

# **Are United Kingdom expenditure data well-behaved?**

**Graeme Chamberlin<sup>1</sup>**

## **Abstract**

Data users are aware that early estimates of Gross Domestic Product are likely to be revised as later vintages of the same data are published. This article analyses revisions to the expenditure measure of GDP and its main components in the UK, using a longer sample than previous studies which incorporates a number of economic cycles. I find that the revisions process is affected by cyclical factors and frequent structural breaks often resulting from the ad-hoc way important methodological changes are introduced to the National Accounts. Therefore care should be taken when using patterns and trends in past data to predict future revisions as empirical relationships tend to break down in longer samples.

**Mathematics Subject Classification:** 91B82

**Keywords:** Gross Domestic Product, revisions, bias, signal to noise, entropy, cointegration, efficient forecast hypothesis, data rationality

---

<sup>1</sup> London South Bank University. E-mail: [chambeg4@lsbu.ac.uk](mailto:chambeg4@lsbu.ac.uk)

## 1 Introduction

The publication of data on Gross Domestic Product (GDP) is a process rather than an event. In the UK the first estimate of GDP is published by the Office for National Statistics (ONS) approximately 23 days after the end of the reference quarter. However, this estimate may be subject to revision in future data releases. It is now generally accepted revisions to GDP data are not usually about correcting errors or mistakes but about updating the figures as more information becomes available. For instance, initial data from quarterly surveys of household and business activities may be later benchmarked to larger and more complete information from annual surveys and administrative data sources such as tax returns. In the UK it is thought that data-driven revisions generally come to an end about 18-24 months after the reference quarter. GDP data will continue to be revised as the ONS introduce new methods and sources to better measure the evolving economy. As methodological changes are open-ended there is in essence no such thing as 'final' data.

Although revisions are now fairly well understood by data users the trade-off between timeliness and accuracy first described by McNees [1] still presents a challenge. In the past macroeconomists had assumed data revisions were small and random, so did nothing more than add some variance to model estimates and not affect the underlying results. More recent analysis of real time data, which shows snapshots of data published at specific points in time, suggests this to be false. Changes to the underlying data can in fact have important consequences for econometric modeling and forecasting.

Croushore and Stark [2] show the key findings in a number of well known and oft-cited published articles can change when models are re-estimated on revised data. Fackler [3] suggests that many researchers, if they were honest with themselves, would admit that previous research when re-estimated on a later vintage of data bore different results. Croushore and Stark [4] find data revisions can affect

forecast models by changing the underlying data, changing the model coefficients or changing the model specifications such as lag lengths and variable selection. A researcher who estimates a new forecast model on revised data may find it relatively easy to improve on a previously estimated model, but it should be acknowledged that the previous model may have been the best given the data available at the time. Kozicki [5] finds the ranking of different forecast models may depend on the data vintage used. Therefore Elliot [6] suggests the econometrician should use real time data when evaluating different models, and Koenig, Dolmas and Piger [7] suggest a number of different data vintages are used when building new forecast approaches. In the next section of this paper I present a number of examples of where UK GDP data has been revised in a way that could have affected the view of the UK economy held by policymakers and other data users.

The study of real time data and the effect of revisions on the understanding of the economy has become a vibrant research area and Croushore [8] provides a good survey of recent findings. This work has been aided by the increasing availability of real time databases for the key macroeconomic time series variables in a number of countries. Croushore and Stark [9] describe the US real time dataset for macroeconomists based on cooperation of Federal Reserve Bank of Philadelphia and University of Richmond. In the UK the Bank of England has constructed a comprehensive real time dataset following earlier work by Castle and Ellis [10] and Eggington, Pick and Vahey [11].

This paper uses an analysis of real time data to assess the behaviour, good or otherwise, of the expenditure measure of UK gross domestic product (GDP(E)) and its main components (private consumption, government consumption, total investment, exports and imports). I have constructed a real time data set which provides a monthly snapshot of each of these variables from 1975Q2 to 2013Q4. The framework for assessing the behaviour of this data is motivated by Boragan Arouba [12] who looks to see if the data satisfy three statistical properties.

The first test is for unbiasedness. This is where the expected value of revisions is

zero. If this test fails then it implies that data revisions are on average revised in a certain direction, and can be reduced through the simple application of a bias adjustment. In section 3 I look at the average revisions for GDP(E) and its main components over time and assess whether a bias adjustment could improve revisions performance.

The second test is for a low variance of revisions. If this test fails, and the variance of revisions is high relative to the variance of the underlying time series, it implies that early data vintages provide a poor signal of the more mature data. In section 4 I look at information windows, which are based on the signal to noise ratio, and show how quickly an accurate view of the data becomes available to data users. In section 5 I look at the entropy of the revisions process which shows how a measure of uncertainty changes as the data matures.

The third test is for the efficiency of revisions. Early data vintages could be thought of a forecast of later data vintages, or a 'nowcast' as the aim is to make an accurate prediction of the present. If subsequent revisions are predictable given the information available at the time early data vintages are published it suggests that these early data vintages are inefficient forecasts of more mature data. Making use of that information in early data vintages could reduce subsequent revisions.

In section 6, I test whether different data vintages are cointegrated, that is whether they share the same stochastic trend. Failure to find cointegration between the different data vintages would imply additional information would be required to predict more mature data given early or preliminary data vintages. In section 7 I look at the statistical evidence concerning whether revisions can be described as 'news' or 'noise'. Finally, in section 8 I test the efficiency of first published data by seeing if subsequent revisions are predictable given a set of business and consumer survey data and financial market indicators available in real time.

An important consideration in this analysis is the choice of 'final' data. This is the data vintage that earlier vintages are compared with in the above assessments.

Many researchers use the most recently published data but, following Boragan Arouba [12], this is not always the best data to use as methodological changes and benchmark revisions made many years later are unlikely to contain much information relevant to past data and could actually distort features of the time series. Therefore I use data published five years or 60 months after the first published estimate as the final data, but also look at data published two years or 24 months after the first vintage as this is an important milestone in the maturity of UK economic time series.

Boragan Arouba [12] concludes US data are not well-behaved, failing each of his three tests. My work though is based on a longer time series and importantly includes a number of economic cycles and has led to different conclusions. I find that data revisions are subject to important cyclical influences and a number of structural breaks over time. Therefore, the measurement of bias and the predictability of revisions can also change over time, making the practical implementation of bias adjustments and the estimation of better nowcasts unreliable. There is however evidence that early data vintages do provide a good signal of the mature data and that the signal has improved over time as the economy has become less volatile. It is also the case that non data-driven revisions such as methodological and benchmark revisions have a significant impact on total revisions, but these are not introduced into official statistics in a systematic way so are very difficult to predict. Therefore care should be taken by those who attempt to use past revisions performance to predict future revisions as the revisions process can be unstable in longer samples.

## **2 Some notable revisions to UK GDP(E)**

In this section, I briefly present four recent cases of where revisions to GDP(E) data could have had an important impact on the reading of the economy and contributed to mistakes in economic policy.

## 2.1 The late 1980s boom missing in the early data

In March 1987 the first estimate for growth in 1986 was recorded at 2.4%. However, by September 1993 the estimate had been revised to over 4%. As Figure 1 shows, annual growth in 1987 and 1988 was also subject to strong upward revisions in data published several years later. Growth in 1987 was initially recorded at 3.6% and was later revised to 4.8%, while growth in 1988 was revised up sharply from 2.6% to 5%. Taken together the economy was much stronger through the mid to late 1980s than early data suggested.

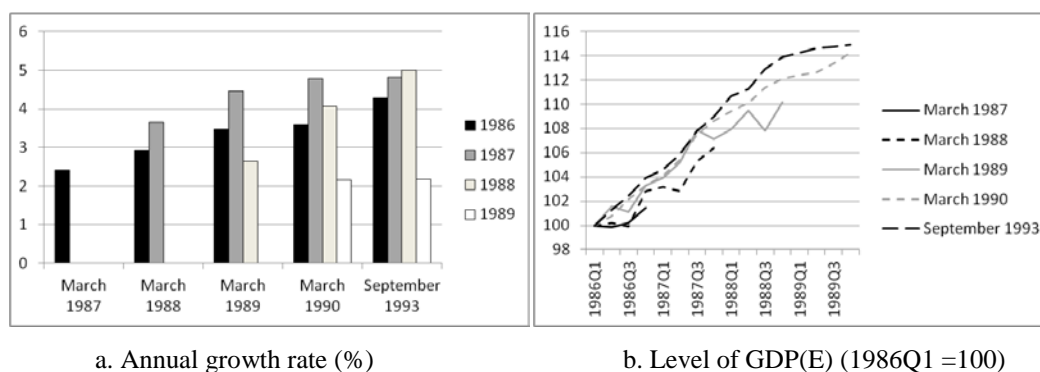


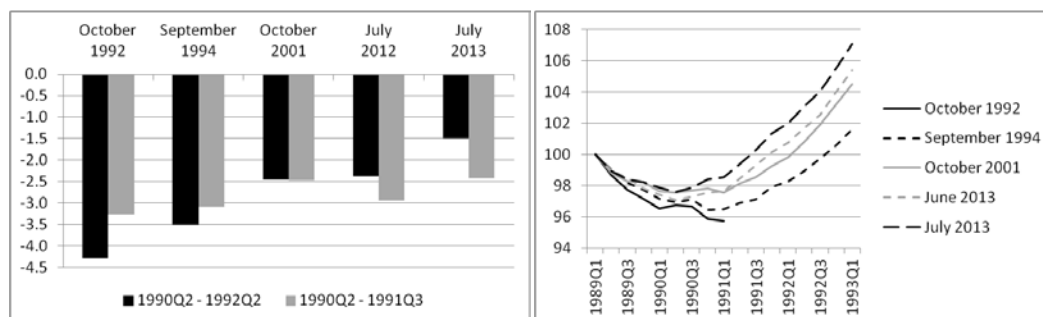
Figure 1: Revisions to GDP(E) growth in the late 1980s

The then Chancellor of Exchequer, Nigel Lawson, would blame these data revisions for the mistakes in economic policy that contributed to the boom and bust in the UK economy. He argued that had the true strength on the economy been known earlier fiscal policy would not have been loosened to the extent it was. Nelson and Nicolov [13] conclude the high inflation at the end of the decade could be traced back to the misreading of the output gap, with the weaker early vintage data giving the impression of more spare capacity in the economy. This was a low point for National Accounting in the UK and resulted in the Pickford Review [14] aimed at improving the quality of economic statistics.

## 2.2 Where did the 1990s recession go?

The UK recession in the early 1990s was associated with a sharp increase in interest rates, a rise in the number of unemployed to above 3 million and the deflation of the house price bubble. In October 1992 the first estimate of the peak to trough fall in GDP(E) between 1990Q2 and 1992Q2 was 4.3%.

Later estimates were to suggest the severity of the recession was less than initially measured. In September 1994 the peak to trough fall had been revised down to 3.6%, partly reflecting a rebasing of the National Accounts data from 1985 prices to 1990 prices. By October 1998, when the National Accounts had been rebased to 1995 prices the peak to trough fall was 2.8%. In October 2001 the trough date was brought forward to 1991Q3 with the total fall in GDP(E) then measured at 2.5%. Real time estimates of the peak to trough fall in GDP(E) are shown in Figure 2.



a. Peak to trough fall in GDP(E) (%)

b. Level of GDP(E) (1989Q1 =100)

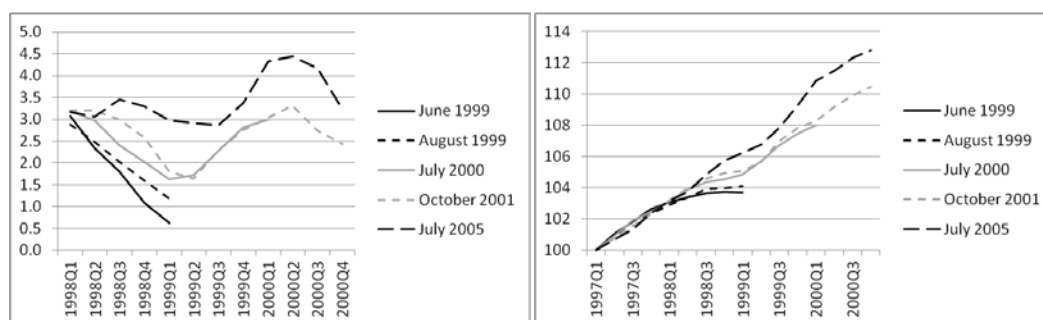
Figure 2: The early 1990s recession

This episode may confirm the suspicion raised by Borogan Arouba [12] that benchmark revisions can alter history in a way that does not necessarily improve data accuracy. Here it looks like the rebasing of the National Accounts to later price indices has smoothed out the depth of the recession. This of course can have important implications for subsequent analysis, for example in comparisons of this recession against that which followed the financial crisis in 2008.

### 2.3 The slowdown that wasn't – the late 1990s

Data published in late 1998 showed UK economic growth slowing abruptly. This was in line with business survey data and consistent with the uncertainty reeked on the global economy by the Asian financial crisis. Data published in June 1999 showed the annual growth rate in GDP(E) had fallen to 0.5%.

This prognosis of the economy was to change quickly (see Figure 3). The following year growth between 1998Q1 and 1999Q1 was recorded at a healthier 2%. By July 2005 it was estimated the economy had actually grown by a buoyant 3% over this period. Growth in the following years was also revised significantly upwards, so far from entering a soft patch the UK economy was actually booming. Early vintages of data had failed to pick up the rapid growth in private consumption that had been supported by a cut in interest rates in 1998 along with the strong housing and labour markets.



a. Four-quarter growth rate (%)

b. Level of GDP(E) (1997Q1 = 100)

Figure 3: Growth in GDP(E), 1998-2000

### 2.4 The start of the 2008 recession

Economists have been widely criticized for failing to see the financial crisis, but it also appears statisticians were slow to pick up on the downturn in the economy. Figure 4 shows a number of different vintages of growth rates for 2008Q2. In August 2008 the first estimate recorded growth at 0.2%, yet four years later this had



been revised to a sizeable contraction of -1.0%. The preliminary data had suggested that the UK economy would experience a slowdown in growth but not enter recession.

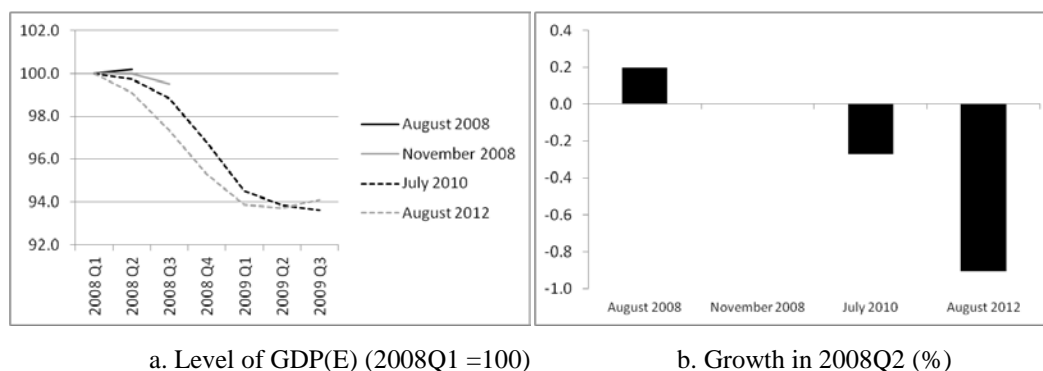


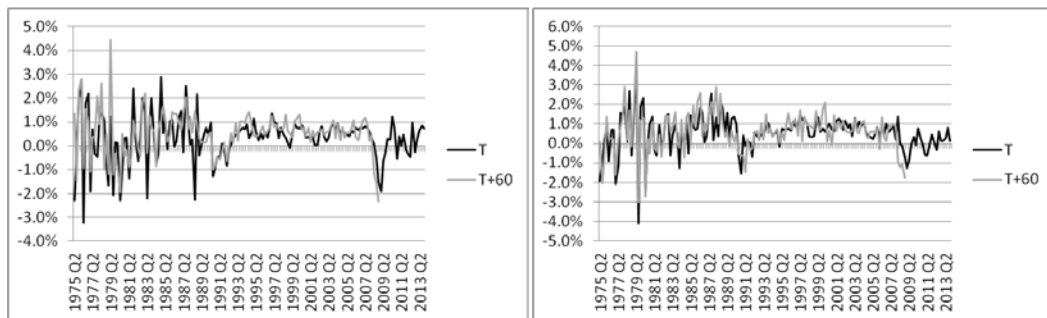
Figure 4: The start of the recession in 2008

It is a general trait of economic statistics that they are slow to record turning points in the economy. Croushore [8] finds that the largest revisions typically occur around turning points which are when policymakers need the data to be both timely and accurate. He shows that in the US annualized growth in the final quarter of 2008 was revised down from an initial -3.8% to -6.2%. Dynan and Elmendorf [15] find early estimates of GDP are unreliable around turning points, showing that revisions are typically larger when economic growth accelerates and decelerates. Fixler and Grimm [16] report that preliminary data in the US tend to do a reliable job at signaling cyclical peaks in GDP, but less so at troughs by tending to overstate declines and understate the beginning of a recovery. Likewise, Swanson and van Dijk [17] report a significant increase in the volatility of revisions around turning points in a number of economies.

### 3 Average revisions to GDP(E) and its expenditure components

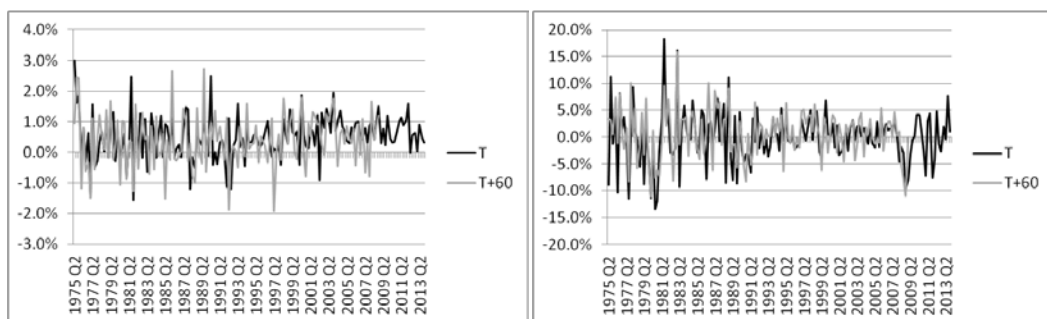
#### 3.1 Total revisions between first published data and those five years later

I define a particular vintage of a macroeconomic times series by  $y_t^i$  where  $y_t$  is the times series for a certain macroeconomic variable and  $i$  the  $T+i$  monthly vintage. In this case the first published data vintage is  $y_t^0$  and the data published five years or 60 months later is  $y_t^{60}$ . The quarterly growth rates of the first published data vintage and that five years later are  $\Delta y_t^0$  and  $\Delta y_t^{60}$  respectively and these two data vintages (marked as T and T+60) for GDP(E) and the five main expenditure components are plotted in Figure 5.



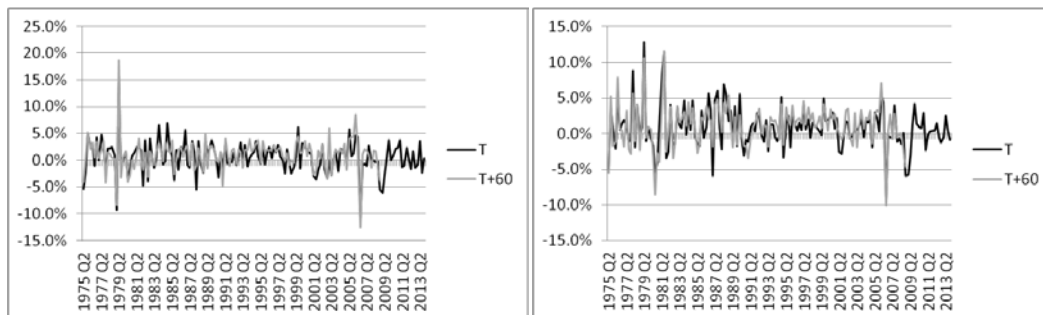
a. GDP(E)

b. Private consumption



c. Government consumption

d. Total investment



e. Exports

f. Imports

Figure 5: Quarter on quarter growth rates: 1975Q2–2013Q4 (%)

There are two features of the data presented in Figure 5 worth noting. First, the quarter on quarter growth rate data is quite volatile, especially in the 1970s and 1980s. From the 1990s onwards the macroeconomic data becomes less volatile reflecting greater stability and lower inflation in the UK economy. Second, although revisions can often be significant, when looking at a long span of the data the first published estimate appears to be a generally good guide to where the data ends up five years later.

### 3.2 Average revisions

The total revision between the first published estimate of the quarter on quarter growth rate and a given vintage  $T+i$  is given by:

$$R_t^{i,0} = \Delta y_t^i - \Delta y_t^0$$

In this section I look at the path of average revisions from the first published data to the  $T+24$  ( $R_t^{24,0}$ ) and  $T+60$  ( $R_t^{60,0}$ ) vintages.

The mean average revision in a given sample of  $n$  observations is simply:

$$MAR_i = \frac{1}{n} \sum_{t=1}^n R_t^{i,0}$$

I study the average of revisions for GDP(E) and its main components over a rolling

five year window (n=60). These are shown in Figure 6, note that the rolling average is centred so the average is plotted on the mid-point of the five year rolling sample.

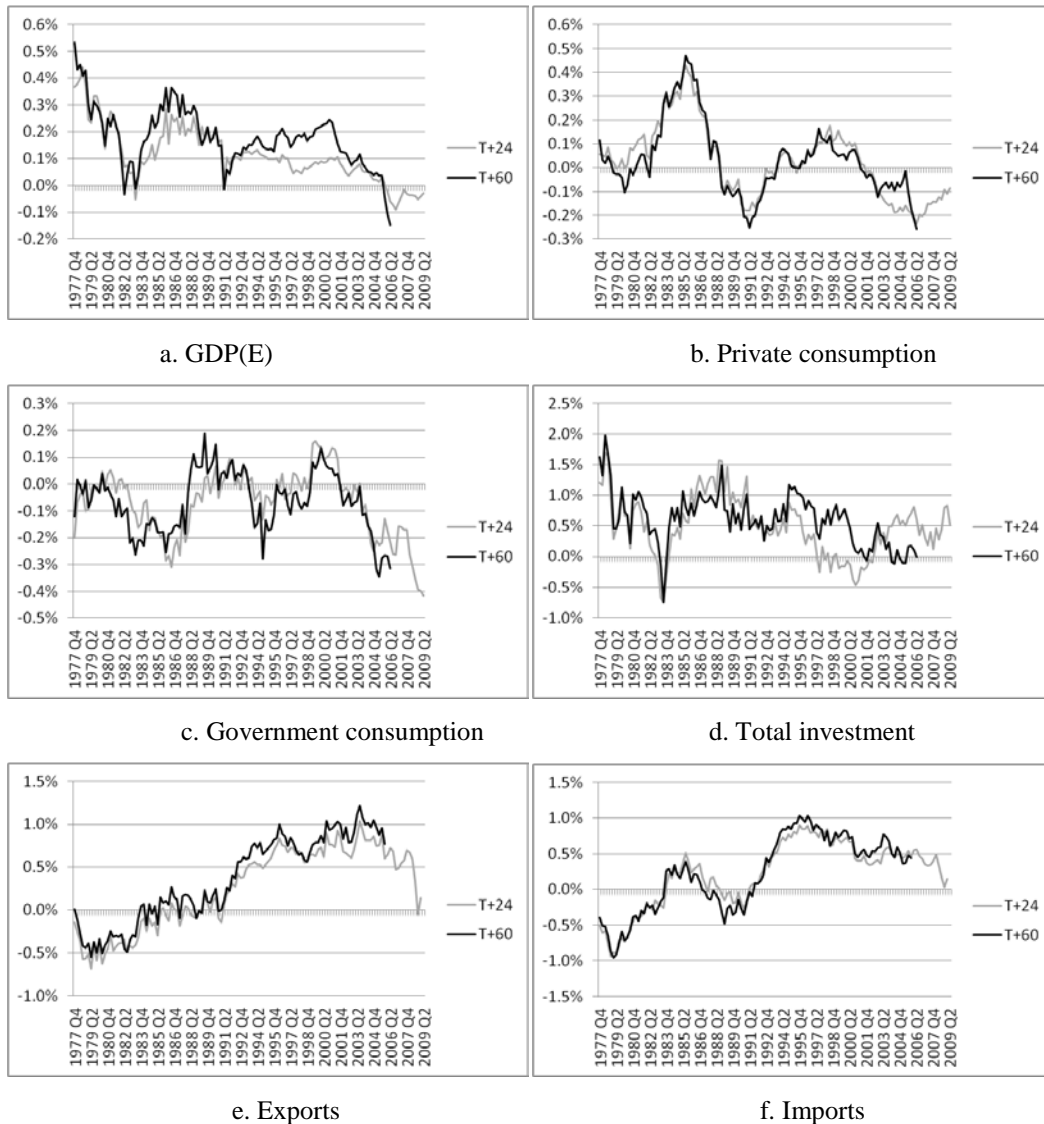


Figure 6: Average revisions in a rolling five year sample (% point change to the quarter on quarter growth rate)

Figure 6 makes clear that average revisions are unstable over time. There is

some evidence of a cyclical pattern, with average revisions more likely to be positive in periods of stronger economic growth and negative or smaller in periods of slow growth or recession. This finding is similar to that reported in Castle and Ellis [10] who also look at a rolling average of revisions to GDP(E) in the UK.

Second, the revisions process itself is subject to a number of structural breaks reflecting changes in the underlying economy and the way GDP(E) is measured. For instance, data used to be benchmarked to a new price index every 5 years but since 2003 the weights are now updated each year through annual chain linking. Methodology changes are also introduced in an ad hoc way so the timing and impact on time series data is typically hard to predict.

### **3.3 Should we use bias adjustments?**

If early data vintages were revised in a consistent way then average revisions to mature data could be reduced by adding or subtracting a bias adjustment. During the Great Moderation period (1993-2008) GDP data was consistently revised upwards and the Bank of England, in forming their assessment of the UK economy, implicitly applied an adjustment based on the history of recent revisions to the official GDP data (see Cunningham, Eklund, Jeffery, Kapetanios and Labhard [18]).

During the 1980s the UK Central Statistics Office officially made a bias adjustment to correct early estimates of the Index of Manufacturing. The data was largely collected from a panel of the same large engineering firms so under-represented new and fast growing companies and led to positive bias in the statistics. The adjustment was removed when the sample design was improved and sample size increased.

This example highlights a problem where data collection in official surveys may lag changes in the economy. These changes, such as the growth and demise of certain industries, are often likely to be driven by the cyclical path of the economy, so it is unsurprising that the direction of revisions can too vary with the economic

cycle. Therefore bias adjustments that are based on backward moving averages (such as the Bank of England in Cunningham et al [18]) run the risk of over-adjusting should the economy move through a turning point as the UK economy did in 2008. The calculated bias adjustment may also be affected by large, unique, and seldom repeated revisions from the recent past. For example, large revisions to past data from the introduction of annual chain-linking were unlikely to be repeated in the future meaning a bias adjustment might over-adjust in subsequent years. Revisions to GDP(E) are generally positive but the very uneven path of average revisions seen in Figure 6 could make the practical implementation of bias adjustments difficult.

## 4 Quantifying the quality of macroeconomic data using information windows

It is difficult to attach a metric to the quality of data and most national statistics institutions generally refrain from trying. Oller and Teterukovsky [19] and subsequently Kholodilin and Siliverstovs [20] put data revisions at the heart of the quality measurement by asking how quickly accurate information becomes available to data users. They show this graphically using an information window which indicates how the signal to noise ratio changes with the publication of each successive data vintage.

The first step in the construction of the information window is to identify a ‘mature’ data vintage which is considered to be an accurate measure of economic activity. As set out in the introduction this paper uses the T+60 vintage for this. Next the mean squared revision (MSR) of each data vintage T+i relative to the T+60 vintage, starting with the first published data i=0, is calculated:

$$MSR_i = \frac{1}{n} \sum_{t=1}^n (\Delta y_t^i - \Delta y_t^{60})^2$$

Clearly the MSR of i=60 would be equal to zero and we would expect the MSR

to converge towards this with each successive data vintage. The next step is to standardize the MSR by scaling with the variance of the T+60 data vintage so that:

$$\overline{MSR}_i = \frac{MSR_i}{var(\Delta y_t^{60})}$$

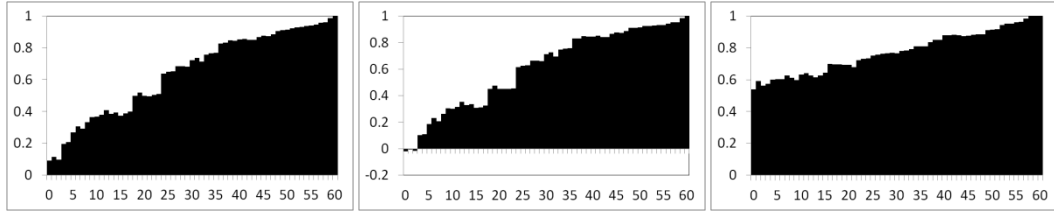
This takes into account the underlying volatility in the data when judging the size of revisions. Finally the signal to noise ratio of each data vintage is given by 1 minus the standardized MSR.

$$SNR_i = 1 - \overline{MSR}_i$$

The information window is then just a plot of the  $SNR_i$  for  $i=0, \dots, 60$ . Clearly the smaller the MSR the larger the SNR, meaning that data vintage provides a more accurate portrayal of the T+60 vintage. The signal to noise ratio should converge towards a value of unity but data users will be interested just how quickly this happens and we can answer this by looking at the information window. It is possible for the standardized MSR to exceed one and for the SNR to be negative. This would imply that data vintage to be a useless measure, and in this case the data user would be better off using the unconditional mean of the past data series rather than that particular vintage.

The information windows for GDP(E) and the main expenditure components are presented in Figure 7. I have also split the sample into two sub-samples reflecting the highly volatile period up to the early 1990s and then the relatively low volatile period from there onwards. Table 1 is a summary table recording the vintage when the SNR reaches the 0.5, 0.7 and 0.9 levels.

There are a number of observations that can be made from Figure 7 and Table 1. First the SNR of early data vintages of GDP(E) improves considerably in the second sub-period reflecting a significant fall in mean squared revisions. Closer inspection suggests that the improvement in the GDP(E) signal to noise ratio lies in the better quality of the total investment data. Private consumption and government consumption data quality actually deteriorated in the second sub-sample, but because total investment is a far more volatile series its improvement is more highly weighted.

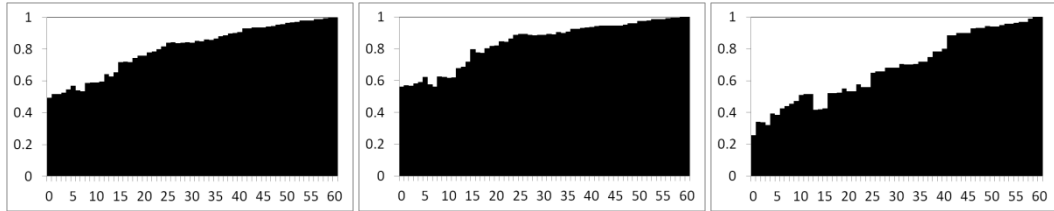


i) 1975Q2 – 2008Q4

ii) 1975Q2 – 1992Q4

iii) 1993Q1 – 2008Q4

a. GDP(E)

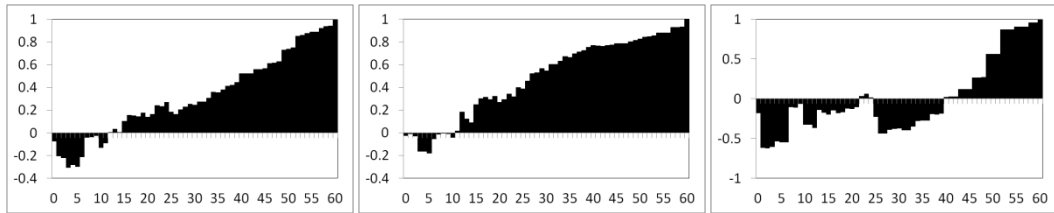


i) 1975Q2 – 2008Q4

ii) 1975Q2 – 1992Q4

iii) 1993Q1 – 2008Q4

b. Private consumption

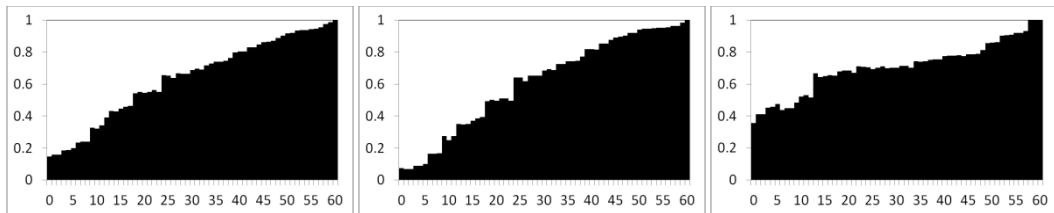


i) 1975Q2 – 2008Q4

ii) 1975Q2 – 1992Q4

iii) 1993Q1 – 2008Q4

c. Government consumption

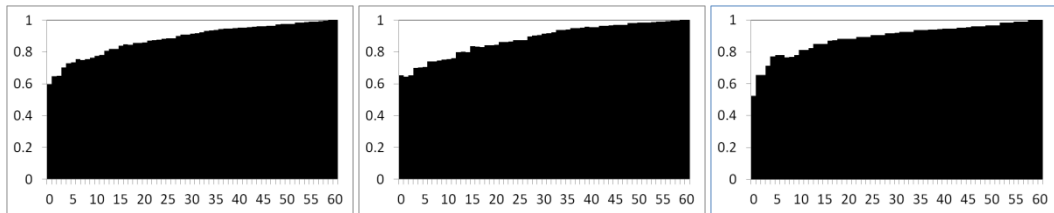


i) 1975Q2 – 2008Q4

ii) 1975Q2 – 1992Q4

iii) 1993Q1 – 2008Q4

d. Total investment



i) 1975Q2 – 2008Q4

ii) 1975Q2 – 1992Q4

iii) 1993Q1 – 2008Q4

e. Exports



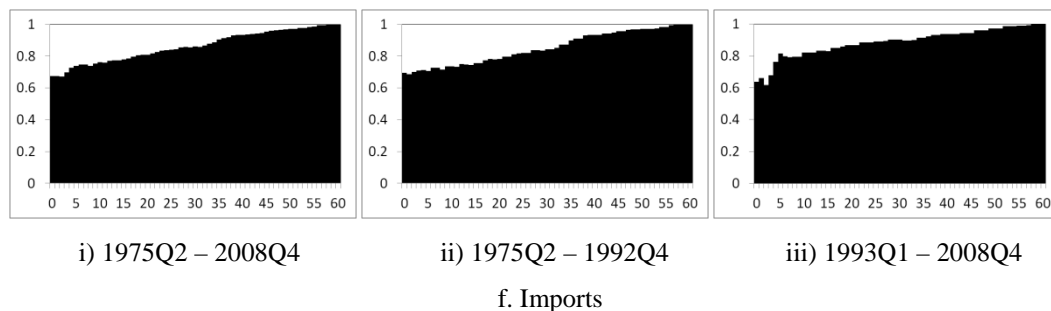


Figure 7: Information windows

Government consumption data is generally poor. In the second sub-sample the SNR is predominately below zero for the first three years and does not reach 0.5 until the 49th vintage. Kholodilin and Siliverstovs [20] also find early government spending vintages to be the least reliable of the expenditure components, largely because of the time it takes in publishing full government accounts. On the other hand the trade data reports a high SNR for early data vintages in both sub-samples which is an indicator of high data quality.

The analysis of information windows suggests for GDP(E) overall early data vintages give a fairly good signal of the mature data, and that this has improved in recent years as the lower volatility of the economy has made the economy easier to measure and lowered revisions. For the individual expenditure components the signal to noise ratio may often be lower for earlier vintages than GDP(E) overall. This reflects the impact of revisions which change the composition of GDP(E) rather than the overall level or growth rate.

Table 1: Information windows: vintage when the SNR reaches a given level

		0.5	0.7	0.9
GDP(E)	1975Q2 to 2008Q4	19	33	48
	1975Q2 to 1992Q4	24	33	48
	1993Q1 to 2008Q4	0	22	49
Private consumption	1975Q2 to 2008Q4	1	15	39
	1975Q2 to 1992Q4	0	14	35
	1993Q1 to 2008Q4	16	31	43
Government consumption	1975Q2 to 2008Q4	40	49	57
	1975Q2 to 1992Q4	27	37	57
	1993Q1 to 2008Q4	49	52	55
Total investment	1975Q2 to 2008Q4	18	33	49
	1975Q2 to 1992Q4	21	33	47
	1993Q1 to 2008Q4	10	22	52
Exports	1975Q2 to 2008Q4	0	3	27
	1975Q2 to 1992Q4	0	3	28
	1993Q1 to 2008Q4	0	3	25
Imports	1975Q2 to 2008Q4	0	4	35
	1975Q2 to 1992Q4	0	2	36
	1993Q1 to 2008Q4	0	4	34

## 5 Entropy and the information gain of revisions

Entropy is a concept in thermodynamics used to describe the degree of disorder in a system but in the statistical world it is often used as a measure of data uncertainty. In the analysis of revisions Patterson and Heravi [21] look at the information gain between different data vintages which is simply the difference in entropies. It therefore provides a measure of the reduction in uncertainty as data matures. The reduction in uncertainty is once again relative to a mature data vintage which, as before, is taken as the T+60 vintage. I consider three different but closely

related measures of entropy to calculate the information gain between different data vintages.

The first measure is based on the entropy of the normal distribution.

$$H(\Delta y_t^{60} | \Delta y_t^i) = 0.5 \ln(\exp 2\pi\sigma_i^2)$$

where the variance is based on the revision between the T+60 (final) vintage and vintage T+i.

$$\sigma_i^2 = \text{var}(\Delta y_t^{60} - \Delta y_t^i)$$

So the information gain between two vintages T+i and T+i+ $\delta$  is given by the reduction in entropy:

$$H(\Delta y_t^{60} | \Delta y_t^i) - H(\Delta y_t^{60} | \Delta y_t^{i+\delta}) = \ln(\sigma_i / \sigma_{i+\delta})$$

We would expect the later vintages to have a lower variance of revisions with the final data in which case the information gain should usually be, but not necessarily, positive between successive vintages.

The second measure takes into account the possible presence of non-zero means in the revision process by deducting the sample means from each data vintage. This acts to remove the effect of (squared) bias in the variance calculation so the measured variance of data revisions is a cleaner indicator of uncertainty.

$$\sigma_i^2 = \text{var}((\Delta y_t^{60} - \mu_{60}) - (\Delta y_t^i - \mu_i))$$

where  $\mu_i$  is the expected value of  $\Delta y_t^i$ . In this case the information gain between vintages T+i and T+i+ $\delta$  is:

$$H(\Delta y_t^{60} | \Delta y_t^i) - H(\Delta y_t^{60} | \Delta y_t^{i+\delta}) = 0.5 \ln[(\sigma_i^2 - \mu_i^2) / (\sigma_{i+\delta}^2 - \mu_{i+\delta}^2)]$$

The final measure of information gain relaxes the assumption of normality in the distribution of revisions. This requires a direct measure of entropy through an approximation of the probability density function and Beirlant, Dudewicz, Gyorfi and van der Meulen [22] provide an overview of the available methods. Here I follow Kholodilin and Siliverstovs [20] and Patterson and Heravi [21] in using the Vasicek [23] approach based on sample spacing.

If the cumulative density function is  $F(y) = p$  then entropy can be estimated by:

$$H = \int_0^1 \ln[dF^{-1}(P)/dp] dp$$

A consistent estimator of H (denoted  $H_{mn}$ ) can be constructed by putting the  $n$  data observations of revisions between the  $T+60$  and  $T+i$  vintages in ascending numerical order  $j=1,2,\dots,n$  so that  $R^{60,i}(1) < R^{60,i}(2) < \dots < R^{60,i}(n)$  and using the space between observations, known as the  $m$  space, as a rough estimate of the reciprocal of the probability density in that region (the derivative of  $F^{-1}(p)$ ).

$$H_{mn}(R^{60,i}) = n^{-1} \left[ \sum_{j \in I_1} \ln(R^{60,i}_{(j+m)} - R^{60,i}_{(j-m)})/2m \right. \\ \left. + \sum_{j \in I_2} \ln(R^{60,i}_{(j+m)} - R^{60,i}_{(1)})/(j+m-1) \right. \\ \left. + \sum_{j \in I_3} \ln(R^{60,i}_{(n)} - R^{60,i}_{(j-m)})/(n-j+m) \right]$$

Where:

$$j \in I_1 \quad j = m + 1 \dots \dots, n - m$$

$$j \in I_2 \quad j = 1 \dots \dots m$$

$$j \in I_3 \quad j = n - m + 1 \dots \dots n$$

The  $m$  space is a positive integer smaller than  $n/2$ . Noughabi and Arghami [24] suggest using an  $m$  space given by  $\sqrt{n}+0.5$ . Given the sample size of  $n=133$  this would suggest an  $m$  space of 12, but I choose to use an  $m$  space of 15 as this avoids having to calculate  $\ln(0)$  which can be a problem when calculating the entropy of revisions for later data vintages. This is why the information gain between later vintages cannot be calculated so I can only present the information gain up to vintage  $T+48$ . Therefore the information gain between two vintages  $T+i$  and  $T+i+\delta$  is again given by the change in entropy  $H_{mn}(R^{60,i}) - H_{mn}(R^{60,i+\delta})$ .

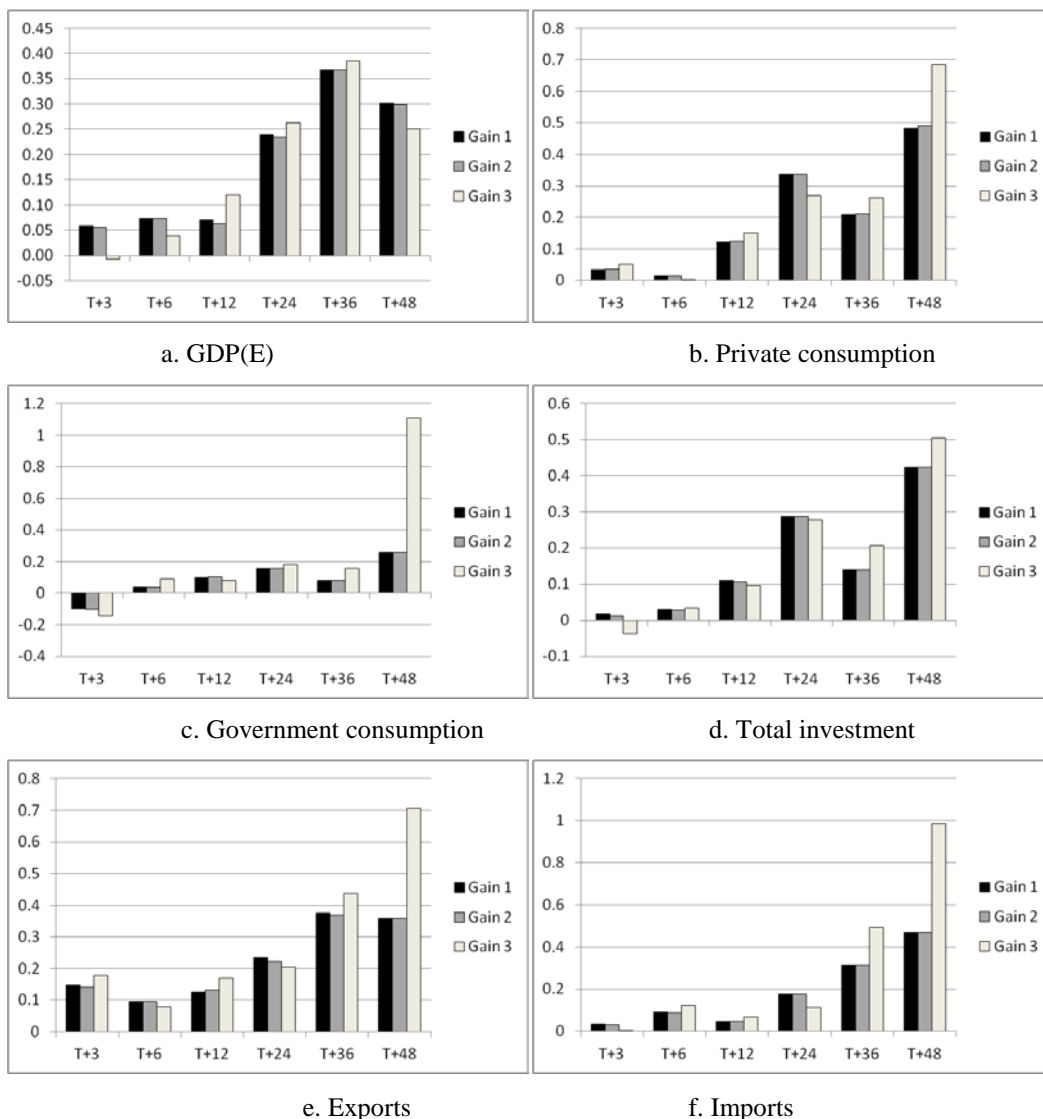


Figure 8: Information gain between successive vintages (starting at T=0)

The  $m$  space is a positive integer smaller than  $n/2$ . Noughabi and Arghami [24] suggest using an  $m$  space given by  $\sqrt{n} + 0.5$ . Given the sample size of  $n=133$  this would suggest an  $m$  space of 12, but I choose to use an  $m$  space of 15 as this avoids having to calculate  $\ln(0)$  which can be a problem when calculating the entropy of revisions for later data vintages. This is why the information gain between later vintages cannot be calculated so I can only present the information gain up to

vintage T+48. Therefore the information gain between two vintages T+i and T+i+ $\delta$  is again given by the change in entropy  $H_{mn}(R^{60,i}) - H_{mn}(R^{60,i+\delta})$ .

The three measures of information gain are presented in Figure 8 and a number of observations can be made. First it appears for most variables the largest information gains are made after the T+24 vintage which suggests methodological and benchmark changes are a larger source of uncertainty than data-driven revisions. Second, for GDP, private consumption and total investment the three different measures of information gain give broadly similar results which implies data revisions characterize a normal distribution. For government consumption, exports and imports there seems to be non-normality in the revisions and thus a particularly large information gain between the T+36 and T+48 vintages.

## 6 Cointegration analysis

### 6.1 Cointegration and the data measurement process

If different vintages of data were cointegrated it would be powerful evidence in support of a well-behaved data measurement process for GDP(E). Following Stock and Watson [25] if there are  $v$  data vintages in a vector autoregressive model (VAR) then  $r+s=v$  where  $r$  is the cointegrating rank of the system and  $s$  the number of independent stochastic trends. If  $r=v-1$  then  $s=1$ , which suggests the data vintages are driven by a single stochastic trend.

Therefore, Patterson and Heravi [26] suggest a failure to find cointegration between different data vintages has a number of implications. First, it might tell us that early data vintages are biased predictors of later data in terms of growth rates. Second, early data vintages cannot be viewed as an efficient forecast of later data vintages because other variables may be required to render the revisions process as a stationary or  $I(0)$  process. Third, care should be taken when using preliminary or early data in forecasting and modeling in place of more mature data because of the

presence of non-stationary or I(1) errors.

I use the procedure set out in Patterson [27], [28] to test for the presence of cointegration in four key data vintages: T, T+12, T+24 and T+60.

$$\mathbf{y}_t = [y_t^0 \quad y_t^{12} \quad y_t^{24} \quad y_t^{60}]'$$

More vintages could have been included but by increasing the dimension of the VAR model it would have limited the available degrees of freedom and make estimation more difficult.

The cointegrating framework is:

$$\Delta \mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Pi} \mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t$$

Cointegration implies:

$$\boldsymbol{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'$$

Where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are each  $4 \times r$  matrices each of rank  $r$  such that  $0 < r < 4$ . If  $r=0$  then no cointegration exists between the 4 data vintages and each variable is driven by its own separate stochastic trend. Alternatively, if  $r=4$ , then all the data is I(0) so no cointegrating relation exists and each variable is best represented by a simple dynamic model.

When  $0 < r < 4$  there are either 1 ( $r=3$ ), 2 ( $r=2$ ) or 3 ( $r=1$ ) common stochastic trends. Clearly the best outcome would be for  $r=3$  as it implies a single stochastic trend governs the set of data vintages. In this case Johanson and Juselius [29] state that there need to be 2 independent restrictions on each of the 3 cointegrating vectors giving a total of 6 restrictions for exact identification of the cointegrating framework.

The easiest way to do this is to set two elements to zero in each of the three rows of  $\boldsymbol{\beta}'$ . Also if the  $i$ -th column of  $\boldsymbol{\beta}$  on the  $i$ th vintage is normalized then we could have the following identifying restrictions:

$$\boldsymbol{\beta}' = \begin{bmatrix} 1 & \beta_{21} & 0 & 0 \\ 0 & 1 & \beta_{32} & 0 \\ 0 & 0 & 1 & \beta_{43} \end{bmatrix}$$

Here each vintage is pairwise cointegrated with the next vintage and if sequential

revisions are stationary then  $\beta_{21} = \beta_{32} = \beta_{43} = -1$ .

Or we could have each data vintage pairwise cointegrated with the final vintage T+60. In this case:

$$\boldsymbol{\beta}' = \begin{bmatrix} 1 & 0 & 0 & \beta_{41} \\ 0 & 1 & 0 & \beta_{42} \\ 0 & 0 & 1 & \beta_{43} \end{bmatrix}$$

Again revisions are stationary if  $\beta_{41} = \beta_{42} = \beta_{43} = -1$ .

Following Gonzalo and Granger [30], the system can be subject to a permanent-transitory (P-T) decomposition

$$\begin{aligned} \mathbf{y}_t &= \mathbf{y}_t^P + \mathbf{y}_t^T \\ \mathbf{y}_t^P &= \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}'_\perp \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}'_\perp \mathbf{y}_t \\ \mathbf{y}_t^T &= \boldsymbol{\alpha} (\boldsymbol{\beta}' \boldsymbol{\alpha})^{-1} \boldsymbol{\beta}' \mathbf{y}_t \end{aligned}$$

As pointed out by Patterson [28] a powerful result arises if we find the final vintage  $y_t^{60}$  is weakly exogenous through a block test on the  $\boldsymbol{\alpha}$  matrix.

This means:

$$\begin{aligned} \boldsymbol{\alpha}'_\perp &= [0 \quad 0 \quad 0 \quad 1] \\ \boldsymbol{\beta}'_\perp &= [1 \quad 1 \quad 1 \quad 1] \\ \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}'_\perp \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}'_\perp \mathbf{y}_t &= [y_t^{60} \quad y_t^{60} \quad y_t^{60} \quad y_t^{60}] \\ \boldsymbol{\alpha} (\boldsymbol{\beta}' \boldsymbol{\alpha})^{-1} \boldsymbol{\beta}' &= \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

In this case the common factor and permanent component is equal to the final data vintage and the transitory components are the 3 total revisions relative to the final vintage so the complete P-T composition is:

$$\mathbf{y}_t = [y_t^{60} \quad y_t^{60} \quad y_t^{60} \quad y_t^{60}] + [y_t^0 - y_t^{60} \quad y_t^{12} - y_t^{60} \quad y_t^{24} - y_t^{60} \quad 0]$$

Under these conditions there is a single stochastic trend and this can be identified as the final or mature data vintage. This could be accepted as a well-behaved data measurement process.

This article takes the following approach. First I conduct unit root tests on the four data vintages to confirm they are non-stationary or I(1). I then run an



unrestricted VAR model to confirm the lag length using information criterion. The Johansen trace and eigenvalue tests are used to find the cointegrating rank of the system and in this case  $r=3$  would imply one common factor and a well-behaved data measurement process. If this result is confirmed I can test identifying restrictions for sequential revisions, and then use a weak exogeneity test to see if the final vintage can be accepted as the common factor or stochastic trend.

## 6.2 Unit root tests

Table 2 shows the results of the Augmented Dickey Fuller (ADF) and Philips-Peron tests for unit roots in GDP(E) and the main expenditure components. These show that all data vintages for all variables are non-stationary I(1) variables which is a prerequisite as I(0) variables cannot cointegrate.

Table 2: Unit root tests

a. GDP(E)

Vintage	Augmented Dickey-Fuller				Philips-Peron			
	t-statistic	5% c.v.	Prob.	I(0)/I(1)	adj. t-statistic	5% c.v.	Prob.	I(0)/I(1)
T	0.903	-2.883	0.995	I(1)	0.936	-2.883	0.996	I(1)
T+12	0.555	-2.883	0.988	I(1)	0.488	-2.883	0.986	I(1)
T+24	0.112	-2.883	0.966	I(1)	0.034	-2.883	0.959	I(1)
T+60	-0.995	-2.883	0.754	I(1)	-0.582	-2.883	0.87	I(1)

b. Private consumption

Vintage	Augmented Dickey-Fuller				Philips-Peron			
	t-statistic	5% c.v.	Prob.	I(0)/I(1)	adj. t-statistic	5% c.v.	Prob.	I(0)/I(1)
T	0.366	2.883	0.981	I(1)	1.088	2.883	0.997	I(1)
T+12	-0.669	-2.883	0.85	I(1)	0.262	2.883	0.976	I(1)

T+24	-1.289	-2.883	0.633	I(1)	-0.026	2.883	0.954	I(1)
T+60	-1.336	-2.883	0.612	I(1)	-0.203	2.883	0.934	I(1)

## c. Government consumption

Vintage	Augmented Dickey-Fuller				Philips-Peron			
	t-statistic	5% c.v.	Prob.	I(0)/I(1)	adj. t-statistic	5% c.v.	Prob.	I(0)/I(1)
T	2.453	-2.883	1	I(1)	1.617	-2.883	1	I(1)
T+12	0.692	-2.883	0.992	I(1)	0.733	-2.883	0.993	I(1)
T+24	2.286	-2.883	1	I(1)	2.421	-2.883	1	I(1)
T+60	1.164	-2.883	0.998	I(1)	1.111	-2.883	0.998	I(1)

## d. Total investment

Vintage	Augmented Dickey-Fuller				Philips-Peron			
	t-statistic	5% c.v.	Prob.	I(0)/I(1)	adj. t-statistic	5% c.v.	Prob.	I(0)/I(1)
T	-2.409	-2.883	0.141	I(1)	-2.403	-2.883	0.1427	I(1)
T+12	-0.798	-2.883	0.816	I(1)	-0.865	-2.883	0.797	I(1)
T+24	-1.019	-2.883	0.745	I(1)	-1.007	-2.883	0.75	I(1)
T+60	-1.365	-2.883	0.598	I(1)	-1.395	-2.883	0.583	I(1)

## e. Exports

Vintage	Augmented Dickey-Fuller				Philips-Peron			
	t-statistic	5% c.v.	Prob.	I(0)/I(1)	adj. t-statistic	5% c.v.	Prob.	I(0)/I(1)
T	-2.242	-2.883	0.193	I(1)	-2.026	-2.883	0.275	I(1)
T+12	-0.671	-2.883	0.849	I(1)	-0.594	-2.883	0.867	I(1)
T+24	-0.328	-2.883	0.916	I(1)	-0.118	-2.883	0.944	I(1)
T+60	-0.328	-2.883	0.916	I(1)	-0.094	-2.883	0.947	I(1)

## f. Imports

Vintage	Augmented Dickey-Fuller				Philips-Peron			
	t-statistic	5% c.v.	Prob.	I(0)/I(1)	adj. t-statistic	5% c.v.	Prob.	I(0)/I(1)
T	-1.502	-2.883	0.53	I(1)	-1.771	-2.883	0.394	I(1)
T+12	-0.795	-2.883	0.817	I(1)	-0.813	-2.883	0.812	I(1)
T+24	-0.202	-2.883	0.934	I(1)	-0.161	-2.883	0.939	I(1)
T+60	-0.208	-2.883	0.933	I(1)	-0.176	-2.883	0.938	I(1)

### 6.3 Lag length tests

Table 3 shows the results of a number of information criterion on the recommended order of the VAR model. The Schwarz Criterion and the Hannan-Quinn criterion provide the most stable results and recommend using a first order VAR model, i.e. with one lag in the cointegrating framework for each variable.

Table 3: Lag length criteria tests (p)

	LR	FPE	AIC	SC	HQ
GDP(E)	7	7	7	1	1
Private consumption	5	1	1	1	1
Government consumption	2	2	2	1	1
Total investment	3	3	3	1	1
Exports	5	5	5	1	1
Imports	5	5	5	1	1

LR: sequential modified likelihood ratio test statistic (each test at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hannan-Quinn information criterion

## 6.4 Cointegration tests

Table 4 shows there is very little evidence of cointegration between the four data vintages for GDP(E) and its components. For private consumption, total investment and imports the cointegrating rank  $r=0$ , implying each data vintage is driven by its own stochastic trend. For the other variables, GDP(E), government consumption and exports the cointegrating rank  $r=1$  implies there are 3 different stochastic trends among the four data vintages. These results reject the notion that one vintage solely accounts for the long memory component in GDP(E) or its components. In fact the revisions process adds stochastic trends to the VAR.

Table 4: Johansen trace and eigenvalue test results for the cointegrating rank ( $r$ )

	Intercept and no trend		Intercept and trend	
	Trace	Max eigenvalue	Trace	Max eigenvalue
GDP(E)	1	1	1	1
Private consumption	0	0	0	0
Government consumption	1	1	0	0
Total investment	0	0	0	0
Exports	1	1	1	1
Imports	0	0	0	0

## 6.5 Structural breaks and cointegration

The presence of structural breaks in the revisions process may account for some of the failure to find cointegration between different data vintages. This is because structural breaks push towards accepting the null of a unit root when looking at the difference between the levels of different data vintages. Following Joyeux [31] I used a number of intervention dummies in the VAR model to try and account for any shifts in intercepts and trends but the results were largely unchanged and still did

not imply a well-behaved data measurement process.

These results are not out of canter with other literature. Paterson and Heravi [32] find evidence of a well-behaved data measurement process in US industrial production data but the data vintages they use are only a few months apart. In his UK work, Patterson [28] finds the data measurement process is well-behaved for 8 data vintages but when the latest data vintage is added the process is no longer found to be governed by a single stochastic trend. The addition of new data vintage published sometime after the other vintages introduced a new stochastic trend into the VAR. These findings support the conclusions reached in Siklos [33] which shows cointegration between data vintages breaks down as a result of benchmark and other significant revisions.

Therefore the failure to find cointegration between different data vintages could reflect two factors. First that the revisions process changes the underlying stochastic trend in the data. This is consistent with the view explored in the next section that the revisions process incorporates news into the data that was absent from earlier data vintages. The key question in this regard is how much of that news was known beforehand, i.e. how predictable is the revisions process, and this is investigated in the final section of the paper. Second, instability in the revisions process itself may bias against finding cointegrating relationships between different data vintages.

## 7 Are revisions news or noise?

Following Mankiw and Shapiro [34] and Patterson and Heravi [35] there are two hypotheses used to describe the revisions process for economic time series. Each has different implications for the statistical relationship between different data vintages.

The noise hypothesis assumes earlier vintages of data are afflicted with measurement error which is reduced in later vintages as estimates are refined.

$$\Delta y_t^i = \Delta y_t^{i+1} + \varepsilon_t$$

Therefore the revision between two successive data vintages ( $R^{i+1,i} = \Delta y_{t,i+1} - \Delta y_{t,i}$ ) reflects the reduction in measurement error  $-\varepsilon_t$ .

This process implies that the variances of successive data vintages should fall as measurement error is removed.

$$\text{var}(\Delta y_t^0) > \text{var}(\Delta y_t^1) > \text{var}(\Delta y_t^2) > \dots > \text{var}(\Delta y_t^{60})$$

Also revisions should be correlated with earlier data vintages but not later ones where the measurement error has been removed.

The news hypothesis suggests that each new vintage adds new information to the previous estimate.

$$\Delta y_t^{i+1} = \Delta y_t^i + \varepsilon_t$$

In this case the revision between successive data vintages reflects a positive innovation ( $\varepsilon_t$ ) which reflects the news absent in previous data vintages. This of course results in the opposite statistical properties between data vintages than the noise hypothesis. First, the variance of later vintages should be greater than earlier vintages.

$$\text{var}(\Delta y_t^0) < \text{var}(\Delta y_t^1) < \text{var}(\Delta y_t^2) < \dots < \text{var}(\Delta y_t^{60})$$

Second, revisions should be more strongly correlated with later data vintages as they incorporate the information missing from earlier data vintages.

It is of course very likely, as emphasized by Fixler and Grimm [16], that the actual revisions process will be driven by both news and noise. However, a simple analysis of variances and correlations can provide some introspection on which hypothesis may be the more relevant.

In Figure 9 I plot the standard deviation of a number of different data vintages for GDP(E) and its main components. I also break the full sample 1975Q2-2008Q4 into two sub-samples (1975Q2-1992Q4 and 1993Q1-2008Q4) to see if the variance relationships between different data vintages have changed over time.

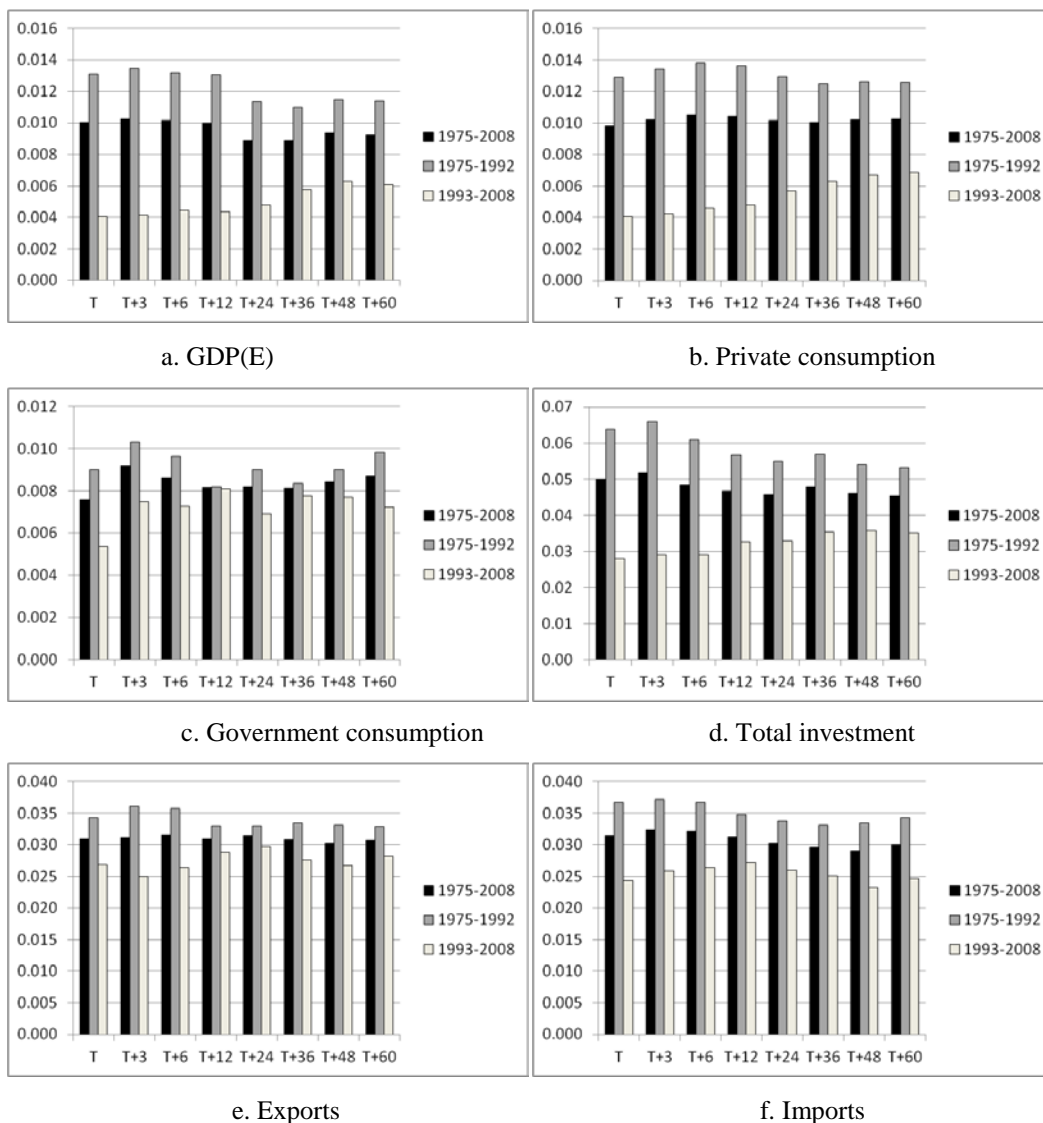


Figure 9: News or noise: the standard deviation of different data vintages

The results are interesting. On the full sample the noise hypothesis seems slightly more applicable for vintages up to T+24, especially for GDP(E) and total investment. However, in the second sub-sample there is clear support for the news hypothesis for GDP(E) overall and its main components with the exception of the trade data.

Table 5: News or noise: correlations

		Revision between T and T+24		Revision between T and T+60		Revision between T+24 and T+60	
<b>GDP(E)</b>	1975-2008	Correlation with T	-0.51	Correlation with T	-0.52	Correlation with T+24	-0.13
		Correlation with T+24	0.25	Correlation with T+60	0.38	Correlation with T+60	0.36
	1975-1992	Correlation with T	-0.54	Correlation with T	-0.57	Correlation with T+24	-0.15
		Correlation with T+24	0.24	Correlation with T+60	0.33	Correlation with T+60	0.31
	1993-2008	Correlation with T	0.05	Correlation with T	0.11	Correlation with T+24	0.11
		Correlation with T+24	0.54	Correlation with T+60	0.75	Correlation with T+60	0.61
<b>Private consumption</b>	1975-2008	Correlation with T	-0.26	Correlation with T	-0.31	Correlation with T+24	-0.16
		Correlation with T+24	0.35	Correlation with T+60	0.42	Correlation with T+60	0.24
	1975-1992	Correlation with T	-0.29	Correlation with T	-0.36	Correlation with T+24	-0.20
		Correlation with T+24	0.30	Correlation with T+60	0.30	Correlation with T+60	0.09
	1993-2008	Correlation with T	0.02	Correlation with T	-0.10	Correlation with T+24	-0.15
		Correlation with T+24	0.70	Correlation with T+60	0.81	Correlation with T+60	0.57
<b>Government consumption</b>	1975-2008	Correlation with T	-0.43	Correlation with T	-0.46	Correlation with T+24	-0.11
		Correlation with T+24	0.55	Correlation with T+60	0.63	Correlation with T+60	0.49
	1975-1992	Correlation with T	-0.46	Correlation with T	-0.47	Correlation with T+24	-0.11
		Correlation with T+24	0.46	Correlation with T+60	0.59	Correlation with T+60	0.49
	1993-2008	Correlation with T	-0.41	Correlation with T	-0.45	Correlation with T+24	-0.09
		Correlation with T+24	0.71	Correlation with T+60	0.75	Correlation with T+60	0.54
<b>Total investment</b>	1975-2008	Correlation with T	-0.48	Correlation with T	-0.52	Correlation with T+24	-0.14
		Correlation with T+24	0.29	Correlation with T+60	0.34	Correlation with T+60	0.28
	1975-1992	Correlation with T	-0.54	Correlation with T	-0.59	Correlation with T+24	-0.17
		Correlation with T+24	0.21	Correlation with T+60	0.25	Correlation with T+60	0.24
	1993-2008	Correlation with T	-0.23	Correlation with T	-0.22	Correlation with T+24	-0.02
		Correlation with T+24	0.56	Correlation with T+60	0.63	Correlation with T+60	0.39
<b>Exports</b>	1975-2008	Correlation with T	-0.25	Correlation with T	-0.32	Correlation with T+24	-0.17
		Correlation with T+24	0.30	Correlation with T+60	0.30	Correlation with T+60	0.10
	1975-1992	Correlation with T	-0.33	Correlation with T	-0.36	Correlation with T+24	-0.11
		Correlation with T+24	0.19	Correlation with T+60	0.22	Correlation with T+60	0.16
	1993-2008	Correlation with T	-0.11	Correlation with T	-0.25	Correlation with T+24	-0.30
		Correlation with T+24	0.44	Correlation with T+60	0.39	Correlation with T+60	-0.01
<b>Imports</b>	1975-2008	Correlation with T	-0.32	Correlation with T	-0.35	Correlation with T+24	-0.09
		Correlation with T+24	0.17	Correlation with T+60	0.20	Correlation with T+60	0.19
	1975-1992	Correlation with T	-0.40	Correlation with T	-0.38	Correlation with T+24	-0.04
		Correlation with T+24	0.06	Correlation with T+60	0.15	Correlation with T+60	0.26
	1993-2008	Correlation with T	-0.11	Correlation with T	-0.25	Correlation with T+24	-0.24
		Correlation with T+24	0.37	Correlation with T+60	0.30	Correlation with T+60	0.01

These findings are corroborated by the correlations presented in Table 5. Looking at the full sample, we can see total revisions between the first published estimate and the T+24 and T+60 vintages are supportive of the noise hypothesis in the overall sample and the earlier sub-sample. Revisions between the T+24 and



T+60 vintages though are less strongly supportive of the noise hypothesis.

In the second sub-sample the correlations are more strongly supportive of the news hypothesis with the exception of the trade data where revisions between the T+24 and T+60 vintages appear to be more noise than news.

The news hypothesis is consistent with the results from cointegration testing which showed that revisions between different data vintages introduced a new stochastic trend to the system. Basically the difference between the levels of different data vintages were non-stationary I(1) processes and additional information or news would be required to render this to a stationary I(0) process consistent with cointegrating relationships. The issue addressed in the next and final section is whether the news in data revisions is available to data producers before first or preliminary estimates are published. If so it would imply that first published data were inefficient or irrational estimates of later data vintages as available information that could lower subsequent revisions were not being incorporated into estimates.

## 8 Are early data vintages rational?

### 8.1 The efficient forecast hypothesis

For early data vintages to be an efficient forecast of later vintages they should be incorporating all the relevant information available at the time. Therefore, data producers are acting rationally if revisions are only a result of previously unknown news. This proposition is easily tested by the econometrician using a Mincer and Zarnowitz [36] regression to test the predictability of revisions:

$$R_t^{i,0} = a + b * \Delta y_t^0 + \mathbf{X}'_t \mathbf{Y} + \varepsilon_t$$

where  $\mathbf{X}_t = [x_{1,t}, x_{2,t}, \dots, x_{n,t}]$  is a vector of indicators available to the modeler ahead of the publication and  $\mathbf{Y} = [\gamma_1, \gamma_2, \dots, \gamma_n]$  the corresponding set of coefficients.

As all the right hand side variables are known to the data producer at the time

of publication they should not systematically predict revisions to future data vintages. If they were to do so then future revisions could be reduced on average by adjustments based on the model's estimated coefficients. The efficient forecast hypothesis (EFH) would imply:

$$\text{EFH } a = b = \gamma_1 = \gamma_2 = \dots = \gamma_n = 0$$

There are a number of research articles investigating the EFH for different macroeconomic time series and the results are inconclusive.

Mankiw and Shapiro [34] using seasonal dummies, past growth rates and financial market data find revisions to US GNP are not forecastable and so preliminary estimates satisfy the EFH. Faust, Rogers and Wright [37] using a data set consisting of lagged preliminary estimates, equity prices, short term interest rates, oil prices, and seasonal and General Election dummies find the degree of predictability varies across G7 countries. While US revisions were generally unpredictable they report for the UK half the variance of longer term (> 24 months) and a quarter of the variance of shorter-term (<24 months) revisions could be accounted for by their augmented regression model. Richardson [38] reports revisions to UK GDP as both biased and predictable, especially for longer-term revisions. Garratt and Vahey [39] also find revisions to be predictable in the UK.

The Bank of England has for a long time analyzed a range of other data alongside preliminary estimates of GDP in forming their view of the economy. This was previously done informally as described by Britton, Cutler and Wardlow [40]. Their recent approaches, see Ashley, Driver, Hayes and Jeffery [41] and Cunningham et al [18] have both attempted to produce models which use a range of survey data and financial indicators to predict the revisions process in the UK.

Many of these studies are based on a limited span of data. In longer samples the instability in the time series and revisions processes makes it harder to find general predictability. Garratt, Koop and Vahey [42] find evidence of a structural break in the UK revisions process after the 1980s with a sharp fall in the probability of large revisions. Castle, Fawcett and Hendry [43] conclude that the success of nowcasting

depends on the responsiveness to structural breaks in the data, in particular the identification of turning points. Swanson and van Dijk [17] argue that knowledge of the current cyclical position of the economy is a key issue in predicting future data revisions.

## **8.2 Testing the rationality of UK first published data**

In this sub-section I estimate a Mincer-Zarnowitz regression on UK GDP(E) and the main expenditure components data to test for the predictability of revisions between the first published estimate (vintage T) and that two years later (vintage T+24). I decided not to do this analysis on the T+60 vintage as I thought this approach more applicable to data-driven rather than methodology-driven revisions. Also, because at any given point in time the data user will not know future revisions the model coefficients are estimated on past data. Revisions data between the T and T+24 vintages are available with a 24 month lag, while that between the T and T+60 vintages would incur a rather excessive 60 month lag.

There is a potentially large set of indicators available. The key concern is that they are available in real time and before the first data vintage is published. It also helps, although not mandatory, if the indicators themselves are not subject to revision. I use consumer and business sentiment data from Eurostat and the OECD, plus financial market data such as interest rates, exchange rates and stock market prices. In order to account for instability I estimate the regression model on a moving window of 60 observations (12 years). Given data availability this enables me to form nowcasts of the T+24 data vintage over the sample 2001Q4 to 2011Q4, a total of 41 observations. Note I could also have produced nowcasts for 2012Q1 to 2013Q4 but would not have actuals for the T+24 data vintage to compare with.

The main difficulty with this approach is the set of available indicators is large relative to the 60 observations used to estimate the model. As a result it is easy to run out of a sufficient degrees of freedom. To do this I follow Castle, Fawcett and

Hendry [43] by using automatic model selection methods for general to specific modeling (GETS). This starts with a general model containing all the available indicators, and through the process of deleting insignificant variables end up with a smaller model of significant regressors. The general problem with GETS type modeling approaches is that the order of deletion matters. A simple rule whereby the least significant variable is deleted at each stage based on a t-test provides just one of many deletion approaches and is found to be rarely the most efficient. Repeated testing may allow Type 1 errors to accumulate. Insignificant variables may remain in the regression by acting as a proxy for others that are deleted. However increasing the size of the test may increase the likelihood of the chance deletion of relevant variables.

Table 6: Key variables from automatic regression models for each data variable

	GDP(E)	Private consumption	Government consumption	Total investment	Exports	Imports
First estimate (T)	39	26	38	36	16	26
Constant		41			12	
<i>Eurostat survey</i>						
Economic sentiment	25	41	11	11		28
Construction confidence	11					
Industrial confidence		37				
Industrial new orders			12			
Consumer confidence		40				
<i>OECD survey</i>						
Manufacturing tendency			19			
Manufacturing future tendency			11			
Construction business situation	13					
Construction order books			11			
Consumer prices		34				
<i>Financial indicators</i>						
FTSE					12	
Bank of England base rate					15	
Effective exchange rate	18	11				
Broad effective exchange rate	21		11			

To circumvent some of these problems I use the automatic regression approach outlined by Hoover and Perez [44] and Hendry and Krolzig [45]. This commences with a General Unrestricted Model (GUM) consisting of relevant variables chosen by the researcher. The automatic algorithm then explores multiple deletion paths, acting like a series of data sieves. Data reductions are only permitted if the subsequent model passes specification tests. Encompassing tests and selection criteria are then used to choose from among the final non-nested models.

Table 6 reports the indicators that are significant in the Mincer-Zarnowitz regression models for each variable, showing the indicators that were selected by the automatic regression approach at least 10 times out of the 41 total model estimates. The results of the nowcasting exercise are plotted in Figure 10 which shows the T and T+24 vintages along with a predicted T+24 vintage based on this modeling approach.

Table 7 provides summary statistics on the revisions between the T and T+24 vintages, and the predicted T+24 and actual T+24 vintages. Evidence of irrationality would show if estimated models lowered average revisions. However, Table 7 in fact shows the opposite is true. For each expenditure component and GDP(E) itself the nowcasts result in higher mean absolute and mean squared revisions than simply just taking the first published estimate.

Table 7: Revisions between the T :T+24 and T:t+24(predicted) data vintages

	Revisions T:T+24			Revisions T+24 (predicted):T+24		
	Mean	Mean absolute	Mean squared	Mean	Mean absolute	Mean squared
GDP(E)	0.01	0.20	0.06	0.02	0.28	0.15
Private consumption	-0.13	0.32	0.15	-0.14	0.44	0.28
Government consumption	-0.19	0.59	0.53	0.01	0.70	1.07
Total investment	0.41	2.08	7.58	-0.08	3.23	17.89
Exports	0.49	1.70	5.33	-0.14	1.93	6.73
Imports	0.32	1.11	1.79	-0.46	1.21	2.45

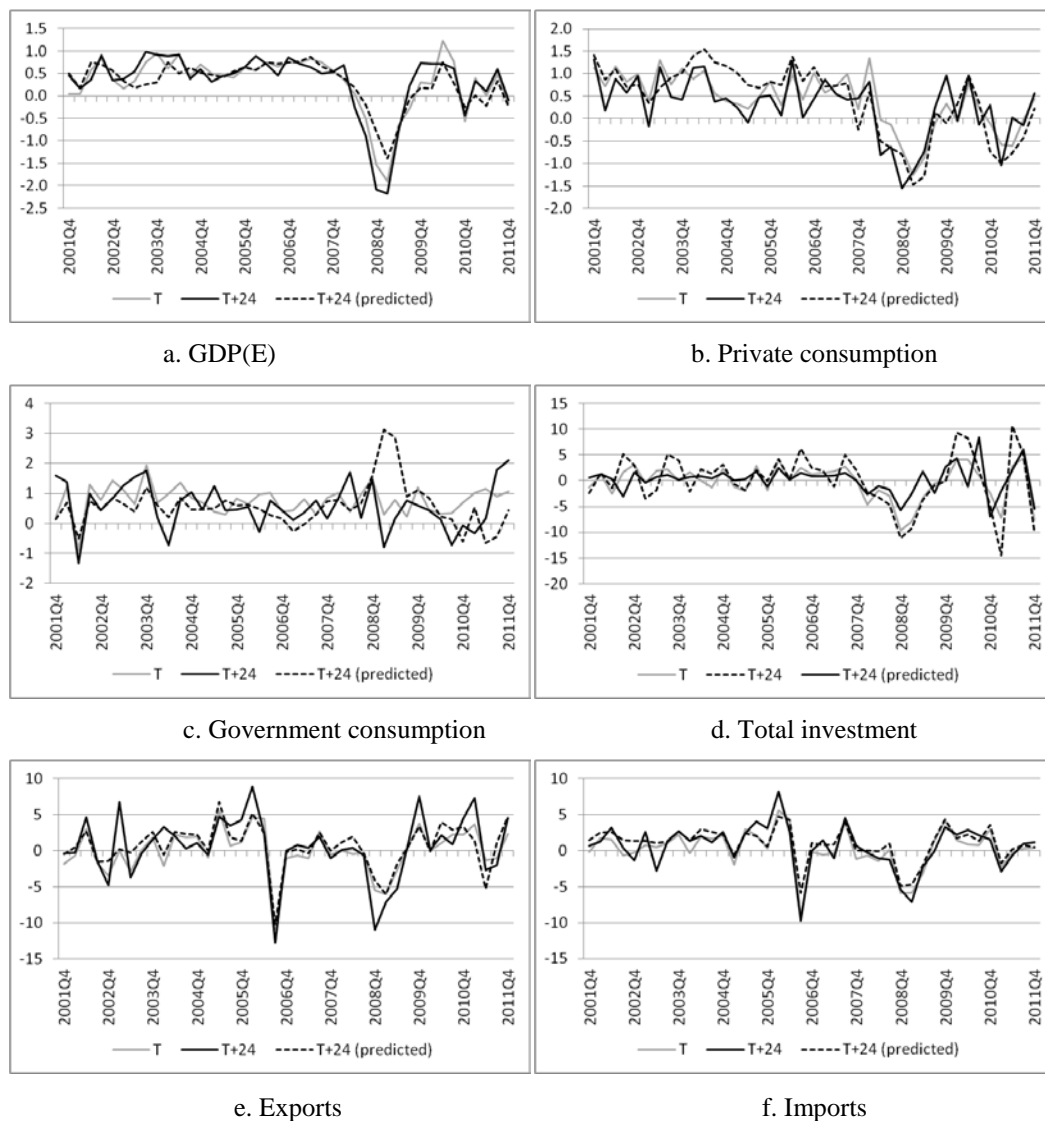


Figure 10: T, T+24 and T+24(predicted) data vintages

The reason for this is easy to see in Figure 10. In most cases the nowcasts do a very bad job of predicting revisions around the recession when data was revised down rather than upward as it had been on average through the Great Moderation period leading up to the recession. Had this exercise been conducted on a sample before the great recession in 2008, as was most of the previous literature, then it is

likely there would have been stronger evidence of predictability of revisions. This however is in essence the problem with nowcasting referred to by Castle, Fawcett and Hendry [43]. It is clear that structural and cyclical changes to the economy and methodology changes in the production of statistics can have an effect on the predictability of revisions so nowcasting models that perform well in some samples are still likely to eventually break down.

Table 8: The number of times the model changed between successive automatic regressions

	Model changes	Empty models
GDP(E)	34	0
Private consumption	15	0
Government consumption	29	2
Total investment	13	5
Exports	18	15
Imports	21	0

Table 8 demonstrates the inherent instability of the nowcasting models used here. It shows the number of occurrences of when the automatic algorithm selected different model variables between runs as well as the number of empty models where no indicators were significant. You will see that the models changed a lot, for instance the GDP(E) model changed 34 times in 41 runs. This may reflect the sensitivity of the algorithm to variables that are relatively similar. Furthermore, automatic regression approaches tend to be less efficient if there is multicollinearity among the variable set. However, I also take it as further evidence of the practical difficulty of producing accurate nowcasts when the underlying data is subject to structural and cyclical changes.

## 9 Concluding comments

This article has provided a detailed analysis of revisions to expenditure data in the UK. Compared to other work the analysis has considered a longer history of revisions that has incorporated a number of economic cycles. I have found the revisions process to be unstable, influenced by cyclical movements in the economy and frequent structural breaks. An important source of revisions are methodological changes to the way the data is compiled and these are often implemented in an ad-hoc way so the timing and impact is both significant and hard to predict.

The trade-off between timeliness and accuracy in the publication of economic statistics is a challenge for policymakers, especially in the field of monetary policy which has to act preemptively to hit an inflation target in the medium term. As a result there has been a lot of empirical work on the predictability of revisions, whether using bias adjustments or other indicators available in real time. These approaches may offer a short-term improvement to the timeliness versus accuracy trade-off, but because of instability in the revisions process these predominately time series econometrics approaches tend to be unreliable in longer samples.

The key to reducing revisions may therefore lie elsewhere. Larger samples that are more responsive to economic conditions are an obvious starting point. Other areas of investigation might include an earlier balancing of the income, output and expenditure measures of GDP, and better price indices at earlier vintages for the deflation of nominal to real expenditure measures.

## References

- [1] McNees S, Forecasts and actual: the trade-off between timeliness and accuracy, *International Journal of Forecasting*, **5**, (1989), 409-416.
- [2] Croushore D and Stark T, A real-time dataset for macroeconomists: Does the vintage matter?, *Review of Economics and Statistics*, **85**(3), (2001), 605-617.



- [3] Fackler J, Comments on Forecasting with a real-time data set for macroeconomists, *Journal of Macroeconomics*, **24**, (2002), 559-562.
- [4] Croushore D and Stark T, Forecasting with a real-time data set for macroeconomists, *Journal of Macroeconomics*, **24**, (2002), 507-531.
- [5] Kozicki S, Comments on Forecasting with a real-time data set for macroeconomists, *Journal of Macroeconomics*, **24**, (2002), 541-557.
- [6] Elliot G, Comments on Forecasting with a real-time data set for macroeconomists, *Journal of Macroeconomics*, **24**, (2002), 533-539.
- [7] Koenig E, Dolmas S and Piger J, The use and abuse of real-time data in economic forecasting, *Review of Economics and Statistics*, **85**(3), (2003), 618-628.
- [8] Croushore D, Frontiers of real-time data analysis, *Federal Reserve Bank of Philadelphia mimeo*, (2009).
- [9] Croushore D and Stark T, A real-time dataset for macroeconomists, *Journal of Econometrics*, **115**, (2001), 111-130.
- [10] Castle J and Ellis C, Building a real time database for GDP(E), *Bank of England Quarterly Bulletin*, (Spring, 2002), 42-49.
- [11] Eggington D, Pick A and Vahey S, Keep it real: A real-time UK macro data set, *Economic Letters*, **77**(1), (2002), 15-20.
- [12] Boragan Aruba S, Data revisions are not well behaved, *Journal of Money, Credit and Banking*, **40**(2-3), (2008), 320-340.
- [13] Nelson and Nicolov, UK inflation in the 1970s and 1980s: the role of output gap mismeasurement, *Journal of Economics and Business*, **55**(4), (2003), 353-370.
- [14] Pickford S, Cunningham J, Lynch J, Radice R, White G, Government economic statistics: a scrutiny report, Cabinet Office, London, (1989).
- [15] Dynan K and Elmendorf D, Do provisional estimates of output miss turning points, *Federal Reserve Bank of Washington working paper*, (November, 2001), (2001).

- [16] Fixler D and Grimm B, Revisions, rationality and turning points in GDP, Bureau of Economic Analysis, **WP2003-01**, (2003).
- [17] Swanson N and van Dijk D, Are statistical reporting agencies getting it right? Data rationality and business cycle asymmetry, *Journal of Business and Economic Statistics*, **24**, (2006), 24-42.
- [18] Cunningham A, Eklund J, Jeffery C, Kapetanios G and Labhard V, A state space approach to extracting the signal from uncertain data, *Bank of England working paper*, **336**, (2007).
- [19] Oller L-E and Teterukovsky A, Quantifying the quality of macroeconomic variables, *International Journal of Forecasting*, **23**, (2007), 205-217.
- [20] Kholodilin K and Siliverstovs B, Do forecasters inform or reassure? Evaluation of the German real-time data, *Applied Economics Quarterly*, **55**(4), (2009), 269-293.
- [21] Patterson K and Heravi S, Revisions to the national income accounts: a multiplicative decomposition of mean square errors, *Discussion papers in macroeconomics, forecasting and econometrics*, **1**, University of Reading, (1990).
- [22] Beirlant J, Dudewicz E, Györfi L and van der Meulen E, Nonparametric entropy estimation: an overview, *Journal of Mathematical and Statistical Sciences*, **6**, (1997), 17-39.
- [23] Vasicek, A test for normality based on sample entropy, *Journal of the Royal Statistical Society Series B*, **38**, (1976), 54-59.
- [24] Noughabi and Arghami, A new estimator of entropy, *Journal of Iranian Statistical Society*, **9**(1), (2010), 53-64.
- [25] Stock J and Watson M, Testing for common trends, *Journal of the American Statistical Association*, **83**, (1988), 1097-1107.
- [26] Patterson K and Heravi S, Are different vintages of data on the components of GDP cointegrated?, *Economics Letters*, **35**, (1991), 409-413.

- [27] Patterson K, Which vintage to use when there are multiple vintages of data? Cointegration, weak exogeneity and common factors, *Economics Letters*, **69**, (2000), 115-121.
- [28] Patterson K, The data measurement process for UK GNP: stochastic trends, long memory and unit roots, *Journal of Forecasting*, **21**, (2002), 245-264.
- [29] Johanson S and Juselius K, Testing structural hypotheses in a multivariate cointegration analysis of the PP and UIP for UK, *Journal of Econometrics*, **53**, (1992), 211-244.
- [30] Gonzalo J and Granger C, Estimation of common long-memory components in cointegrated systems, *Journal of Business and Economic Statistics*, **13**, (1995), 27-35.
- [31] Joyeux R, How to deal with structural breaks in practical cointegration analysis, *School of Economic and Financial Studies, Macquaire University mimeo*, (2001).
- [32] Patterson K and Heravi S, Revisions to official data on US GNP: a multivariate assessment of different vintages, *Journal of Official Statistics*, **20**, (2004), 573-602.
- [33] Siklos P, What can we learn from comprehensive data revisions for forecasting inflation? Some US evidence, *Department of Economics, Wilfrid Laurier University mimeo*, (2006).
- [34] Mankiw N G and Shapiro M, News or noise? An analysis of GNP revisions, *NBER Working Paper*, **1939**, (1986).
- [35] Patterson and Heravi, Efficient forecasts or measurement errors? Some evidence for revisions to United Kingdom growth rates, *Manchester School*, **60**(3), (1992), 249-263.
- [36] Mincer J and Zarnowitz V (1969), *The evaluation of economic forecasts*, J Mincer, ed. *Economic forecasts and expectations*, NBER New York, 1969.
- [37] Faust J, Rogers J and Wright J, News and noise in G7 GDP announcements, *Journal of Money Credit and Banking*, **37**(3), (2005), 403-419.

- [38] Richardson C, Revisions analysis: a time series approach, *Economic Trends*, (December, 2003), 86-89.
- [39] Garratt A and Vahey S, UK real time data characteristics, *Economic Journal*, **116**, (2006), F119-F135.
- [40] Britton E, Cutler J and Wardlow A, The Bank's use of survey data, *Bank of England Quarterly Bulletin*, (February, 1999), 177-182.
- [41] Ashley J, Driver R, Hayes S and Jeffery C, Dealing with data uncertainty, *Bank of England Quarterly Bulletin*, (Spring, 2005).
- [42] Garratt A, Koop G and Vahey S, Forecasting substantial data revisions in the presence of model uncertainty, *Economic Journal*, **118**, (2008), 1128-1144.
- [43] Castle J, Fawcett N and Hendry D, Nowcasting is not just contemporaneous forecasting, *National Institute Economic Review*, **210**, (2009), 71-89.
- [44] Hoover K and Perez, Data mining reconsidered: a general to specific approach to specification search, *Econometrics Journal*, **2**, (1999), 167-191.
- [45] Hendry D and Krolzig H-M, Computer automation of general to specific model selection procedures, *Journal of Economic Dynamics and Control*, **5**(6-7), (2000), 831-866.