J. (2004) Periodic time series models: a structural approach. *Technical Report*. London School of Economics and Political Science, London.

Vidakovic, B. (1999) *Statistical Modeling by Wavelets*. New York: Wiley.