# Methodological Review of Smoothing and Commuting Adjustments in Regional Accounts

March 2012

Duncan Elliott and Joseph Winton

## Executive Summary

Following a UK Statistics Authority assessment of Regional Accounts, this report investigates adjustments made to produce residence based and workplace based estimates and the smoothing of regional data. This report is intended to provide information on methods currently used, inform debate on some of the issues that face users of regional accounts data, recommend areas of further research and address some of the requirements of the UK Statistics Authority assessment report.

**Recommendation 1:**
Regional Accounts should consult users about the proposal to extend commuting adjustments to all regions.

**Recommendation 2:**
Dual run the Regional Gross Value Added system with and without the current smoothing method to derive data at the NUTS2 and NUTS3 level without the potentially distorting effects of smoothing and constraining. Review the current method of smoothing and constraining Regional Gross Value Added at the NUTS2 and NUTS 3 levels.

**Recommendation 3:**
Consult users of the data on their requirements for additional interpretation of year-on-year movements in the raw series at NUTS3, NUTS2 and NUTS1 levels, by industry classifications where appropriate, and for the totals and separate components of Regional Gross Value Added and Regional Gross Disposable Household Income. Consultation should request information from users about their perception of volatility and reliability of estimates and the need for a smoothed series.

**Recommendation 4:**

Review the estimation methods used for deriving estimates of Regional Gross Value Added and Regional Gross Disposable Household Income. Consideration should be made of techniques that reflect user requirements.

**Recommendation 5:**
Revisions information should be available for all published raw and smoothed data for all NUTS and industrial classifications.

## 1. Introduction

The UK Statistics Authority Assessment Report 143 (UK Statistics Authority, 2011) assesses Regional Gross Value Added (RGVA) and Regional Gross Disposable Household Income (RGDHI), against the Code of Practice for Official Statistics (UK Statistics Authority, 2009). The assessment judges that these are designated National Statistics subject to eight requirements. This paper addresses requirement four as set out in the report

**Requirement 4:**

a) Review the methods used to produce residence-based estimates of GVA and smoothed estimates of the Regional Accounts and consult users about the findings; and

b) Review the way in which the smoothed and unsmoothed estimates of the Regional Accounts are presented and explained (para. 3.9)

RGVA is produced on a workplace basis and a residence basis. The residence-based estimates are produced using residence based data to adjust the workplace based estimates. There is some concern over the validity of the assumptions used in this method and that the resulting RGVA on a residence basis may be misleading. Section 2 of this paper reviews the methods for producing residence based estimates.

Estimates for RGVA and RGDHI are produced for a number of different regions based on the Nomenclature of Units for Territorial Statistics (NUTS). These regional estimates are produced by apportioning UK level components of GVA using the most appropriate data available. In 2003 an internal review of Regional Accounts proposed smoothing RGVA and RGDHI over time using a weighted five point moving average (known as a 3x3 MA), as the data were considered to be highly volatile. The UK Statistics Assessment Report 143 remarks that the Scottish Government uses the unsmoothed data in economic analysis. Section 3 of this paper reviews the methods used for producing smoothed estimates of RGVA and RGDHI, while section 4 reviews the way in which smoothed and unsmoothed estimates of the Regional Accounts are presented and explained.

### Background

This section briefly discusses some of the methods and processes used to produce RGVA and RGDHI that are relevant to this review. For more information on methods and concepts see ONS (2010a). Both RGVA and RGDHI are made up of a number of component series and are available for a number of different regions based on NUTS1, NUTS2, and NUTS3 classifications. RGVA is also available for a number of different industrial splits. In general regional data is derived by apportioning UK level data using appropriate data sources. There is a requirement of consistency in aggregation of the data such that the sum of the regional totals for any point in time should be equal to that of the UK level data.

RGVA is available on a workplace-basis and residence-basis. This is discussed further in section 2. RGVA and RGDHI are both currently smoothed. The smoothing takes place at component level. The smoothed series are then constrained so that the sum of the smoothed series over the regions is equal to the unsmoothed UK total for that particular component (by industry in the case of RGVA).

RGVA is smoothed at the NUTS1 level (by component and industry) first. These smoothed data are constrained to the UK unsmoothed total (by component and industry). There are broadly six components of RGVA (Compensation of Employees, Non-Market Capital Consumption, Total Rent, Mixed Income, Gross Trading Profits & Gross Trading Surplus, and Taxes). However, taxes are smoothed by a number of components rather than as a total tax series. There are 13 regions at NUTS1 and 20 industries, which implies (ignoring taxes), that there are some 1,300 time series directly smoothed for RGVA at the NUTS1 level. NUTS2 level data are then derived by apportioning the smoothed, constrained NUTS 1 level data. NUTS2 level data are then smoothed and constrained for 37 regions by 20 industries and two components (GVA and Compensation of Employees) – a total of 1,480 at NUTS2 level. There are two versions of unsmoothed NUTS3 level data; unsmoothed NUTS3 data derived by apportioning unsmoothed NUTS2 level data; and unsmoothed NUTS3 level data derived by apportioning smoothed and constrained NUTS2 level data. The latter version of the unsmoothed NUTS3 level data are smoothed and constrained by 139 regions and 10 industries – 1,390 time series. This means that there are no true raw data at the NUTS2 and 3 levels as they have been derived by apportioning smoothed constrained series.

Raw, unsmoothed data for RGDHI are available at a NUTS3 level for 139 regions by 10 components. These 1,390 time series are smoothed and constrained. The NUTS 2 level raw and smoothed constrained series are simply aggregations of the raw and constrained series from the NUTS3 level, while NUTS1 level data are aggregations of the NUTS2 level.

In total some 5,560 series are smoothed and constrained. The smoothing method was adopted following an internal review that considered the data to show volatility that does not seem a credible reflection of true year-to-year regional economic activity (ONS, 2003).

## 2. Commuting Adjustment Methods

For NUTS1 regions, Regional Accounts produce workplace based and residence based estimates of RGVA. Workplace based RGVA is GVA generated by people who work but do not necessarily live in that region. Residence based RGVA is GVA generated by people who live in that region but do not necessarily work in that region. The component of RGVA that is adjusted to provide these two different estimates is Compensation of Employees (CoE), which is the largest component of RGVA. The difference between workplace based and residence based CoE depends on the level of commuting between regions.

A commuting adjustment has historically been applied to three regions only, the South East, East of England, and London; Regional Accounts assumes that commuting is small or cancels out between all other regions.

A 2009 study (Abramsky, 2009) using Labour Force Survey (LFS) and Annual Survey of Hours and Earnings (ASHE) data found that there was evidence of more significant commuting flows in other regions and proposed that either the commuting adjustment remains as it is, or the adjustment is expanded to include all regions.

### 2.1 Current Methodology

#### Sources of Compensation of Employees

There are different information sources for wages & salaries for different industries. Some of these sources are workplace based (based on business survey data), some residence based (social survey and administrative data) and one which is both[1]. For example, for all industries excluding agriculture, manufacturing, central government and employees of private households, earnings from the Annual Survey of Hours and Earnings (ASHE) is multiplied by the number in employment from the Short-Term Employment Surveys (STES) to get a measure of wages & salaries by region and industry from workplace based sources. A measure of wages & salaries by region and industry from a residence based source is residence-based pay supplied by HMRC, based on a 1% sample of PAYE records[2].

#### Creating Workplace and Residence Based CoE

Regional Accounts start with the workplace based values.

[1] Department for the Environment Food and Rural Affairs provide wages and salaries data for farms. ONS business surveys used to estimate wages and salaries include the Annual Business Survey, the Annual Survey of Hours and Earnings and Short-Term Employment Surveys. The Labour Force Survey (an ONS social survey and therefore residence based information) is also used for estimation purposes in particular industries.

[2] In what follows PAYE-based estimates are distinguished from PAYE records as PAYE forms the basis of residence based estimates. But, for example the regional PAYE-based estimates differ from the raw administrative data due to National Accounts balancing (ONS, 2011a).

For London, East of England and the South East (LEESE) a Greater London commuting adjustment is applied to get residence-based values. For each industry, the total value of workplace based estimates within LEESE is apportioned across the three regions, using the equivalent proportions of the PAYE-based data.

These figures are then adjusted (or constrained) so that national industry totals match the industry totals published in the Blue Book and also so that regional totals match PAYE-based data (which has been appropriately adjusted so that the sum over all relevant industries and regions is equal to the Blue Book national total[3]). This is a process of two-way pro-rating discussed in more detail below.

The result of this process is the residence based CoE.

Workplace based CoE consistent with Blue Book industry totals is then created for LEESE (note that workplace based and residence based CoE are the same for all other regions). For each industry, the total value of residence based CoE within LEESE is apportioned across the three regions, using the equivalent proportions of the unconstrained workplace based wages and salaries data. Calculating the workplace based CoE from the residence based CoE may seem unnecessary as the residence based estimates are calculated from workplace based estimates in the first place. However, the reason for calculating workplace based estimates from residence based estimates is that the residence based estimates have been constrained to Blue Book industry totals and PAYE-based regional totals. The workplace based estimates derived from residence based estimates are then consistent with Blue Book industry totals and workplace based estimates are equal to residence based estimates for regions other than LEESE[4].

### Commuting Adjustment in Algebra

Let:

$x_{ih}^{ASHE \times STES}$ be the wages and salaries value for industry $i$, in region $h$ from workplace based sources (ASHE earnings x STES employment)

$x_{ih}^{PAYE}$ be the wages and salaries value for industry $i$, in region $h$ from residence based sources (PAYE)

Where $\in I$ , $I = \{\text{SIC 07 industries}\}$, $h \in H$ and $H = \{\text{UK NUTS1 Regions}\}$

Define the intermediate step $x_{ih}^{*}$as:

---

[3] The Blue Book estimates of Compensation of Employees at the national level are based on PAYE data, however due to the balancing process and some additional information on wages and salaries they are not equal to the raw PAYE data. For more information on United Kingdom national accounts see for example ONS (2011a).

[4] Note that if an alternative approach were taken whereby workplace based industry totals were constrained to Blue Book industry totals and then residence based estimates derived from these, the resulting estimates for residence based (once constrained to Blue Book industry and PAYE-based regional totals) would differ for regions outside of LEESE, which would not be consistent with the current assumption that commuting between regions cancels out.

$$x_{ih}^{*} = \begin{cases} x_{ih}^{ASHE\times STES} & ,h \notin LEESE \\ \dfrac{x_{ih}^{PAYE}}{\sum_{h\in LEESE} x_{ih}^{PAYE}} \sum\limits_{h\in LEESE} x_{ih}^{ASHE\times STES} & ,h \in LEESE \end{cases}$$

Two way pro-rating is applied to the$x_{ih}^{*}$, to create $x_{ih}^{RES}$such that

For each industry $j$:

$$\sum_{h\in H} x_{jh}^{RES} = x_{j}^{BLUE-BOOK}$$

And for each region $g$:

$$\sum_{i\in I} x_{ih}^{RES} = \sum_{i\in I} x_{ih}^{PAYE}$$

Note that

$$\sum_{i\in I}\sum_{h\in H} x_{ih}^{PAYE} = \sum_{i\in I} x_{i}^{BLUE-BOOK}$$

Where

| | |
|---|---|
| $x_{j}^{BLUE-BOOK}$ | is the total CoE for industry $j$, published in the National Accounts |
| $\sum_{i\in I} x_{ih}^{PAYE}$ | is the total CoE for region $g$, from PAYE-based data |
| $x_{ih}^{RES}$ | is the Residence Based CoE for industry $i$, in region $h$ |

The method of two-way pro-rating is to first constrain the region by industry data to Blue Book industry totals (over all regions and industries). The regional totals (summing across industries), will then be inconsistent with the PAYE-based regional totals. Therefore the data is constrained the PAYE-based regional totals, which again distorts the industry totals and so an iterative procedure is conducted until the above constraints are met.

Now define the Workplace based CoE for industry $i$, in region $h$, $x_{ih}^{WP}$as:

$$x_{ih}^{WP} = \begin{cases} x_{ih}^{RES} & ,h \notin LEESE \\ \dfrac{x_{ih}^{ASHE\times STES}}{\sum_{h\in LEESE} x_{ih}^{ASHE\times STES}} \sum\limits_{h\in LEESE} x_{ih}^{RES} & ,h \in LEESE \end{cases}$$

## 2.2 Proposed Changes to Commuting Adjustments

A 2009 study using LFS and ASHE data found that there was evidence of more significant commuting flows in other regions; specifically from the East Midlands. The report concluded that since commuting from the East Midlands, while significant, was dispersed among many different regions; the two alternatives would be to leave the commuting adjustment as it is, or to expand the adjustment to include all regions.

Expanding the Commuting Adjustment to cover all regions would mean that for all regions, wages and salaries data from PAYE would be put into the adjustment. This means that residence based CoE is PAYE data constrained to Blue Book industry totals and workplace based CoE is the total residence based CoE apportioned across all regions and industries, using the Industry-Region proportions of the raw ASHExSTES data.

The expansion of the adjustment has been applied to three years of CoE data. The likely impact of making this change has been investigated by comparing residence based and workplace based CoE values under the current adjustment and with the expanded adjustment.

## 2.3 Impact of Proposed Changes to Commuting Adjustments

It is difficult to assess the effect of the expanded adjustment since there is no known true value for industry level CoE on either basis at NUTS1.

Four different values for CoE are considered; workplace based under both the current and proposed adjustments and residence based under both the current and proposed adjustments. Due to the nature of the constraining, at the regional level there is little difference between any of the measures, however when looking at the annual growth, the workplace based measure with all regions included in the commuting adjustment behaves differently to the other measures.

### By Region

At a whole region level, there is no difference between the two residence based measures in terms of growth as it is only the industry make up of the regions that will change. However, for the workplace based estimates there is little correlation between the growth in workplace based CoE with LEESE adjustments and the workplace based CoE with adjustments for all regions.

### By Industry

At an industry total level, all four measures are identical as all are constrained to Blue Book totals.

### By Region and Industry

At a region by industry level the workplace based CoE under the current adjustments is highly correlated with the workplace based CoE for all industries – this makes sense as both use ASHExSTES data for the majority of regions.

However, growths in the residence based CoE under the two adjustment methods are not very well correlated – this is due to big differences in the input data.

The significant observation at this level is that the relationship between the workplace based and residence based measures change when the commuting adjustment changes. Under the current adjustment, the two measures are broadly similar (again, both use mostly ASHExSTES data) however when adjusting all regions, the two measures of CoE differ significantly.

## 2.4 Recommendations

This analysis highlights several areas for further investigation.

> **Recommendation 1:**
> Regional Accounts should consult users about the proposal to extend commuting adjustments to all regions.

# 3. Smoothing Methods

In the current approach, many one-dimensional time series are smoothed independently of one another for RGVA and RGDHI. There are a variety of smoothing methods that could be considered. Alexandrov *et al* (2008) provide an overview of some approaches to smooth one-dimensional time series discussing the advantages and disadvantages of the different approaches in the context of trend estimation. Their paper outlines four groups of approaches that they class as model-based approaches, non-parametric filtering, singular spectrum analysis and wavelets. The 3x3 moving average filter used to smooth RGVA and RGDHI is an example of a non-parametric approach to smoothing a one dimensional time series. Alternative approaches to smoothing could take into account spatial correlation that may exist in the data.

Macaulay describes a smooth curve ‘as one which does not change its slope in a sudden or erratic manner’ (Macaulay, 1931:19), but goes on to say that a suitable measure of smoothness is hard to suggest. Moreover, it is not clear in the current context if a smooth curve is required. An internal review (ONS, 2003) and the background notes to the regional accounts statistical bulletin ONS (2011) both suggest that the purpose of smoothing is to remove volatility that does not seem a credible reflection of true year-to-year regional economic activity. The stated purpose of smoothing RGDHI is to reduce the impact of sampling and non-sampling errors in the data sources (ONS, 2010). The main problem is to understand what a credible reflection of year-on-year economic activity is in an objective sense. It is unlikely that this should always be a smooth curve as step changes or outliers may be a true reflection of year-on-year movements.

RGVA and RGDHI both use what is commonly known as a 3x3 moving average to smooth component regional time series[5]. The smoothed time series are then constrained so that the sum over all regions for each time point is equal to that of unsmoothed UK values for that particular industry and component. That is to say the proportions of the smoothed series across component series is preserved at each time point. The main difference between the methods used for smoothing RGVA and RGDHI is the way in which NUTS1, NUTS2 and NUTS3 data are estimated – RGVA can be considered more top down, while RGDHI is built up from the lowest level of aggregation. This section begins with an overview of the methods for smoothing and constraining in general and presents some measures of smoothness that are used to evaluate the methods as they are applied in the estimation of RGVA and RGDHI. The performances of alternative smoothing methods are reviewed briefly in section 3.1 and 3.2 for RGVA and RGDHI respectively.

Smoothing is achieved by filtering out noise from a time series. The choice of the 3x3 moving average out of other possible filters or smoothing methods seems to have been motivated in part by its use in the seasonal adjustment software X-12-ARIMA (Findley *et al*., 1998) and in part by its simplicity due to the sheer number of series to be smoothed.

A time series of raw regional estimates for a particular component, industrial classification, and regional classification $y_{it}$, for region $i$ ($i \in R$) and year $t$ ($t = 1,2,\ldots,T$) is smoothed using a weighted moving average such that

$$z_{it} = \sum_{j=-m_t}^{n_t} w_j y_{it+j}$$

where

$$\sum_{j=-m_t}^{n_t} w_j = 1$$

The weights, $w_j$, are symmetric if $m_t = n_t$, and in the case of the 3x3 moving average used for smoothing RGVA and RGDHI $m_t = n_t = 2$ if $2 < t < T - 1$

[5] The 3x3 moving average is to smooth data across annual frequencies in the X-12-ARIMA seasonal adjustment software used by ONS. For more information on the range of filters available see for example Ladiray and Quenneville (2001)

Asymmetric weights are used if $t \leq 2$ or if $t \geq T-1$

**Table 1: Symmetric and asymmetric weights for a 3 by 3 moving average**

| $\boldsymbol{t}$ | $\boldsymbol{m_t}$ | $\boldsymbol{n_t}$ | $\boldsymbol{w_j}$ |
|---|---|---|---|
| $\boldsymbol{t=1}$ | 0 | 2 | $\frac{1}{27}\{11,11,5\}$ |
| $\boldsymbol{t=2}$ | 1 | 2 | $\frac{1}{27}\{7,10,7,3\}$ |
| $\boldsymbol{2}<t<T-1$ | 2 | 2 | $\frac{1}{27}\{3,6,9,6,3\}$ |
| $\boldsymbol{t=T-1}$ | 2 | 1 | $\frac{1}{27}\{3,7,10,7\}$ |
| $\boldsymbol{t=T}$ | 2 | 0 | $\frac{1}{27}\{5,11,11\}$ |

Table 1 shows the weights used for producing a smoothed estimate at different time points.

The resulting smoothed series are then constrained to provide the final estimate $\hat{y}_{it}$

$$\hat{y}_{it} = \frac{\sum_{i \in R} y_{it}}{\sum_{i \in R} z_{it}} z_{it}$$

where $R$ is the set of regions for a particular NUTS level of classification (1, 2 or 3). Hence

$$\hat{y}_{it} = \sum_{j=-m_t}^{n_t} w_j^* y_{it+j}$$

where

$$w_j^* = a_t w_j$$

$$a_t = \frac{\sum_{i \in R} y_{it}}{\sum_{i \in R} z_{it}}$$

The effect of constraining is therefore to change the weights used for smoothing, multiplying them by some time dependent constant, $a_t$ (for that particular component and industrial classification).

The properties of a filter can be analysed using the gain and phase functions (Jenkins and Watts, 1968). The gain function of a filter shows what happens to the amplitude of cycles when the filter is applied, whereas the phase shows the number of periods that cycles are moved. Chart 1 *(Annex 1; A1)* shows the gain and phase functions for the weights in table 1. The gain functions in chart 1 show that the longer term trend (represented by cycles of greater length) is preserved by the filters while some shorter cyclical movements, that may be interpreted as noise (ABS, 2005) are dampened though less so for the asymmetric filters. The phase functions for these filters show little phase effect for cycles that are predominantly preserved in the data. One problem of this smoothing method is the use of asymmetric weights at the ends of the series, where phase effects are present. The use of asymmetric weights is an implied forecast therefore as new time points are added to the series these points will be subject to revision as they move from being estimated with asymmetric to symmetric weights.

However, the weights that are used to smooth and constrain the data will only be equal to those shown in the table under certain conditions. The effect of constraining on the gain function may be to shift the level or change its shape. The way in which it is changed will be specific to the data. Chart 2 *(Annex 1; A1)* shows a synthetic example of modifying the weights where values of $a_t$ are draws from a random uniform distribution with minimum value of 0.9 and a maximum value of 1.1. As can be seen the shapes of the symmetric and asymmetric versions are not too dissimilar to those shown in chart 1 but the levels of the gain for cycles that may be considered to represent the trend or signal component are dampened or inflated dependent upon whether $a_t$ is less than or greater than one respectively.

The implications of the constraining are that the overall level of the smoothed series ($z_{it}$) can be shifted up or down, and the growth rates of the smoothed series can be distorted as a result of the variability of $a_t$ year-on-year. In some situations constraining may make more sense than in others. For example, if the smoothed series for all regions are lower in value for a point in time than the respective raw series (and by the same proportion) then it makes sense to increase the value by that proportion. Again this is data dependent.

There are many alternative smoothing methods that could be considered. Two simple examples of other non-parametric filters are presented below; a 5-term Henderson filter and kernel smoothing using the Naradaya-Watson estimator[6]. The purpose of these examples is to demonstrate that changes to the smoothing method – while retaining the simplicity and ease of use of a one-dimensional non-parametric filter over time is unlikely to make a big difference to the final series due to the requirement that the sum of the regional estimates is equal to the unsmoothed UK level data. Henderson filters are derived by minimizing the sum of squares of the third difference of the moving average coefficients[7]. There are a range of non-parametric estimators of a smooth regression function that could be considered, the Naradaya-Watson estimator has been chosen arbitrarily as the purpose of this review is not an in depth review of appropriate

[6] See Shumway and Stoffer (2006:74) for example

[7] See for example Ladiray and Quenneville (2001)

smoothing methods, but to consider the effect of smoothing and provide some points of discussion for users of regional accounts data[8].

Chart 3 *(Annex 1; A2)* shows the symmetrical weights for the 3x3 MA, 5-term Henderson filter and the kernel smoothing using the Naradaya Watson estimator with a bandwidth of $\frac{50}{27}$, a series of 15 periods, and a normal kernel. As can be seen the weights differ slightly. The kernel smoothing gives some weight to all time points whereas the 3x3 MA and 5-term Henderson filter are both weighted 5-point moving averages. Note that the kernel filter weights are only symmetrical for the centre of the series in this instance, while the 3x3 MA and 5-term Henderson are both symmetrical for the periods as shown in table 1.

The gain and phase functions for these three filters are shown in chart 4 *(Annex 1; A3)* (note the charts do not distinguish between symmetric and asymmetric filters, and are presented to give an overall picture of the weights used to smooth the series). It can be seen from chart 4 that the kernel filter should be smoother followed by 3x3 MA and then the 5-term Henderson. The 5-term Henderson filters have a smaller phase effect for the cycles preserved by the filter.

In what follows raw series are data that has not been smoothed, smoothed series will refer to those that have had particular smoothing methods applied (without constraining), while constrained series will refer to series that have been smoothed and then constrained as outlined above.

It is assumed that smoothness is a desired property of the series, based on statements about the volatility of year-on-year movements in the series. Given that the final smoothed estimates are a combination of smoothing and constraining the following measures are used to assess the data. Two measures of smoothness used by Dagum and Morry (1984) include

$$S = \sqrt{\frac{1}{T-1}\sum_{t=2}^{T}(r_t - \bar{r}_t)^2}$$

$$D = \frac{1}{T-2}\sum_{t=3}^{T}|r_t - r_{t-1}|$$

where

$$r_t = \frac{x_t - x_{t-1}}{x_{t-1}}$$

8 Note for example that Clark (1977) presents a new family of non-parametric estimators that may have advantages over the Naradaya-Watson estimator in small samples, which may be more appropriate in this context.

$$\bar{r}_t = \frac{1}{T-1}\sum_{t=2}^{T} r_t$$

and $x_t$ denotes the series for which the measure is being calculated (raw, smoothed or constrained).

Lower values of $S$ and $D$ indicate smoother series. The above smoothness measures allow comparison of the smoothness of raw, smoothed and constrained series. As the purpose of smoothing is to reduce volatility in the year-on-year movements it is assumed that the growth rates of the smoothed series are considered to be more informative than the raw. Therefore if constraining distorts the growth rates of the smoothed series this is considered to be undesirable. A further measure of evaluating the smoothing and constraining is therefore to count the number of times in the series where the growth rates in the smoothed series are in the opposite direction to that of the constrained series for the same time points. The following measure is also computed to measure the effects of smoothing and constraining

$$S_1 = \sqrt{\frac{1}{T-1}\sum_{t=2}^{T}(r_t - v_t)^2}$$

where $r_t$ and $v_t$ are growth rates of raw, smoothed or constrained series (for example comparing the growth rate of the smoothed series to constrained series).

It has been assumed that smoothness is a desired property and no attempt has been made to consider the effects of outliers or level shifts that may represent events that should be shown in the series. When smoothing the data if there are particular events then the effect of these will appear in periods other than the ones that they occurred. The filters considered here are not robust to outliers, and this issue should be considered when deciding on the desirability of smoothing, or appropriate smoothing methods.

A final measure that could be considered when comparing filters is the implicit forecast error that arises due to the use of the asymmetric filters. This has not been done in this review, but if further work is required on smoothing methods such a measure may be appropriate to consider.

## 3.1 Smoothing RGVA

Given the sheer number of series based on the number of components and industries covered, some example series are presented where the combination of smoothing and constraining may cause some problems for interpretation of movements. It should be noted that GVA is not directly smoothed in practice, instead the component series are as discussed in above. Nevertheless the indicative results hold.

Chart 5 *(Annex 1; A4)* shows that the effect of constraining is to exacerbate the volatility in the series and the constrained series bears little relationship to the smoothed series. If the smoothed growth rates are considered more representative of actual movements in the series then constraining is causing a significant problem.

Chart 6 *(Annex 1; A4)* shows the same series as in chart 5, though smoothed using a 5-term Henderson filter. As can be seen the smoothed series appears less smooth compared to the 3x3 MA, as is expected based on the discussion above, but the same problem occurs where the constraining distorts the message of the smoothed series. However, as the 5-term Henderson does not smooth as much as the 3x3 MA it means that the distortion from smoothed to constrained is less severe than is the 3x3 MA case.

The same problem is again present in the case of kernel smoothing, using the kernel smoothing filter discussed above, as shown in chart 7 *(Annex 1; A5).* If greater smoothing (longer filters such as a 3x5, 3x9 etc or 7-term Henderson, or a larger bandwidth in the case of kernel smoothing) were to be applied the smoothed series would obviously be smoother but the final constrained series would likely be much the same as shown in charts 5 and 7.

These are extreme examples from the data analysed, and chosen to show what problems can occur, and there are a number of series where this effect is less extreme. To summarise the effects for a greater number of series charts 8 to 12 (*Annex 1; A6-A10*) present boxplots of the smoothness measures outlined above, including the counts of changing the direction of growth rates when moving from smoothed to constrained series. For each panel in the charts the measures are calculated for all time points and industrial classifications by region within a stated component of RGVA. In each case the top left panel of the chart shows the ratio of the mean absolute deviation of growth rates ($D$) for the raw series to that of the constrained series. Therefore values of one imply that they have the same smoothness. Values of less (more) than one imply the raw (constrained) series is smoother. The top right panel shows the ratios of root mean square difference ($S_1$), where the numerator is the root mean square difference between raw and smoothed series and the denominator is the root mean square difference between the growth rates of the constrained and smoothed series. Therefore a value of one suggests that the raw and constrained series are on average equally different to the smoothed series. The bottom left panel shows the ratios of the standard deviation ($S$) of growth rates of raw to constrained data. Finally the bottom right panel shows the distribution of occurrences in series (of 14 time points in length) where the growth rates of the smoothed and constrained for a particular time point are showing the opposite direction. Again due to the number of components that are directly smoothed only a few results are presented.

Chart 8 shows that for the Compensation of Employees the process of smoothing and constraining is on average for most industrial classifications having some effect, although there are some series where smoothing and constraining is making the series very marginally more volatile, but not consistently by all measures of smoothness. For each region over 50 per cent of series experience one or more changes in the direction of the growth rates when moving from smoothed to constrained.

Chart 9 shows that for Non Market Capital Consumption there are a number of occurrences where the process of smoothing and constraining is not making the final constrained series any smoother than the raw series. The constraining process also seems to be distorting the interpretation in terms of direction of growth rates for a number of series. This is more pronounced in some regions that others, notably region N (Northern Ireland.).

Chart 10 shows that the effect of smoothing and constraining for the majority of series results in series that are no smoother than the raw series, while the constraining changes the interpretation of growth rates from that of the smoothed for two or more time points for over 50 per cent of series.

Chart 11 shows that in the case of Mixed Income, smoothing and constraining makes the series marginally more smooth, but the constraining effects the interpretation of growth rates relatively often. Chart 12 shows a very similar situation to char 11 for Total Gross Trading Profit & Gross Trading Surplus.

Box plots of measures of smoothness are not presented here for the alternative examples of non-parametric filters discussed as the results are not dissimilar to those in charts 8 to 12. As might be expected the 5-term Henderson filter produces final constrained series that are in general smoother than the raw series (though not always, as seen in charts 11 and 12 in particular), but marginally less smooth compared to the 3x3 filter. The 5-term Henderson filter also reduces slightly the occurrence of growth rates of the smoothed and constrained in the opposite direction (as might be expected given that the smooth series is closer to the raw compared to the 3x3 MA). The opposite is therefore true for the kernel smoothing which produces marginally smoother estimates than the 3x3 MA and slightly increases frequency of growth rates in the opposite direction for smoothed and constrained series.

For the case of smoothing RGVA only the NUTS1 level has been considered, due to the complexity of the system for apportioning NUTS2 and NUTS3 level data. While the results of the NUTS1 level seem to suggest that smoothing with constraining is perhaps distorting the information in the series, it has not been possible as part of this review to examine the results for NUTS2 and NUTS3.

> **Recommendation 2:**
> Dual run the Regional Gross Value Added system with and without the current smoothing method to derive data at the NUTS2 and NUTS3 level without the potentially distorting effects of smoothing and constraining. Review the current

method of smoothing and constraining Regional Gross Value Added at the NUTS2 and NUTS 3 levels.

### 3.2 Smoothing RGDHI

The process of smoothing and constraining for RGDHI is the same as RGVA for the individual series that are smoothed and constrained. However, the main difference between the approaches is that RGDHI smoothes and constrains the NUTS3 level data first. NUTS2 level data raw and constrained is simply the sum of the respective NUTS3 level data, and NUTS1 level is the sum of the NUTS2 level data. This means that there is true raw data at each level, unlike RGVA where the pseudo-raw series for NUTS2 and 3 level data are apportioned versions of the NUTS1 series.

Chart 13 *(Annex 1; A11)* shows smoothed and constrained estimates for Gross Operating Surplus (a component of RGDHI) at the NUTS 3 level for UKL14 (South West Wales) using the three different filters described above. This is just one example from the 133 regions at the NUTS3 level, but is fairly representative of the effect of smoothing and constraining in all regions for this component. Generally it is found that the constrained and raw series are not very different from one another and the constraining can cause some large differences compared to the smoothed series.

As is expected from the discussion above the 5-term Henderson filter is less distorted by the constraining as the smoothed series more closely follows the raw series.

None of the Gross Operating Surplus series at the NUTS3 level seem to be particularly volatile, and in most cases the constrained series seem to follow the raw series quite closely. As the raw and constrained NUTS2 and NUTS1 levels are aggregations from this level of data the same results can be expected to follow up the aggregation. Chart 14 *(Annex 1; A12)* shows similar results to chart 13 but for GDHI (note that GDHI is not directly smoothed as it is component series that are smoothed, but it serves the purpose of comparing results through aggregation).

Given the bottom up nature of the approach different results could be obtained if smoothing and constraining NUTS1 and NUTS2 level directly rather than indirectly estimating them by aggregation. Chart 15 *(Annex 1; A13)* shows the NUTS3 level data aggregated to NUTS1 level for the North West Region. As can be seen the same result as presented in chart 14 is present, where constraining effects the 3x3 MA and kernel smoothing estimates more than the 5-term Henderson filter.

An alternative approach to deriving the NUTS1 level data would be to directly smooth and constrain it rather than derive it indirectly. Chart 16 *(Annex 1; A14)* shows direct estimates. As can be seen there is very little difference between the smoothed and constrained versions whether directly or indirectly estimated.

From the series analysed (GDHI and Gross Operating Surplus) the series in general do not seem to be very volatile even at the NUTS3 level. Moreover the process of smoothing and then constraining leaves the final estimate – the constrained series – as very close to the raw series. This is confirmed for the smoothness measures shown in chart 17 *(Annex 1; A15)* which plots similar information to that given in charts 8 to 12, although not presented as box plots as there is no break down by industry, so only one value of smoothness per region. Chart 17 is not replicated for the alternative filters discussed above as the results are very similar. There is a slight reduction in the frequency of opposite growth rates between the smoothed and constrained series when using the 5-term Henderson filter.

**Recommendation 3:**
Consult users of the data on their requirements for additional interpretation of year-on-year movements in the raw series at NUTS3, NUTS2 and NUTS1 levels, by industry classifications where appropriate, and for the totals and separate components of Regional Gross Value Added and Regional Gross Disposable Household Income. Consultation should request information from users about their perception of volatility and reliability of estimates and the need for a smoothed series.

**Recommendation 4:**

Review the estimation methods used for deriving estimates of Regional Gross Value Added and Regional Gross Disposable Household Income. Consideration should be made of techniques that reflect user requirements.

## 4. Presentation of smoothed and unsmoothed Regional Accounts data

The relevant principles and practices related to the presentation of smoothed and unsmoothed Regional Accounts data highlighted by UK Statistics Authority (2011) include Principle 4, Practice 1 and Principle 8 Practices 1 and 2. This review comments on the current presentation under each of these principles below

***Principle 4, Practice 1:*** *Ensure that official statistics are produced according to scientific principles. Publish details of the methods adopted, including explanations of why particular choices were made.*

This review is intended to provide some further information than is currently available on the methods of smoothing that are currently applied. Some attempt has been made to explain the reasons for the current smoothing and constraining methods. However, this review generally recommends that the current smoothing and constraining process should not be continued and that users should be consulted on their requirements for additional interpretation and analysis of year-on-year movements in series.

***Principle 8, Practice 1:*** *Provide information on the quality and reliability of statistics in relation to the range of potential uses, and on methods, procedures, and classifications.*

The production of quality measures for National Accounts is not a straightforward issue due to the range of data sources and processes involved in producing final estimates. This is explained for example in the summary quality report for RGDHI (ONS, 2010). Revisions are used as a measure of accuracy, which is reasonable. However, the revision information produced in the summary quality report for RGDHI is for revisions to raw RGDHI at NUTS1 level. Revisions are published in the statistical bulletin for RGVA for the UK and the NUTS1 level regions, but not for the estimates of England. The revisions information presented in the table should follow the information published in the tables provided in the bulletin.

> **Recommendation 5:**
> Revisions information should be available for all published raw and smoothed data for all NUTS and industrial classifications.

***Principle 8, Practice 2:*** *Prepare and disseminate commentary and analysis that aid interpretation, and provide factual information about the policy or operational context of official statistics. Adopt formats for the presentation of statistics in graphs, tables and maps that enhance clarity, interpretability and consistency.*

This review is intended to provide information to users on the potential effects of smoothing and constraining. If it is believed there is volatility in the year-on-year movements in series that is problematic for the uses to which the data are put then appropriate analysis should be made to highlight this. The regular publication of commentary and analysis of the regional accounts data should be adjusted appropriately following user consultation.

**References:**


Abramsky J. (2009) ‘2009 Commuting Report’ Internal ONS Paper.

Alexandrov, T., S.Bianconcini, E.B. Dagum, P. Maass, T. McElroy (2008) ‘A Review of Some Modern Approaches to the Problem of Trend Extraction’ *Research Report Series (Statistics #2008-3),* U.S. Census Bureau http://www.census.gov/ts/papers/rrs2008-03.pdf

ABS (2005) ‘An Introductory Course on Time Series Analysis – Electronic Delivery’, *Information Paper* 1346.0.55.001*,* Australian Bureau of Statistics http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1346.0.55.001Jan%202005?OpenDocument

Clark, R.M. (1977) ‘Non-Parametric Estimation of a Smooth Regression Function’ *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 107-113

Dagum, E.B. and M. Morry (1984) 'Basic Issues on the Seasonal Adjustment of the Canadian Consumer Price Index' *Journal of Business & Economic Statistics*, Vol. 2, No. 3 (Jul., 1984), pp. 250-259

Findley, D, B.C. Monsell, W.R. Bell, M. C. Otto, B. Chen (1998) 'New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program' *Journal of Business & Economic Statistics*, Vol. 16, No. 2 (Apr., 1998), pp. 127-152

Jenkins, G.M. and D. Watts, (1968), *Spectral Analysis and Its Applications*, Holden-Day.

Ladiray, D. and B. Quenneville (2001) *Seasonal Adjustment with the X-11 Method* Springer: New York

Macaulay, F.R (1931) *The Smoothing of Time Series,* National Bureau of Economic Research, http://www.nber.org/books/maca31-1

ONS (2003) 'Options for publishing NUTS1 GVA estimates 1989-2001' Internal ONS paper.

ONS (2010) 'Summary Quality Report for Regional Gross Disposable Household Income' Information Paper http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/economic-statistics/summary-quality-report-for-regional-gross-domestic-household-income--gdhi-.pdf

ONS (2010a) 'Regional Accounts Methodology Guide' http://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/regional-accounts/regional-accounts-methodology-guide.pdf

ONS (2011) 'Regional, sub-regional and local Gross Value Added 2010' Statistical Bulletin http://www.ons.gov.uk/ons/rel/regional-accounts/regional-gross-value-added--income-approach-/december-2011/sbd-regional-gva-dec-2011.pdf

ONS (2011a) 'United Kingdom National Accounts: The Blue Book' http://www.ons.gov.uk/ons/rel/naa1-rd/united-kingdom-national-accounts/2011-edition/bod---blue-book-2011.pdf

Shumway, R.H. and D.S. Stoffer (2006) *Time Series Analysis and Its Applications With R Examples* Springer Science+Business Media: New York

UK Statistics Authority (2009) 'Code of Practice for Official Statistics' http://www.statisticsauthority.gov.uk/assessment/code-of-practice/code-of-practice-for-official-statistics.pdf

UK Statistics Authority (2011) 'Assessment of compliance with the Code of Practice for Official Statistics: UK Regional Accounts' Assesment Report 143 November 2011 http://www.statisticsauthority.gov.uk/assessment/assessment/assessment-reports/assessment-report-143---uk-regional-accounts.pdf

## ANNEX 1 - Charts

**Chart 1: Gain function for symmetric and asymmetric weights of a 3x3 MA**

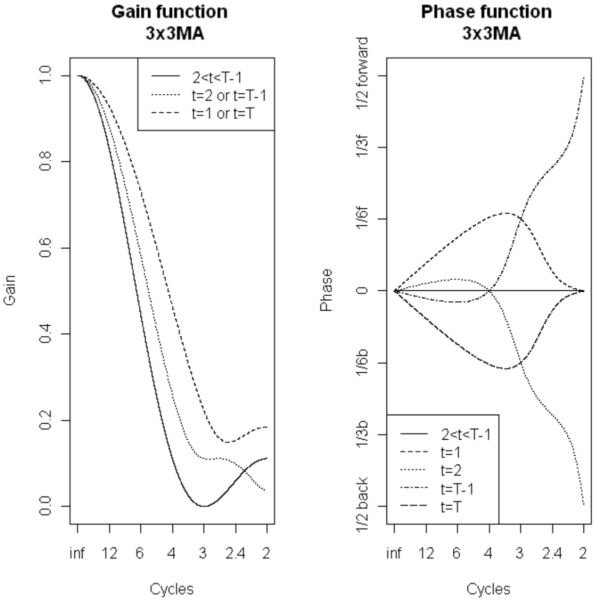



**Chart 2: Gain function for symmetric and asymmetric weights of a synthetically constrained 3x3 MA**

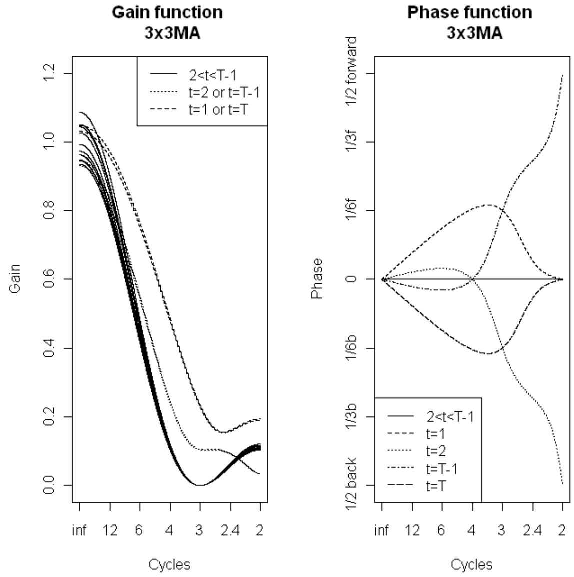


**Chart 3: Symmetric weights for 3x3 MA, 5-term Henderson and a kernel smoothing filter (normal kernel, bandwidth=$\frac{50}{27}$ and 15 time points)**

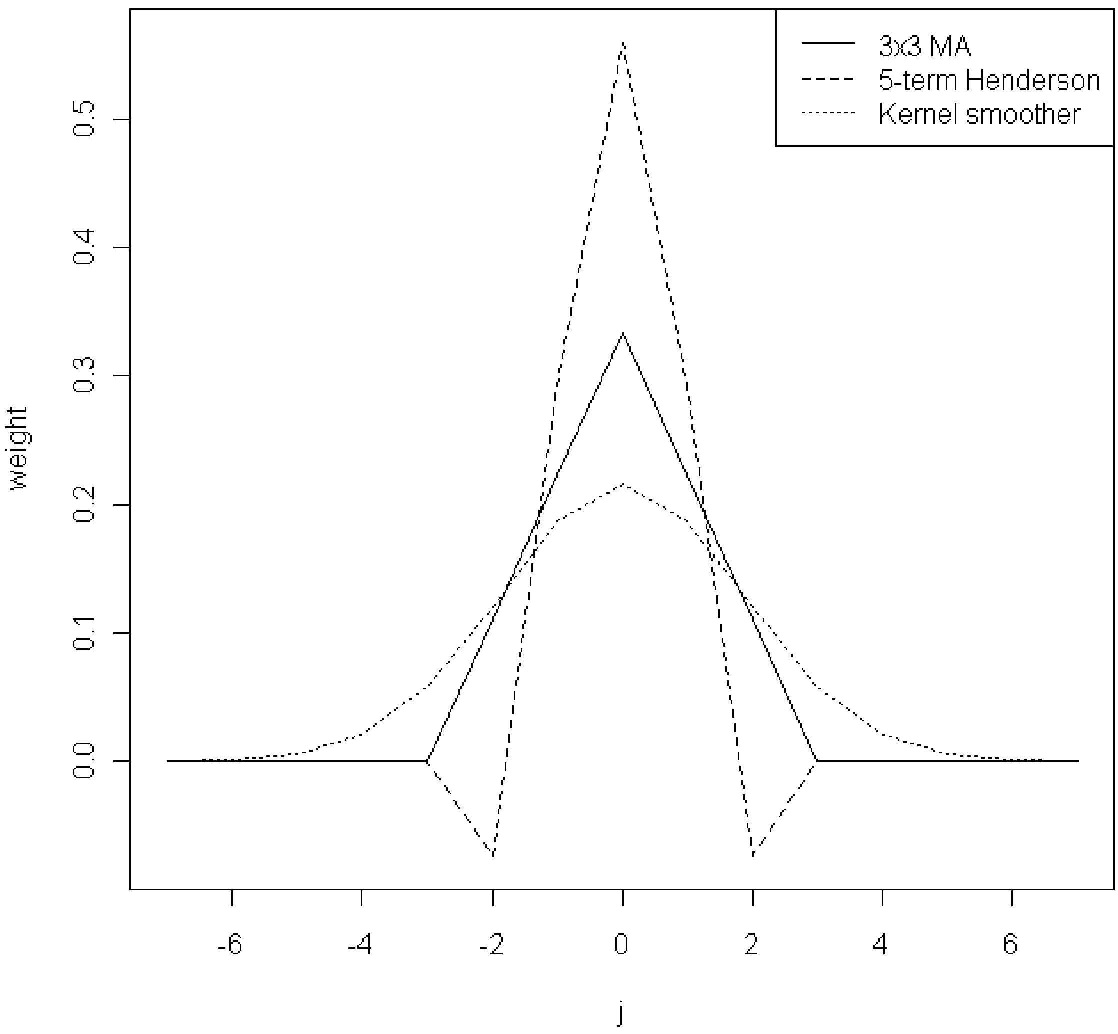


**Chart 4: Gain and phase functions for for 3x3 MA, 5-term Henderson and a kernel smoothing filter (normal kernel, bandwidth=$\frac{50}{27}$ and 15 time points)**

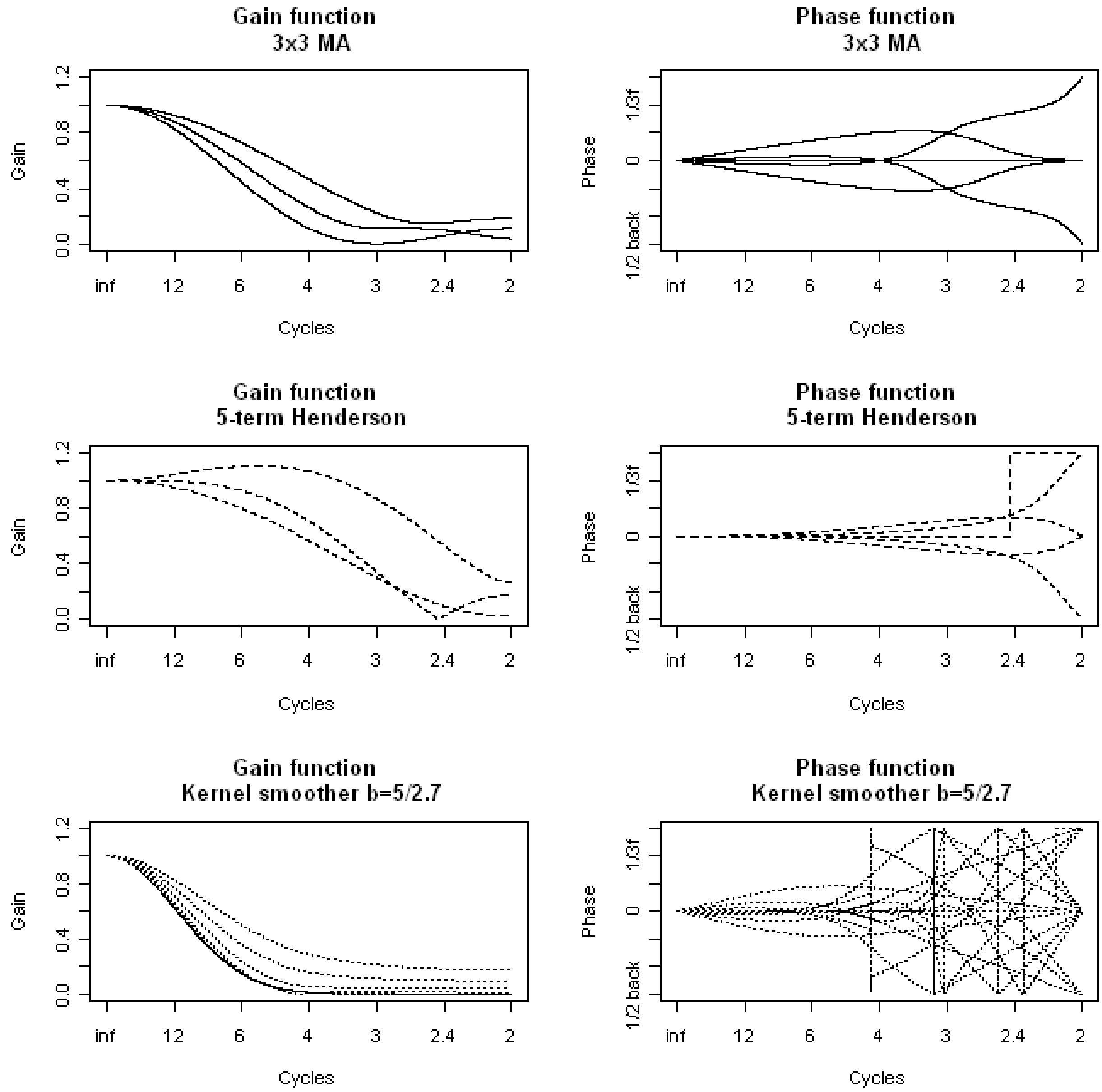


**Chart 5: Raw, smoothed and constrained series using 3x3 MA for RGVA NUTS1 region D (North West), industry A (agriculture)**

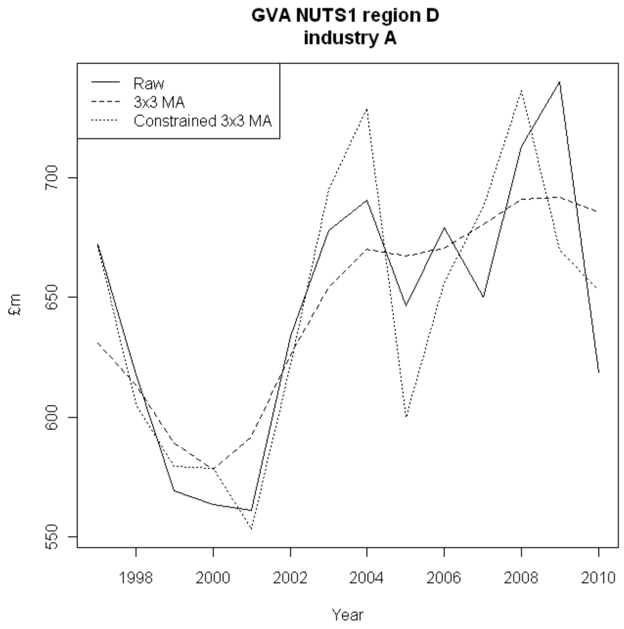



**Chart 6: Raw, smoothed and constrained series using 5-term Henderson filter for RGVA NUTS1 region D (North West), industry A (agriculture)**

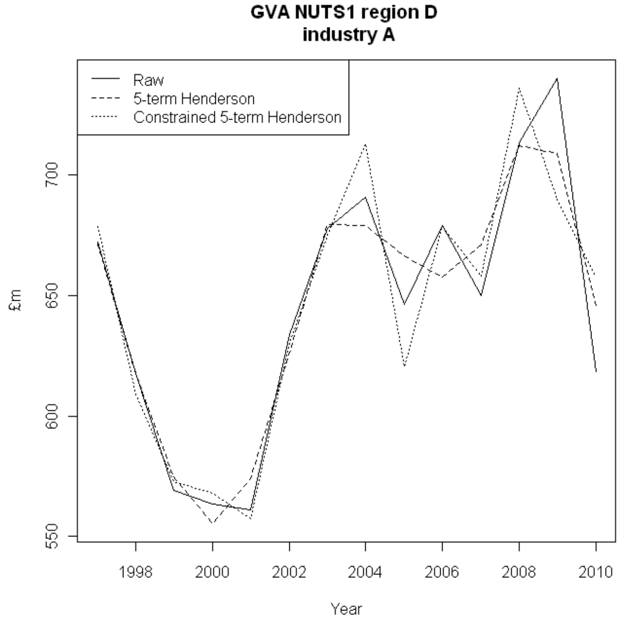


**Chart 7: Raw, smoothed and constrained series using kernel smoothing filter for RGVA NUTS1 region D (North West), industry A (agriculture)**

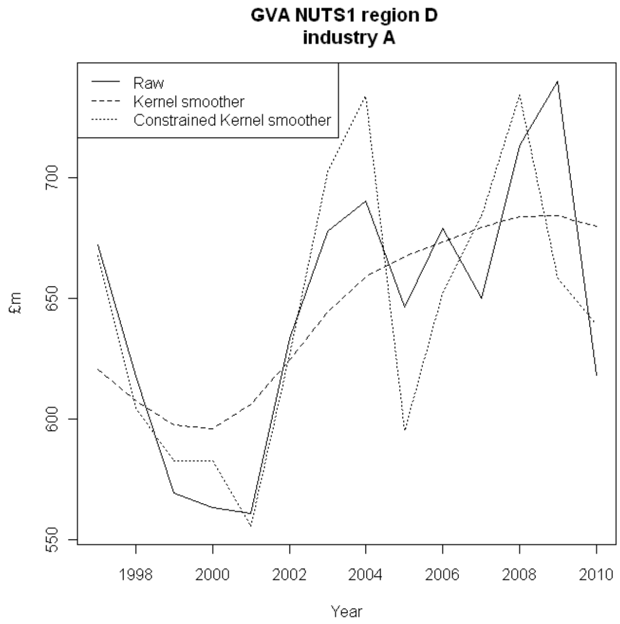


**Chart 8: Box plots of the ratios of smoothness measures for the component Compensation of Employees at the NUTS 1 level which covers 21 industries and 13 regions.**

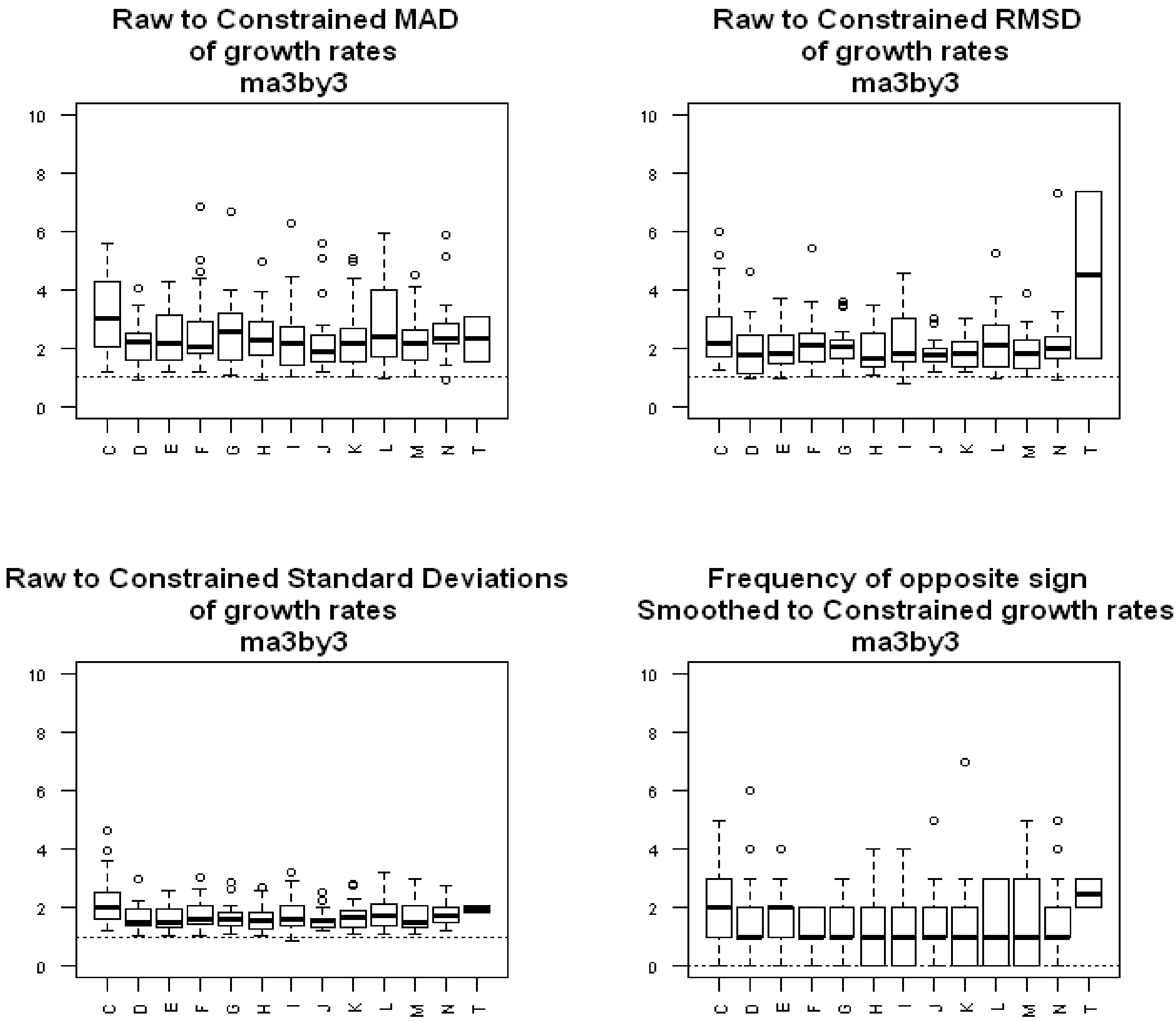


**Chart 9: Box plots of the ratios of smoothness measures for the component Non Market Capital Consumption at the NUTS 1 level which covers 21 industries and 12 regions**

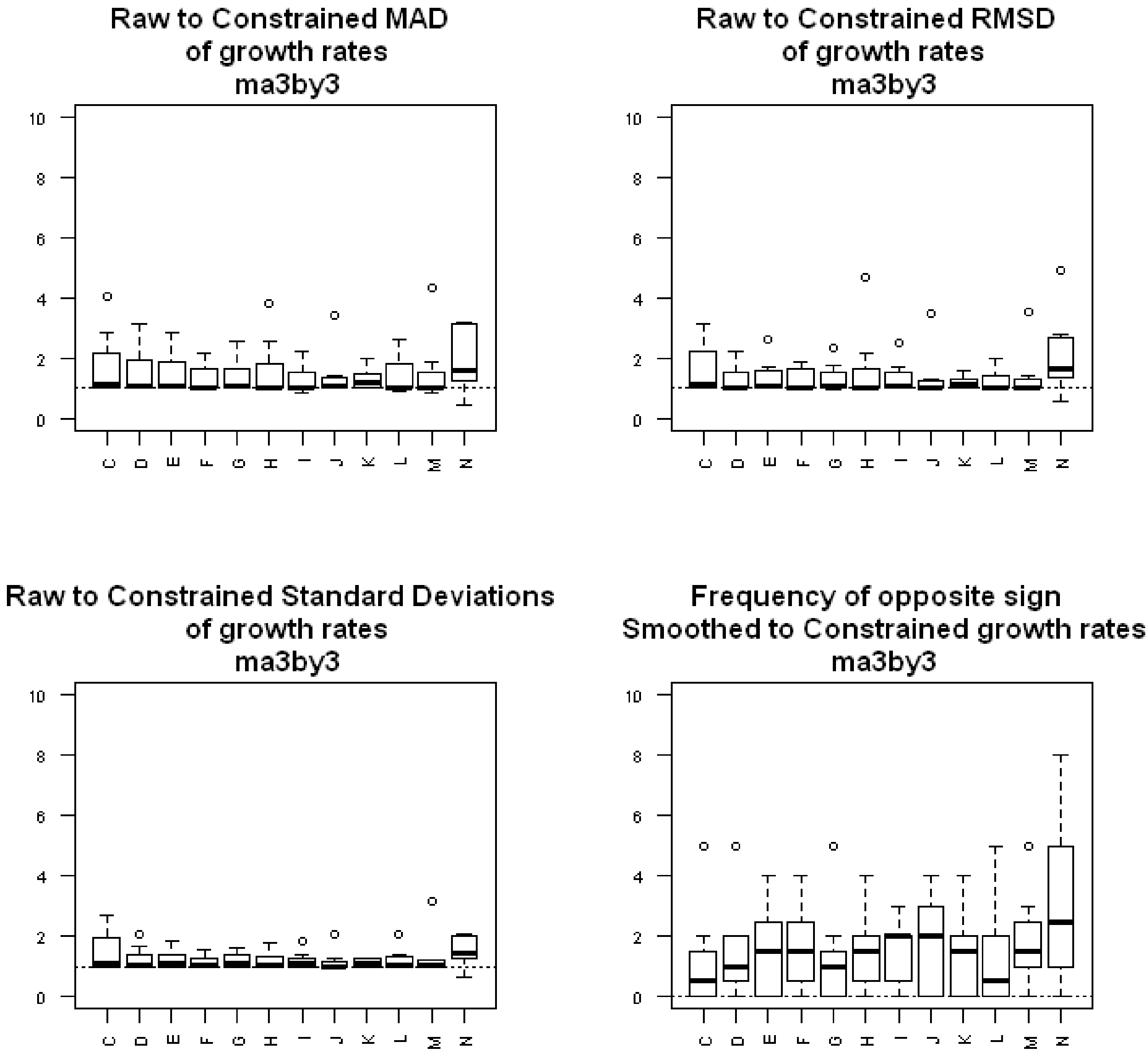


**Chart 10: Box plots of the ratios of smoothness measures for the component Total Rent at the NUTS 1 level which covers 21 industries and 13 regions.**

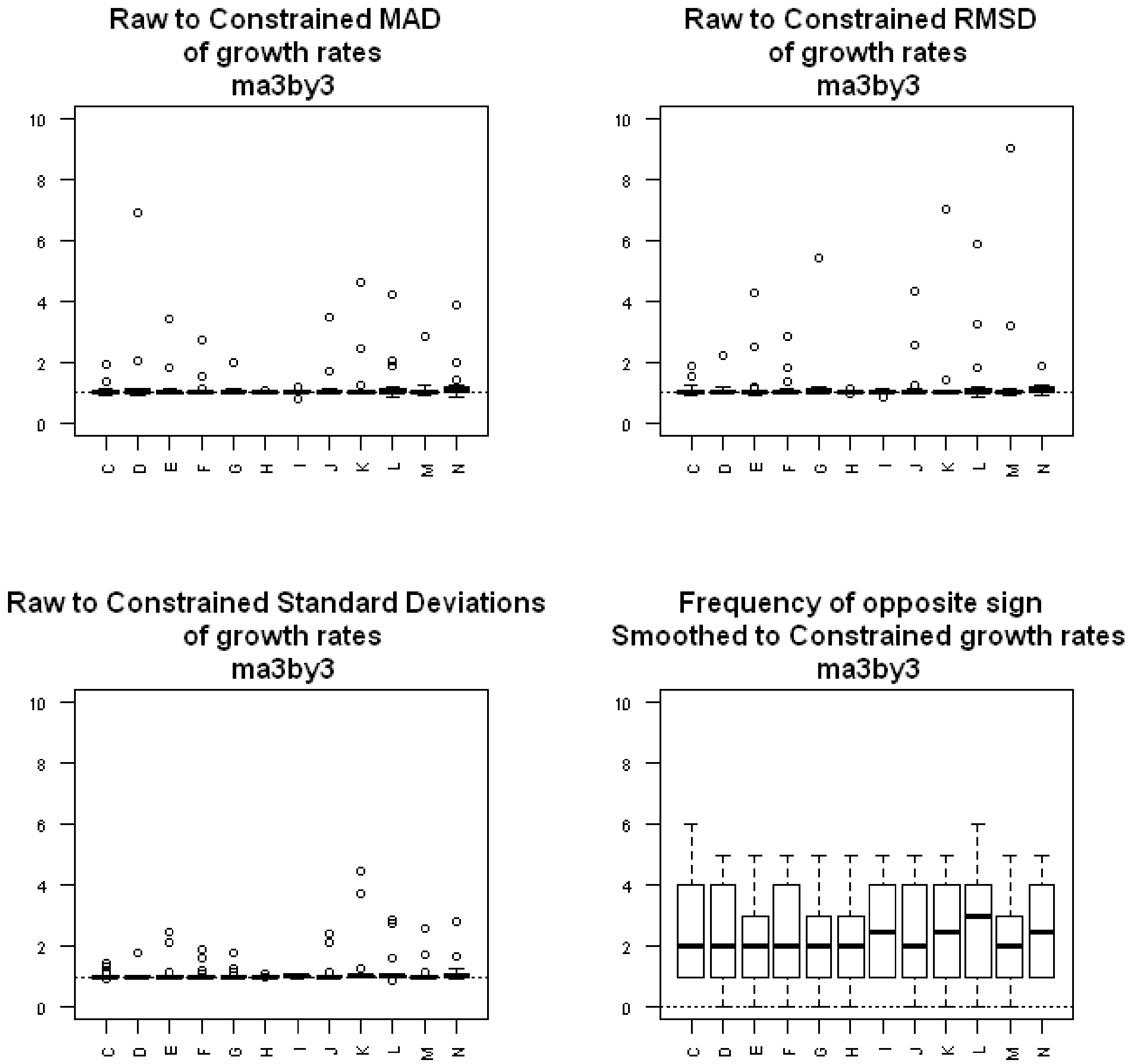


**Chart 11: Box plots of the ratios of smoothness measures for the component Mixed Income at the NUTS 1 level which covers 21 industries and 12 regions.**

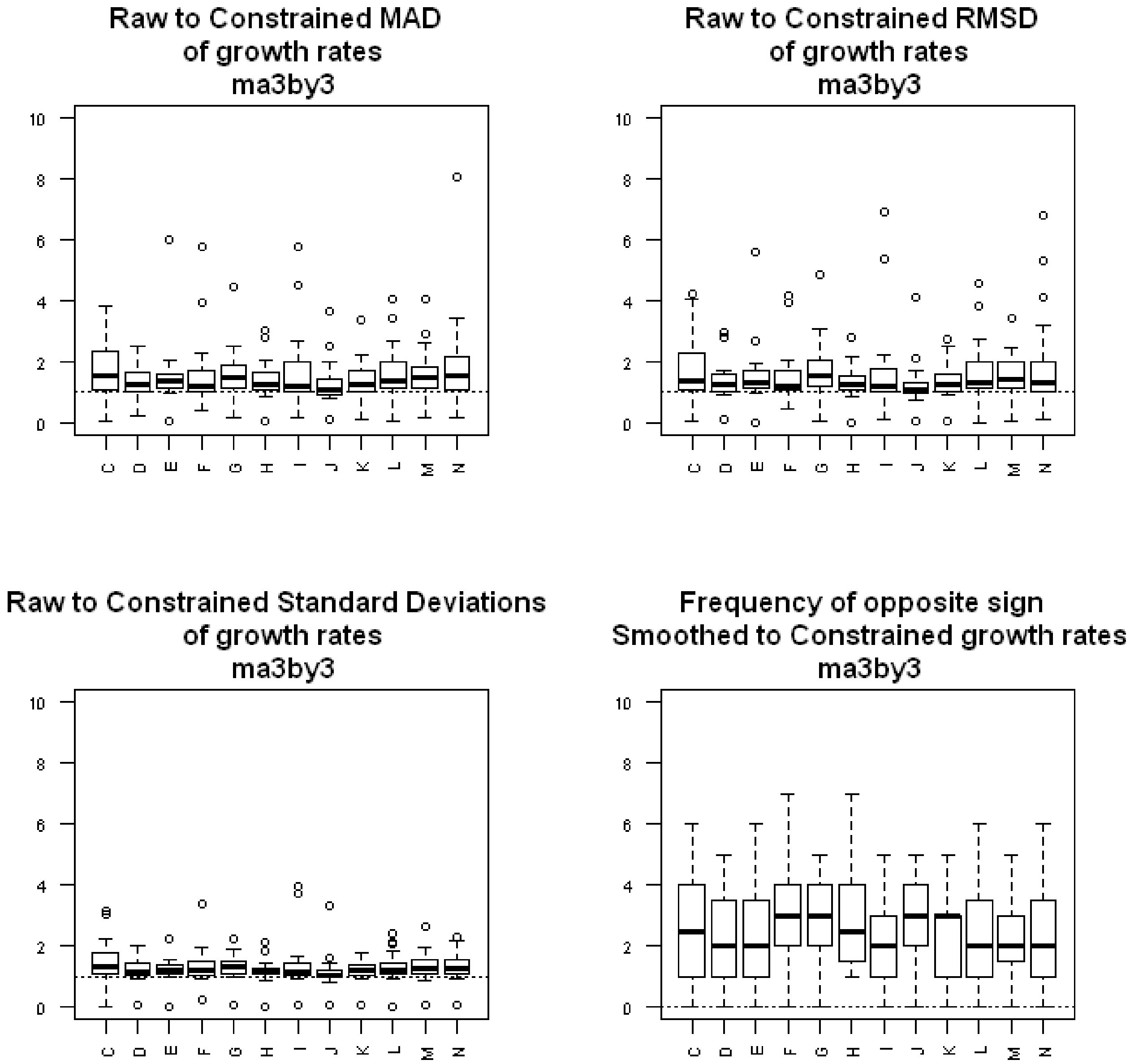


**Chart 12: Box plots of the ratios of smoothness measures for the component Total Gross Trading Profit & Gross Trading Surplus at the NUTS 1 level which covers 21 industries and 13 regions.**

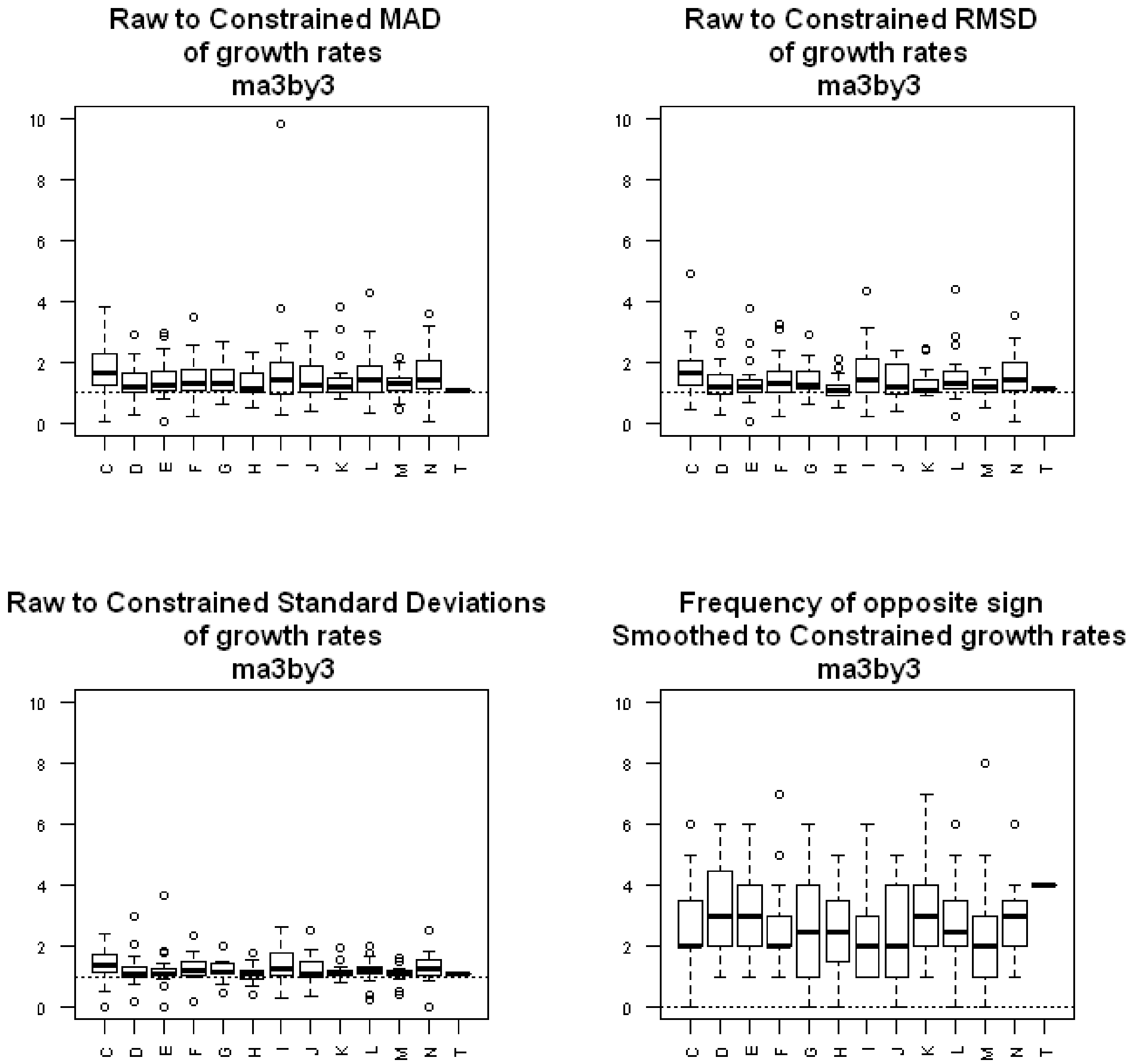
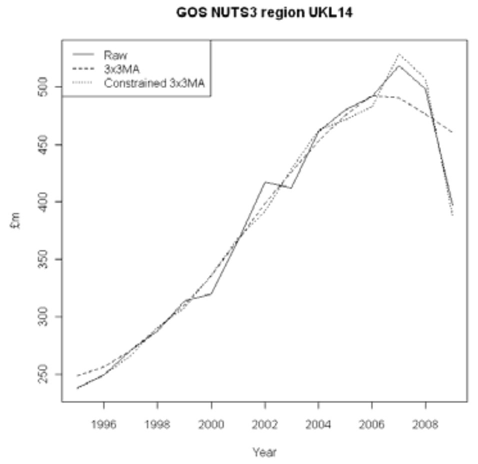
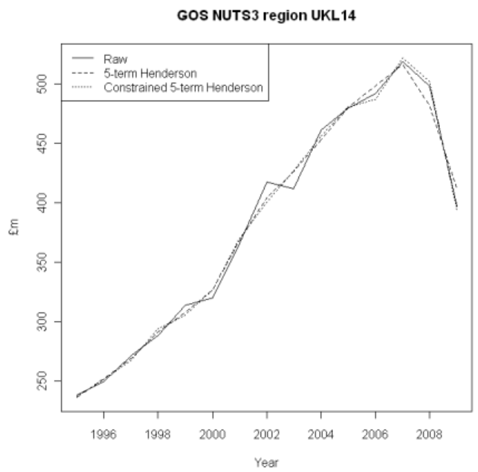


**Chart 13: Raw, smoothed and constrained data for Gross Operating Surplus for NUTS3 region UKL14 (South West Wales). Top panel 3x3 MA, centre panel 5-term Henderson filter, bottom panel kernel smoothing using parameters described above.**



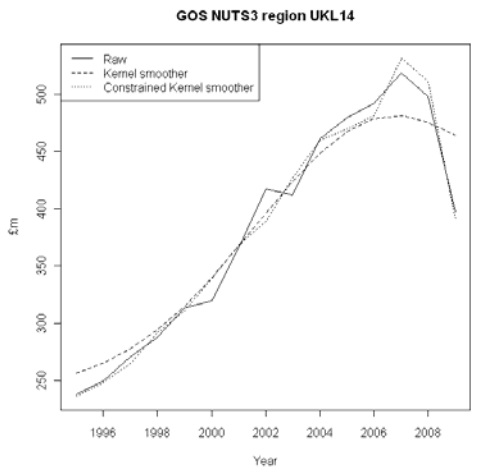



**Chart 14: Raw, smoothed and constrained data for GDHI for NUTS3 region UKD61 (Warrington). Left panel 3x3 MA, centre panel 5-term Henderson filter, right panel kernel smoothing using parameters described above.**

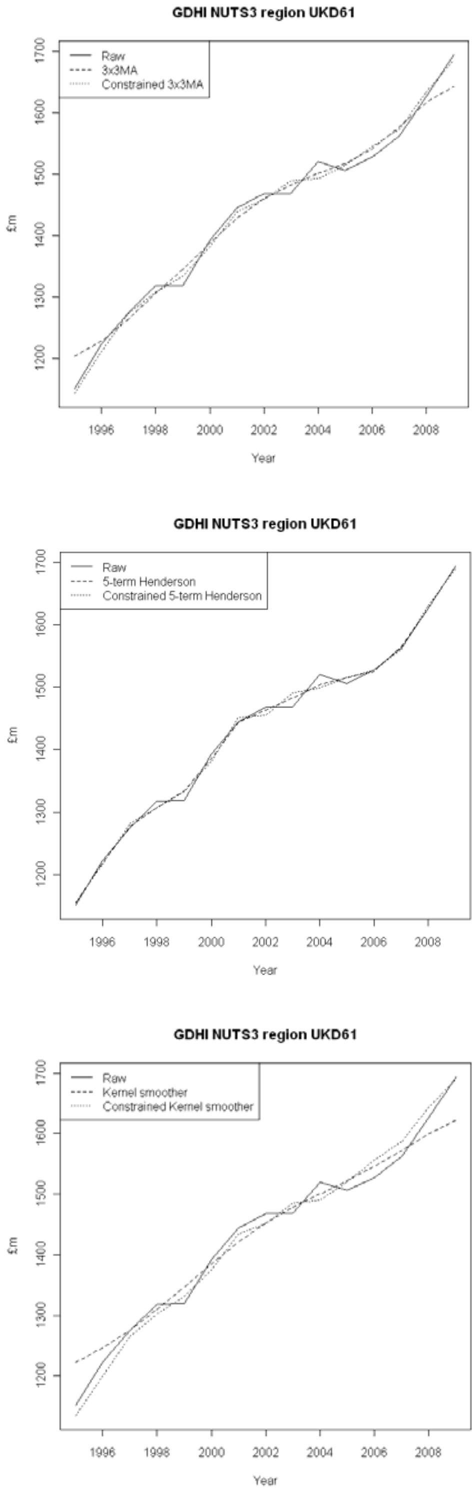


**Chart 15: Raw, smoothed and constrained data for GDHI for NUTS1 region UKD (North West) indirect estimates (sum of NUTS3 level data). Left panel 3x3 MA, centre panel 5-term Henderson filter, right panel kernel smoothing using parameters described above**



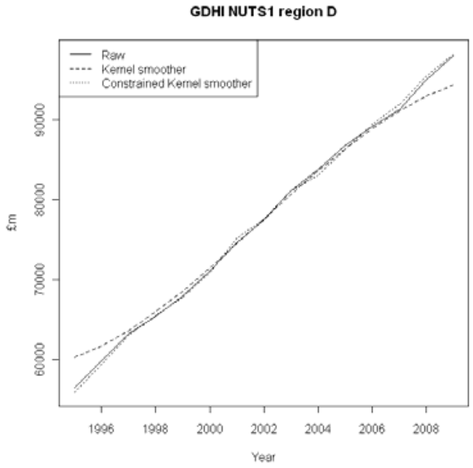


**Chart 16: Raw, smoothed and constrained data for GDHI for NUTS1 region UKD (North West) direct estimates (sum of NUTS3 level data). Left panel 3x3 MA, centre panel 5-term Henderson filter, right panel kernel smoothing using parameters described above**

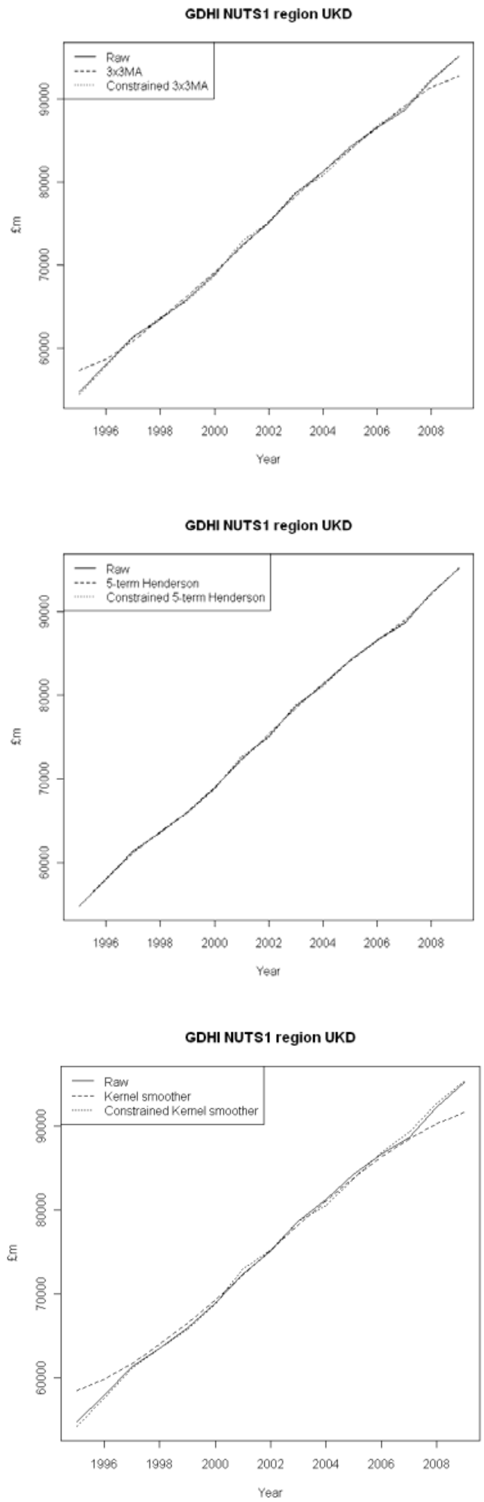


**Chart 17: Plots of the ratios of smoothness measures for the component Gross Operating Surplus at the NUTS 3 against 133 regions**