STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS

NOWCASTING UK GDP DURING THE DEPRESSION

BY

PAUL SMITH

NO 16-06

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STRATHCLYDE
GLASGOW

Nowcasting UK GDP during the Depression

Paul Smith *

April 2016

Abstract

Nowcasting UK GDP during the Depression reviews the performance of several statistical techniques in nowcasting preliminary estimates of UK GDP, particularly during the recent depression. Traditional bridging equations, MIDAS regressions and factor models are all considered. While there are various theoretical differences and perceived advantages for each technique, replicated real-time out-of-sample testing shows that, in practice, there is in fact little to choose between methods in terms of end-of-period nowcasting accuracy.

The analysis also reveals that none of the aforementioned statistical models can consistently beat a consensus of professional economists in nowcasting preliminary GDP estimates.

This inability of statistical models to beat the consensus may reflect several factors, one of which is the revisions and re-appraisal of trends inherent in UK GDP statistics. The suggestion is that these changes impact on observed relationships between GDP and indicator variables such as business surveys, which impairs nowcasting performance. Indeed, using a synthetic series based purely on observed preliminary GDP estimates, which introduces stability to the target variable series, the nowcasting accuracy of regressions including closely-watched PMI data is improved by 25-40 percentage points relative to a naive benchmark.

JEL Classification: C22, C32
Keywords: Nowcasting, Forecasting, Real-Time Data, GDP, MIDAS

*Department of Economics, University of Strathclyde, Sir William Duncan Building, 130 Rottenrow, Glasgow G4 0GE, United Kingdom: Tel: +44 1491 461 038 e-mail: paul.smith.101@strath.ac.uk

1 Introduction

This paper reviews the performance of several statistical techniques in nowcasting preliminary estimates of UK GDP, particularly during the recent depression.

The basis for conducting this research is based on the observation of a lack of clarity from the literature on which method is the "best" for nowcasting GDP.

The paper therefore attempts to offer some guidance to practitioners by conducting a nowcast "horserace" between methods. Traditional bridging equations, MIDAS regressions and factor models are all considered.

However, while there are various theoretical differences and perceived advantages for each technique, replicated real-time out-of-sample testing shows that, in practice, there is in fact little to choose between methods in terms of end-of-period nowcasting accuracy.

The analysis also questions some of the literature that has suggested particular methods could outperform the "wisdom" contained within a consensus. By conducting real-time analysis, the paper shows that, for the UK at least, none of the aforementioned statistical models can consistently beat a consensus of analysts in nowcasting preliminary estimates of quarter-on-quarter changes in GDP.

This inability of statistical models to beat the consensus may reflect several factors, one of which is the revisions and re-appraisal of trends inherent in UK GDP statistics.

The suggestion is that these changes impact on observed relationships between GDP and indicator variables such as business surveys, which then impairs nowcasting performance.

Subsequently the paper offers an alternative to practitioners by suggesting focus could be placed on a series that is based purely on preliminary estimates of quarter-on-quarter changes in GDP. By introducing a new stability to the target variable series, the nowcasting accuracy of regressions including closely-watched PMI data is found to be improved by 25-40 percentage points relative to a naive benchmark.

1.1 Nowcasting

In July 2014 the UK Office for National Statistics (ONS) reported that, after six-and-a-half years, the longest UK post-war depression was over. Gross

Domestic Product (GDP) in the second quarter of 2014 was estimated to be 0.2% higher than its previous peak in Q1 2008. From peak to the trough in 2009 the economy was estimated to have shrank 7.2%. It subsequently took over five years to recover the ground lost, although at the time of writing subsequent revisions now put the peak to trough fall at 6.0% while the depression is now estimated to have finished in Q3 2013.

Given the largely unprecedented swings in economic output throughout this period, not just in the UK, but around the globe, interest and demand in understanding how the economy was performing in a timely manner heightened, especially as GDP data are published with a lag and subject to considerable revision post publication. This led to a growing body of academic work in a sub-field of forecasting commonly referred to as "nowcasting".

Generally speaking, the aim of nowcasting is to link GDP to the flow of information emanating from some kind of heterogeneous dataset.

As an example, the preliminary estimate of UK GDP covering the first quarter of 2014 was available on the 29th April 2014, nearly a month following the end of the quarter. Being quarterly, this is the first comprehensive update on the performance of the economy for the first three months of the year as a whole. Previously available information only went up to the end of 2013.

But throughout the quarter, data for several other variables that offer a steer on economic performance are also available. These include direct "hard" indicators that may be used to compile the GDP statistics, such as monthly industrial production figures. In early March, for example, the ONS reported figures for the performance of industry in January. "Soft" indicators such as business surveys are also available, and in a more timely manner, being typically released around the beginning of the month, but offering a qualitative take on current economic conditions. Developments in the financial, housing and labour markets are also likely to be monitored.

Attempts to successfully exploit the information contained within these variables leads to a number of challenges from the perspective of the econometrician interested in predicting GDP growth.

Firstly, the dependent variable is quarterly, whereas data for many of the explanatory variables are available on a monthly, weekly or even, in the case of financial markets, daily basis. This creates a mis-match in terms of time frequencies which are not easily handled in traditional forecasting frameworks.

Secondly, there is the so-called jagged edge: the variables contained within the nowcaster's dataset typically have separate release dates and may refer to different reference periods. Maintaining the example of nowcasting UK GDP in Q1 2014, the release of industrial production data covering January was the 9th of March, whereas the PMI business surveys for February were available over the 3-5th March. Such a situation results in missing observations for a number of time series which is especially problematic as the nowcaster typically wishes to update their predictions for GDP on a continuous basis.

Various methods have been proposed to deal with both the mixed-time frequency and ragged edge issues, such as bridging equations, Mixed Data Sampling (MIDAS), mixed frequency VARS and mixed frequency factor models. Several excellent surveys have emerged that provide extensive details of these approaches and associated econometric studies including Bańbura et al. (2013), Camacho, Perez-Quiros, and Poncela (2013), and Forini and Marcellino (2013).

A key takeaway from the literature is a broad agreement that the use of high frequency data can be successfully utilised to reduce uncertainty surrounding GDP estimates compared to some benchmark, especially as information accumulates throughout the nowcasting period.

But less clear is which method is best. While there are various theoretical differences across the model set-ups, which can give rise to user preferences based on theoretical grounds or the nowcaster's general aims, ranking according to perceived strengths and weaknesses is challenging. The question therefore becomes an empirical one, with the usefulness of any approach resting on its predictive accuracy i.e. how well do the models actually nowcast the variable of interest?

1.2 Outline of Paper

With the depression officially over, it seems a good time to review several of the nowcasting techniques and consider their performance in nowcasting UK GDP over this period of economic upheaval. Nowcasts that are produced relatively close to the release of the preliminary estimate of GDP are the primary consideration (lead time is around a week). This reflects the high degree of interest amongst institutions and analysts that surrounds the preliminary estimates of UK GDP. The contribution of the paper is to therefore ask, amongst competing methods, which is the most accurate at predicting preliminary estimates of GDP when conditioned on equivalent levels of information? The exercise is essentially a horserace between methods, an attempt to understand which one actually does best when they converge around a week to go before the release of GDP.

Bridging equations and MIDAS regressions are at the heart of the empirical work.

They are used firstly to provide GDP point nowcasts derived from single individual predictors, which can then also be pooled. Secondly, the bridging and MIDAS models are combined with monthly factors that are extracted from a dataset containing 24 variables (commonly referred to in the literature as "bridging with factors" and Factor-MIDAS modelling). Using a recursive out-of-sample modelling exercise, that is based on real-time data, the average nowcast errors provided by the various models covering the period 2006 to the end of 2013 are compared against a simple AR(1) benchmark. During the so-called Great Moderation such a benchmark was widely viewed as difficult to beat, but to add an additional layer of analysis the performance of a consensus of professional forecasters is also considered. Formal judgement tends to play an important role in the delivery of the consensus view, providing an interesting additional check on whether statistically driven model nowcasts can match, or even surpass, the "wisdom" contained within such polls.[1]

As a prelude, there is little difference to be found between the nowcasting performances of the models despite various differences in set-up and statistical features: simple bridging equation frameworks based on a small select set of indicators seem to perform just as well as models that (arguably) utilise more persuasive and sophisticated econometric frameworks.

However, none of the models are able to perform as well as the consensus nowcast, which exhibits a considerable performance advantage. This suggests judgement plays a role in nowcasting UK economic growth, supporting earlier assertions by Mitchell (2009) and more recently Bell et al. (2014).

There are a number of reasons why judgement may be important, one of which is the considerable revisions that UK GDP experiences. Such revisions for instance have changed the profile of the early years of the depression and may well have an impact on the stability of nowcasting regression equations.

With this in mind, and having outlined model properties, dataset features and empirical results of the nowcasting exercises over the period 2006-2013 in sections 2-5, section 6 provides further empirical results of a re-running of the recursive out-of-sample nowcasting exercises. The difference, however,

[1]For instance, the European Central Bank report that contributors to their Survey of Professional Forecasters considered that forty percent of their short-term GDP forecasts were judgment-based (ECB 2009).

is that the target variable is explicitly the preliminary estimate of GDP growth. With a stable target series, the pooled performance of the models is generally improved as there is a significant reduction in the average errors from the nowcasting equations based on business survey data covering the manufacturing and services sectors.

2 Model Frameworks

The aim of this section is to provide an overview of the various statistical models used to nowcast GDP and overcome several of the hurdles typically faced by practitioners. These of course don't cover all those proposed in the literature, but provide a good cross section spanning relatively simple bridging equations, which remain popular amongst practitioners, to more sophisticated models that include factor extraction. All have common themes in that they attempt to deal with the problems of mixed-time frequencies and missing dataset observations.

The formal benchmark to assess the performance of various statistical models is also introduced in this section, alongside the second important benchmark: the Bloomberg Consensus.

2.1 Bridging Equations and MIDAS Regressions

Model 1 is the AR(1) regression which is used as a benchmark to nowcast UK GDP, Y_t, and is defined as:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \varepsilon_t \tag{1}$$

Note that the time period t is quarterly and GDP is set as a quarter-on-quarter change.

A natural extension on model 1 is to add some kind of explanatory variable, X_t, which could be useful at predicting changes in GDP:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \gamma X_t + \varepsilon_t \tag{2}$$

where X_t is defined as an arithmetic average of the m observations over a single calendar quarter:

$$X_t = \frac{1}{m} \sum_{k=1}^{m} L_{HF}^{k} X_t^{HF} \tag{3}$$

This "two-equation" approach is commonly known as the bridging equation model. It is a parsimonious, popular and easy to implement framework which deals with a principal challenge involved in nowcasting, that of mixed time frequencies: GDP data are available on a quarterly basis, while data for a high frequency explanatory variable X_t^{HF} is available m times over a quarter. A common example is industrial production, a closely-watched supply side indicator that tends to have a close relationship with GDP and is available monthly (i.e. $m = 3$).

An equal weighted average of the high frequency data is taken to transform to the lower frequency sampling rate. The technique was first developed extensively for US GNP by Klein and Sojo (1989), with further examples and derivations seen in (amongst others) Baffigi, Gonelli, and Parigi (2004), Ingenito and Trehan (1996), and Rünstler et al. (2008).

A well worn criticism of bridging equations is that the average of high frequency data is performed with the potential cost of valuable information from individual timing innovations being diluted or lost.

An alternative solution would be to include, on the right-hand side of equation 2, the explanatory variable at its original sampling rate. So all observations have unique coefficients. However, as sampling frequency rises then parameter proliferation can be a problem (imagine daily data, for instance, where conceivably $m = 66$ over a quarter as a whole, assuming 22 trading days per month).

To find some middle ground between the issues of information loss and parameter proliferation, Ghysels, Santa-Clara, and Valkanov (2004) introduce the so-called MIDAS (Mixed Data Sampling) regression, which is built on here by adding an AR(1) term to create an AR-MIDAS model:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \gamma \sum_{j=1}^{p} \phi(k:\theta) L_{HF}^{k} X_{t-h}^{HF} + \varepsilon_t \tag{4}$$

In this set-up, changes in the dependent variable, GDP, are explained by one lag of itself and the high frequency variable, X_t^{HF}, lags of which may also be included. Temporal aggregation of X_t^{HF} is determined by a parametrised polynomial weighting function $\phi(k : \theta)$ to maintain parsimony in the model specification. Whereas MIDAS models generally focused on financial applications in early studies, more recently they have been used to forecast low frequency macroeconomic times series such as GDP using higher-frequency data. Armesto, Engemann, and Owyang (2010) provide an especially intuitive and easy-to-follow introduction. See also Kuzin, Marcellino, and Schumacher (2011) and Kuzin, Marcellino, and Schumacher (2013) for nowcasting the GDP of the eurozone and various industrialised countries using MIDAS.

The distributed lag polynomial weighting functions used in MIDAS regressions could take on many non-linear functional forms and various specifications have been considered. Preference could be dependent on the user's own beliefs such as placing greater weight on the more recent values (see Ghysels, Sinko, and Valkanov (2007) for a discussion). Throughout this paper's empirical work, an unrestricted form of MIDAS, where the weights are estimated without restriction, is used as per Marcellino and Schumacher (2007).

Models 2 and 3 are single explanatory variables approaches, but many data series that could be useful in nowcasting GDP are available e.g. business surveys, labour market statistics or financial markets variables.

A temptation is to add additional regressors to the models. However, the tendency within the literature has been to use individual nowcasting regressions and then take some kind of average of the resulting nowcasts. This helps to a) avoid over parametrization when using many exogenous variables b) any issues of co-linearity that can arise when using several macroeconomic time series in the same regression equation and c) leverage the idea that the pooling of forecasts based on several small forecasting (or nowcast) functions can yield better predictive performance see e.g. Aiolfi, Capistrán, and Timmermann (2010).

Let $\hat{Y}_{i,t}$ represent an individual model nowcast of Y_t. The number of nowcasts made is $i = 1, ...N$ which is equivalent to the number of explanatory variables used to predict Y_t i.e. the number of models created is the same as the quantity of predictive high frequency indicators X_t^{HF}. The overall nowcast $\hat{Y}_t$ is taken as the average of these N nowcasts:

$$\hat{Y}_t = \frac{1}{n} \sum_{i=1}^{n} \hat{Y}_{i,t} \tag{5}$$

2.2 Monthly Factor Model

Macroeconomists can access considerable volumes of data to track economic activity. Using individual bridging equations or MIDAS models for all available series can result in these approaches becoming weildy and difficult to implement in practical terms. Practitioners are subsequently tempted to track only a handful of data series to maintain a manageable modelling framework.

In this paper, five individual variables relating to business surveys and official output series are specifically tracked (and an average of their nowcasts) but it is recognised overall model capability ultimately rests on the performance of these: there is a vulnerability to a breakdown in relationships with GDP. And using a restricted dataset comes at the cost of potentially important information being discarded.

So, rather than pre-selecting indicators, a method used in macroeconomics to help shrink high dimensional datasets are factor models similar to those outlined in Stock and Watson (2011).

Assuming a high degree of co-movement across various series, these models extract r unobservable factors, which capture the bulk of the dynamics within the dataset containing N variables. Crucially $r << N$; the information held within a large volume of predictors is replaced by a much smaller number of estimated factors.

Redefine X_t^{HF} as a dataset containing N high frequency variables, all of which are available monthly. Assume this dataset has some kind of factor structure:

$$X_t^{HF} = \Lambda f_t^{HF} + \varepsilon_t \tag{6}$$

where $f_t^{HF} = (f'_{1,t}, f'_{2,t} \cdots f'_{r,t})$ represents a vector of r factors. Multiplying this vector by the $N \times r$ loadings matrix Λ provides the common component of each variable. The idiosyncratic components not explained by the factors but still part of X_t^{HF} are held in ε_t.

In the empirical application below, the monthly factors are derived from a static principal component analysis (PCA). A dynamic version of PCA was considered, but there seems little statistical difference in either approach when nowcasting see e.g. Marcellino and Schumacher (2007) or Jansen, Jin, and Winter (2014).

A further consideration is the number of principal components (or factors) to retain. One or two have been shown to capture the bulk of variation within macroeconomic datasets used in forecasting applications see e.g. Stock and Watson (2002) and with specific reference to nowcasting for various countries and regions Aastveit and Trovik (2008), Giannone, Reichlin, and Small (2008), and Yiu and Chow (2011). Conversely, some kind of information criterion may be desirable given a lack of theoretical grounding for a seemingly arbitrary number. Bai and Ng (2002) for instance propose a methodology that uses a penalty criteria in a combination with some kind of loss function to "correctly" choose the number of retained factors.

A pluralistic approach is taken, with one, two and a determined number of principal components retained. The deterministic approach is based on Kaiser's criterion. This involves retaining all those principal components with eigenvalues greater than one.

The respective sizes of the eigenvalues for each retained principal component are then used as weights to create a single high frequency factor, f_t^{HF}, which can be used directly in a bridging equation specification, which becomes model 4:

$$\hat{Y}_t = \alpha + \beta_1 Y_{t-1} + \gamma f_t + \varepsilon_t \tag{7}$$

where f_t is as an arithmetic average of the three observations of the monthly factor, f_t^{HF}, over a calendar quarter as per equation 3.

Moreover, the f_t^{HF}'s can also be used in the MIDAS regression to form model 5:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \gamma \sum_{j=1}^{p} \phi(j:\theta) L_{HF}^k f_{t-h}^{HF} + \varepsilon_t \tag{8}$$

Finally, a note on the 24 series used to create f_t^{HF} (descriptions of which are provided in table 1).

These indicators cover a wide range of activities that are likely to offer some kind of inference on economic growth. These include "soft" indicators such as business and consumers surveys, financial markets variables, plus "hard" data that offer monthly updates on the performances of (for example) industry and the service sector. In other words, a cross section of widely used data is contained within these indicators.

With hundreds of data series now available to macroeconomists, it is recognised there is a case to suggest N being equal to 24 seems small. However, there were several motivations for keeping the dataset around this size.

Firstly, there were practical considerations: creating and maintaining a database containing hundreds of data series can be challenging and may lead to some computational burden for the researcher.

Secondly, when the number of available series for analysis from one data source is large (say e.g. a business survey where the number of series could be over 10) is it really useful to include all of these? Experience suggests there tends to be a high degree of cross correlation within such single source datasets resulting in concerns of oversampling and excessive influence in the factor calculation.

Thirdly, Boivin and Ng (2006) have suggested more data is not necessarily better. In simulations to forecast macroeconomic data series, the authors show that in a real-time forecasting exercise using 40 indicators to extract factors resulted in at least equivalent (if not better) results to using nearly 150 data series. In other words, N need not be excessively large for reasonable estimates. All of these concerns lead to questions over the 'sweet' spot for the size and composition of the data used to create the factor estimates. And these were firmly in mind when selecting the 24 indicators used in the empirical applications below.

2.3 The Bloomberg Consensus

In addition to the statistical models, which require no formal judgement, the Bloomberg market consensus view is considered. This may be viewed as a tough benchmark to beat see e.g. Bragoli, Metelli, and Modugno (2014) for a specific example using the poll as a comparator to statistical nowcasting models.

The consensus is the median of various institutional and private sector forecasts of quarterly changes in GDP as provided to Bloomberg's polling unit in the week prior to the preliminary release of GDP. The polling days are typically over Wednesday-Friday, with early responses uploaded to Bloomberg's terminals on the Thursday, with the remaining responses added Friday afternoon.

Again consider nowcasts for Q1 2014. Based on the calendar of releases over this period, the preliminary estimate for GDP was provided by the Office for National Statistics (ONS) on 29 April 2014. The early consensus view is made publicly available on 24th April, with the final reading released on the afternoon of April 25th.

There are several important points to bear in mind surrounding what the consensus forecast is, and how it should be viewed relative to the statistically driven nowcasts:

1. The consensus is commonly referred to as a forecast. But using the Bańbura and Rünstler (2011) definition of a nowcast as "…the prediction of the present, the very near future and the very recent past…"

then the consensus comes under the umbrella of nowcasting.

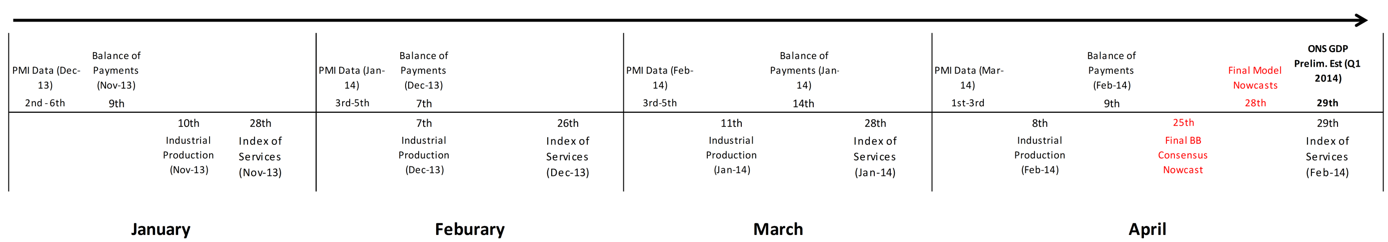
2. This Bloomberg "nowcast" will be based on very similar, if not the same, information (in terms of data) as the automated mechanical models. Looking at the calendar of releases then one of the last "major" pieces of data made available to the public are UK trade statistics. These are released around two-to-three weeks before the GDP preliminary estimates and will be broadly known by those being polled in the Bloomberg survey (see the timeline of major releases outlined in figure 1). However, it is assumed that contributors to the consensus may exercise a degree of judgement in their forecasts, perhaps incorporating soft information such as changes in the weather. Such information is not easily absorbed by automated statistical model procedures.

3. The Bloomberg consensus should be treated as the final GDP estimate for the quarter from the institutions and private sector economists that partake in the poll. Although earlier estimates will have been made, then refined in line with the accumulation of new data through the quarter and may be available in some consensus form via Bloomberg's monthly surveys, the timing of the final poll and the release of its findings means this is the summary of the final best guesses of UK GDP growth for a specific quarter. The statistical models are run as if nowcasting UK GDP around a week prior to the first estimate so these should be viewed as being broadly comparable to the consensus.

3 The Dataset and the Jagged Edge

From the perspective of timing, the nowcasts for the empirical application are performed around a week before the release of the preliminary estimate of GDP, which for the UK is within three-to-four weeks following the close of a calendar quarter. However, the standard bridging equation framework implies that all readings for a full calendar quarter are available. This allows the monthly series to be transformed into quarterly time-series as per equation 3. But due to lags in data publication and the non-synchronous nature of releases, data for the indicators are available at different times and a full set of observations may not be readily available prior to the publication of GDP. The result is what Giannone, Reichlin, and Small (2008) refer to as the dataset's "jagged" edge. There is therefore a need to "fill in the gaps".

Figure 1: Nowcasting UK GDP Timeline for Q1 2014



These issues have been sidestepped to a degree by the timing profile of the indicators that will be used to nowcast GDP. Variables broadly exhibit the characteristics of timeliness and non-revision (or only very minor at best) so a full set of observations for the vast majority of indicators are available at the time the nowcast is run. Moreover, the dataset covers a broad range of economic variables: indirect measures of GDP components (largely through business surveys), developments in the labour market (e.g. surveys and claimant count data), changes in house prices and influences on the economy that seem to be of interest to practitioners at central banks: economic uncertainty (Haddow and Hare 2013) and financial conditions (Angelopoulou, Balfoussia, and Gibson 2013).

However, there are several exceptions. At the time the preliminary estimates of GDP are released, publicly available information for trade and industrial production covers the first two months of a quarter, while the index of services is just one.

A solution to dealing with these missing observations could be to forecast using some kind of auxiliary modelling. Following Rünstler et al. (2008), univariate AR models with a maximum lag of 12 are used (lag lengths are determined by the AIC). The results of the auxiliary forecasts are then combined with observations already available just prior to the preliminary estimate of GDP and then entered into an OLS regression as per equation 2. The high frequency variables are subsequently transformed into a quarterly time series to match the time frequency of GDP observations.

In contrast, the MIDAS regressions can be adopted in line with data availability: there is no need for the extrapolation of "missing values" to deal with the dataset's jagged edge, with rebalancing essentially achieved by shifting the time series of respective explanatory variables forward (the parameters of equation 4 are dependent on h, which reflects the difference between the forecast target period and the most recent observation of the indicator).

In theory, this represents an improvement over the bridging equation methodology where forecasting regressions are used to fill in missing observations. Such an approach may introduce additional uncertainty into the GDP nowcasting equation if the auxiliary models are specified. However, note separate MIDAS regression continuously have to be calculated as h varies and new data points are observed.

For the factor models, which requires the creation of a synthetic monthly series, a slightly more sophisticated approach to dealing with missing observations has been adopted. Based on an imputation method available within the MATLAB statistical software package, missing values are automatically generated and imputed from a weighted average of the normalised values of the top 25% "nearest neighbours" (to perform a principal component analysis note all series are normalised to mean zero and variance of one before extracting component extraction). The "nearest neighbours" are determined by those that have the highest correlations (R^2 statistic) with the target series. Those with the highest correlation subsequently have the largest weight.

As an example, the UK manufacturing PMI tends to have a high correlation with industrial production data. Because of the timeliness of the PMI numbers, observations are available some six weeks before equivalent industrial production data. By using the latest normalised value of the PMI (plus those for other relevant series) this cross sectional information is exploited to support the forecasting of missing industrial production values.

The utilisation of cross sectional data in the estimation of missing information was viewed as an attractive characteristic of an imputation method, and seen as a way to potentially strengthen the estimations made from the relatively naive approach of relying on auto-regressions traditionally used in bridging equations.

Moreover, a broadly automated set-up, especially where the method can make use of existing "off the shelf" software, offers easier computations of the missing data compared to having to set-up and calculate individual regression models for a number of variables.

An alternative to an imputation method could be to use Kalman filtering or the expectation-maximisation algorithm popularised elsewhere in nowcasting see e.g. Giannone, Reichlin, and Small (2008) and Bańbura and Modugno (2014).

It was hard to determine whether there would be a vast practical difference between these approaches in this particular nowcasting application, especially as they share similar characteristics: both are designed to deal with missing data in an automated fashion and utilise the dataset in a cross-dynamic dataset.

However, as with any estimation procedure, additional uncertainty could nonetheless be introduced into the model set-up through such an iterative approach (unlike aforementioned direct estimation tools such as MIDAS regressions which circumvent these).

Still, forward projections of variables were made over a relatively limited time-frame (one or two months of missing data) so it seems likely the impact of model mis-specification errors would be limited.

On balance, a desire to use a method not seen elsewhere in nowcasting (to the best of the author's knowledge) tipped the favour in using imputation techniques in this particular application. There is also the potential for comparing the relative performances of these methods in nowcasting applications, and this is left for future work.

Some further notes on the dataset.

First, data history. Several of the business surveys start around 1996/1997. The sample is therefore split into two parts, with in-sample regressions and models created on data from 1998 to the end of 2005. Nowcasts are then created on a recursive basis once a quarter from 2006-2013. Out-of-sample real-time model nowcasts are subsequently assessed against the equivalent preliminary estimate of GDP through a root mean squared forecasting error (RMSFE) statistic.

Secondly, real-time assessment includes striving to replicate the dataset available at the time the nowcast is made, not just its structure but also in terms of actual data availability. So vintage series are utilised for several indicators which are subject to heavy revision. These include GDP itself, industrial production, the index of services, retail sales and trade statistics. For the first four mentioned, the source of real-time data is the excellent revisions triangles databases provided by the UK Office for National Statistics. For trade, the OECD's revisions database is used.

Finally, all the data are transformed where appropriate to ensure stationarity. A full list of the data sources and transformations is provided in table 1.

To summarize, the following models are ran in replicated real-time, with nowcasts produced on an information set that would be available on the day before the release of the first estimate of UK GDP:

- Five individual bridging equations and the mean of their respective nowcasts. These are conducted with an AR(1) component included in the regression equation, but this feature is also turned off to assess comparative performance and the contribution of this element.

- Similarly, five individual MIDAS equations plus the mean of the nowcasts. Again, the AR(1) component is turned on and off.

- An extracted factor that is used in a bridging equation, with one, two and a rule-determined number of factors retained for comparison. The

Table 1: Dataset Description

Variable Name	Frequency	Transformation	Type	Source
Industrial Trends: Volume of output, next three months	Monthly	n/a	Survey	CBI
Industrial Trends: total order books, current situation	Monthly	n/a	Survey	CBI
Distributive Trades, Retailing, Volume of sales for time of year	Monthly	n/a	Survey	CBI
UK Services PMI: Business Activity	Monthly	n/a	Survey	Markit Economics
UK Services PMI: Business Expectations	Monthly	n/a	Survey	Markit Economics
UK Construction PMI: Business Activity	Monthly	n/a	Survey	Markit Economics
UK Manufacturing PMI	Monthly	n/a	Survey	Markit Economics
Consumer Survey: Total, Confidence Index, Balance, SA	Monthly	n/a	Survey	GfK
Consumer Survey: Unemployment Over Next 12 Months	Monthly	n/a	Survey	GfK
RICS Housing Market, Price, England and Wales	Monthly	n/a	Survey	RICS
Economic Policy Uncertainty Index	Monthly	n/a	Derived Index	Economic Policy Uncertainty
Report on Jobs: Permanent Staff Placements	Monthly	n/a	Survey	KPMG, REC
UK Unemployment Rate: Claimant Count Measure	Monthly	n/a	Labour Market	Office for National Statistics
House Prices, Halifax, SA, Index	Monthly	Annual % Change	Price	Halifax
House Prices, Nationwide, SA, Index	Monthly	Annual % Change	Price	Nationwide
FTSE, All-Share Price Index, Return, Close, GBP	Daily	3-month% Change	Financial Markets	FTSE International Ltd.
United States Volatility Index (VIX), Close, USD	Daily	n/a	Financial Markets	Reuters
Effective Exchange Rate Index	Daily	Annual % Change	Price	Bank of England
Eurozone Composite PMI	Monthly	n/a	Survey	Markit Economics
US Manufacturing PMI	Monthly	n/a	Survey	Institute for Supply Management
Industrial Production	Monthly	3-month % Change	Real, Hard	Office for National Statistics
Index of Services	Monthly	3-month % Change	Real, Hard	Office for National Statistics
Retail Sales	Monthly	3-month % Change	Real, Hard	Office for National Statistics
Export Trade	Monthly	3-month % Change	Real, Hard	OECD

AR component is again included and also excluded.

- The same factor specifications, but also used in MIDAS equations.

Performance of these models, along with the Bloomberg consensus view, is assessed against a simple benchmark AR(1) model via Root Mean Squared Forecasting Error (RMSFE) statistics.

4 Empirical Results

In this section, the accuracy of the various statistical models in nowcasting preliminary estimates of UK GDP, primarily against the benchmark AR(1) model, is presented. This is is done through the ratio of model and benchmark RMSFEs. A reading greater than one signals model under-performance (i.e. nowcasts are, on average, further away from first GDP estimates than the benchmark), while a reading lower than one indicates out-performance (i.e. the model is closer on average than the benchmark in nowcasting GDP).

Section 5.1 assesses the best performing models, and includes a discussion on the benefit of pooled nowcasts alongside a deeper look at the retained principal components within the monthly factor model. Section 5.2 looks specifically at the relative performances of the bridging equations and MIDAS-based models. Section 5 concludes with an examination of model performances relative to the market consensus view.

4.1 Modelling Accuracies

Table 2 shows the results of the real-time nowcast modelling exercises over the full sample period 2006Q1-2013Q4.

With the exceptions of two of the four manufacturing PMI-based models, the benchmark is generally beaten, and at times substantially so. Although there is some variation, the out-performance can be in the region of 20-25%. This indicates the additional information provided by the timely and high frequency explanatory variables, either individually, through pooled forecasts, or within derived factors are useful and add value when nowcasting GDP. Such a finding resonates loudly with the nowcasting literature.

Naturally there tends to be variance in performance across the sample period. For instance, the AR(1) is difficult to beat during periods of economic stability e.g. 2006-2007, the years immediately preceding the financial crisis.

Table 2: Root Mean Squared Forecast Error (RMSFE) Ratios

Bloomberg Consensus	0.44

	Bridging Equations		MIDAS	
Model	Exc AR	Inc AR	Exc AR	Inc AR
Manufacturing PMI	1.08	0.92	1.15	0.94
Construction PMI	0.76	0.75	0.80	0.78
Services PMI	0.91	0.80	0.93	0.82
Industrial Production	0.88	0.68	0.84	0.81
Index of Services	0.92	0.94	0.84	0.82
Pooled Nowcast	0.79	0.76	0.81	0.78
Factor (r=1)	0.85	0.81	0.94	0.82
Factor (r=2)	0.77	0.76	0.81	0.78
Factor (r= rule)	0.80	0.78	0.87	0.83

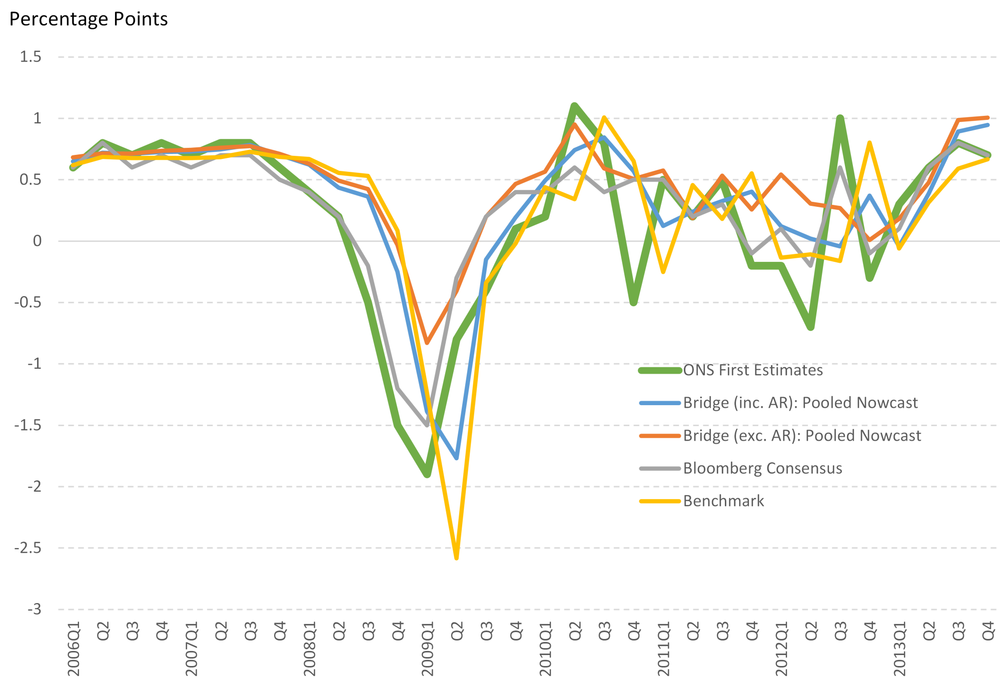
Notes: The table shows the ratio of the Root Mean Squared Forecast Error (RMSFE) for each model to the benchmark RMSFE over the period 2006Q1-2013Q4 when making nowcasts of quarter-on-quarter changes in GDP (the benchmark is a simple AR(1) model). A reading greater than one signals model under-performance (i.e. nowcasts are, on average, further away from first GDP estimates than the benchmark AR(1) model), while a reading lower than one indicates out-performance (i.e. the model is closer on average than the benchmark in nowcasting quarter-on-quarter changes in GDP). The dependent variable that is being nowcast is the first estimate of GDP growth as provided to users in real-time.

As the volatility of GDP increased with the onset of recession from late 2008 onwards, the benchmark performs worse and is easily outperformed by the statistical models (although all show a deterioration in absolute terms). This corroborates findings elsewhere for UK GDP nowcasting e.g. Mitchell (2009). While results are not shown in table 2, figure 2 provides a visualisation of selected model nowcasts relative to ONS preliminary estimates.

The weakness of the AR(1) benchmark is (not surprisingly) especially evident around turning points such as the trough of the severe recession in Q1 2009 and at other times of economic volatility (e.g. 2010-2012). These problems are also evident when an AR(1) component is incorporated into respective nowcasting models (see e.g blue line in figure 2).

When the AR(1) component is turned off, turning points are captured more easily and issues of lag dissipate. But magnitudes of change during

Figure 2: Selected Model Nowcasts Relative to ONS Preliminary GDP Estimates (Quarter-on-Quarter Growth)



large movements in GDP such as the sharp downturn in late 2008/early 2009 are not well captured (see e.g. red line in figure 2). The opposing forces of including and excluding the AR(1) component tend to offset so, over the sample period as a whole, there is relatively little difference in respective accuracies.

Turning to direct comparisons of the various statistical models, there is little performance difference. Pooled forecasts from a small set of individual bridging equations and a factor approach retaining two principal components (with both models including auto-regressive components) registered the lowest RMSFEs relative to the benchmark. Several notable points come to the fore.

Firstly, when using a small sub-set of models, the nowcasting of GDP is enhanced by taking some kind of average. Inevitably, there will be individual models that beat the pooled nowcast over the sample period. These are, notably, the Construction PMI and Industrial Production based models. But relying on relationships between dependent and single explanatory variables to not break down seems dangerous: the accuracy of the Construction PMI model proved to be relatively uneven compared to other models during 2013 and its own performance earlier in the sample period. The safer approach is to take the pooled nowcast, which tends to yield better performance on average. This again resonates with the literature.

Secondly, the results for the factor models indicate that UK macroeconomic performance tends to be summarised best by a small number of derived factors, in this case just two.

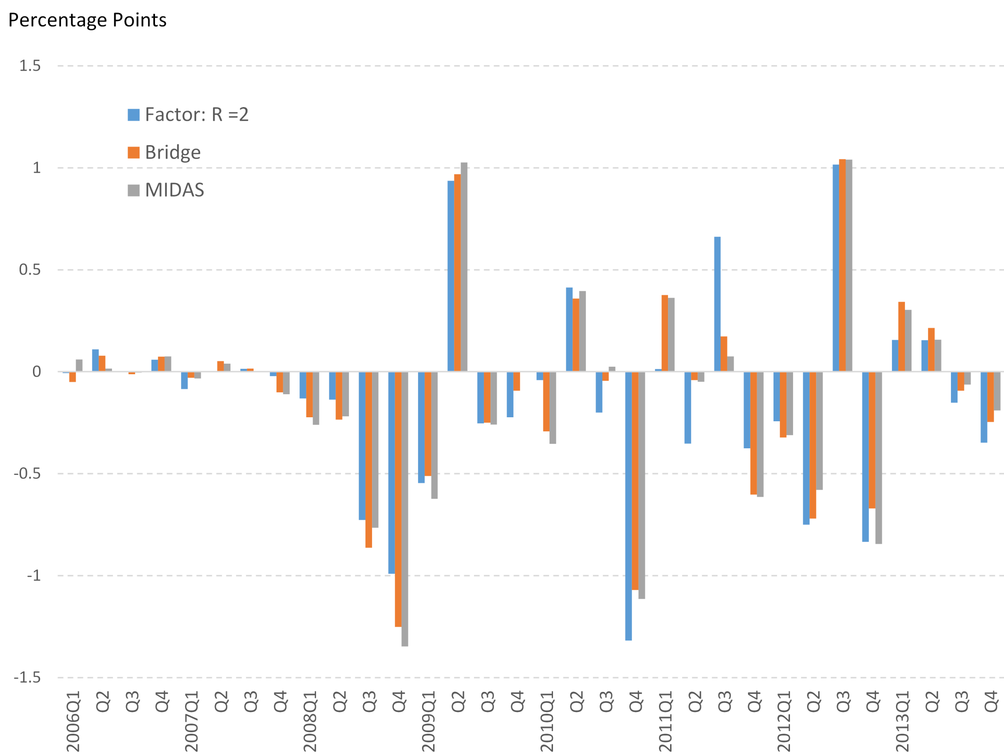
Notably, the first principal component is found to be most closely related to the Markit UK Services PMI and the KPMG/REC Demand for Staff variables, indicating these two diffusion indicators provide excellent summaries of general underlying changes in macroeconomic performance. Andreou, Ghysels, and Kourtellos (2013) and Lombardi and Maier (2011) provide similar conclusions with PMI data. Perhaps reflective of the importance of consumption and the housing market to the UK, the second principal component is found to be most closely related to indicators such as GfK consumer confidence and house price indices provided by Nationwide and Halifax.

Finally, as tends to be the case in any nowcasting application, model performances vary over time. Whereas the factor model with two retained principal components performs best through the more extreme parts of the financial crisis, registering the lowest nowcasting errors from late 2007 through to the emergence from the deep recession in mid-2009, performance thereafter has been rather uneven, particularly through much of 2011 and 2012, a period of notable swings in UK GDP. Errors for the MIDAS and bridging equations were on average lower than the best factor model throughout this period. See figure 3 for an illustration.

4.2 Bridging versus MIDAS Regressions

MIDAS specifications slightly under-perform relative to bridging equation models. They certainly show no out-performance, a finding recently corroborated by Schumacher (2014) when performing a similar model comparison exercise for euro area GDP. This is despite MIDAS arguably having clear desirable statistical features such as the non-forecasting of missing observations and no potential loss of important information from unweighted averaging of

Figure 3: Model Nowcast Errors



high frequency observations: these advantages don't seem to clearly translate when linking between monthly and quarterly variables in this particular real-time application. Indeed, where there is no extrapolation of missing values, such as in the PMI models, the unrestricted weighting scheme of the monthly observations appears if anything to be disadvantaged.

Less clear cut, though, is when the missing observations have to be forecast which was a key feature of the bridging equation framework's ability to make timely GDP nowcasts. Ambiguity flows from the nowcasting results for the IP and IoS models, which respectively required one and two months of observations to be estimated: on the one hand, the IP bridging equation model with an AR component considerably outperforms its MIDAS counterpart. But the roles are reversed when looking at IoS specifications: MIDAS is the better performer.

Perhaps this is a function of the simplistic nature of the forecasting of missing observations: an auto-regressive function was used to "fill in the gaps". Maybe exploiting already available data in a cross sectional sense (such as the timely business survey data) would provide better estimates of these and offer improved nowcast results for the IP and IoS models.

Nonetheless, whenever missing data are forecast, then an additional layer of uncertainty is inevitably imported into the nowcasting regressions. And with the pooled results of the individual bridging and MIDAS models barely distinguishable over the sample period, the benefits and flexibility of utilising MIDAS regressions remain persuasive: there is the option of including higher frequency explanatory data such as weekly or daily data, while lags of the explanatory variables could be easily incorporated in a MIDAS model set-up (although it is not immediately clear why one would include lags when using contemporaneous indicators to measure current changes in GDP).

4.3 The Consensus Nowcast

A standout result from table 2 is that no model outperforms the consensus of economists polled by Bloomberg. The consensus has a 56% out-performance over the benchmark, some 30 percentage points better than the best performing statistical nowcasting models. In absolute terms, the RMSFE is 0.32 percentage points, which compares to 0.66 percentage points for the benchmark and around 0.50 percentage points for the strongest performing statistical models.

This suggests there is considerable value added through the consensus, which contrasts to other studies where model-based, statistically driven, nowcasts are shown to be performing just as well e.g. deWinter (2011) and Bańbura et al. (2013). Notably for the UK, the results are broadly consistent with recent research by the Bank of England (2014) which showed that the Bank's own staff forecasts tend to outperform mechanical nowcasting models. There may be several explanations for the strong performance of the consensus, in particular:

- Evidence of consensus beating performance has generally been rooted on samples that are dominated by the Great Moderation and may not include (or only just include) elements of the financial crisis which had a dramatic impact on the volatility of economic output. For instance, Bańbura et al. (2013) cover the period 1995-2010 for US GDP.

And when conditioning model and consensus-based nowcasts explicitly for information availability, Liebermann (2014) finds the consensus performs just as well, if not slightly better than automated models. Moreover, Bragoli, Metelli, and Modugno (2014) report broadly similar performances between institutional and model nowcasts for Brazil between 2007 and 2013, while Higgins (2014) discovers its hard to beat the consensus view using similar techniques to that of Giannone, Reichlin, and Small (2008) when performing US GDP nowcast horse-races over the period 2011-2014. Barring the first two years, the majority of the out of sample nowcast testing in this paper covers a similar period of unprecedented swings in UK economic performance.

- deWinter (2011) seems a notable exception, where the performance of private sector forecasts against statistical models in nowcasting Dutch GDP is explicitly modelled in periods of crisis. The conclusion is that augmenting a purely statistical procedure with judgement adds little value. Recently, Jansen, Jin, and deWinter (2014) argue that professional forecasts, while offering some positive results tend to perform poorly when compared directly to model nowcasts. However, the research is provided with the caveat that real-time data sets were not utilised (so revisions to variables were not incorporated). This leads to some concern whether comparisons against the consensus view were fair, given that revisions to GDP data can be large. Further exploration of these features is provided in section 6, but note if the consensus performance was compared against the latest GDP vintage (rather than real-time information) the nowcast accuracy quoted in table 2 would deteriorate by nearly 50%. This suggests that statistical model comparisons against consensus views should be conditioned on exactly the same information i.e. the data that was available in real-time to professional forecasters should also be used to construct the statistical model nowcasts and is important when comparing respective predictive GDP accuracies.

- Outside of the financial crisis, there have been several instances in recent years of what may be referred to as UK specific "special events" which led to additional volatility in the quarterly GDP data. Notable special events include the Royal Wedding in April 2012 and the London Olympics which followed in July/August of 2012. These events drove

> sharp changes in output that proved difficult for mechanical models to pick-up. A degree of "judgement" and the drawing of information not easily incorporated into a model set-up probably proved a sounder strategy during this period.

As a final remark here, the consensus is, of course, not correct all of the time: even the experts can be wrong-footed. For example, in Q4 2010 heavy snowfall had a large disruptive impact on economic activity leading to a -0.5% decline in GDP against expectations of a rise in GDP of +0.5%. This nowcast error was the largest recorded for the consensus throughout the sample period.

5 Notes on UK GDP Revisions

When running the real-time nowcasting simulations, a notable observation was that UK GDP experiences substantial revisions. Economic history is constantly being rewritten.

These revisions must inevitably impact on historical relationships with explanatory variables, especially those that are unrevised such as the business surveys. Coefficients in nowcasting equations will be unstable, which could have a detrimental impact on nowcasting model accuracy: seemingly good model performance can turn poor following the release of a new GDP vintage (and vice versa).

In this section, some background is provided on the evolution of trends in quarterly GDP and the sources of revisions. Then the results of re-running some of the statistical models presented in section 3.4 are provided: the difference is contemporaneous changes in GDP are nowcast by using, as the dependent variable, a synthetic series built purely from preliminary estimates of quarterly changes in GDP. Crucially this series is not subject to revisions through time.

5.1 GDP Vintage Evolutions

A visualisation of the evolution of various vintages of quarterly changes in GDP from January 2010 through to September 2014 (the latest vintage) is provided in figure 4. To observe these evolutions, it should be read top row, left to right, followed by the middle row, left to right etc.

Two reference series are also provided in the figure: the first published preliminary estimates of quarterly changes in GDP and quarterly averages of the monthly UK Services PMI, which is never revised and was shown to provide a good overview of underlying macroeconomic conditions.

From the top left quadrant, which shows changes in quarterly GDP as published in January 2010 against equivalent first estimates and the UK Services PMI, there are several observations.

Firstly, revisions from preliminary estimates of GDP in the early years of the plot moves the implied path of the economy further away from that signalled by the business survey data: the preliminary estimates of GDP in 2002-2003 suggested slower growth of the economy, which matched the easing of activity signalled by the PMI. In the January 2010 vintage, however, the economy was estimated to have been growing at an accelerated rate over this period.

Secondly, the profile of the sharp downturn indicated by first estimates and the January 2010 vintage started and peaked later than implied by the PMI. Whereas the official data suggests that the economy continued to grow markedly at the end of 2007, the PMI pointed to a sharp deceleration which pre-empted the onset of recession the following year. The low point of the recession was signalled by official data in Q1 2009, with a quarterly fall in GDP of over 2%. But the PMI, in contrast, indicated the business cycle had already turned up in early 2009.

Finally, the PMI pointed to an earlier emergence from recession than the official data, with the PMI implying that the economy was growing strongly in the second-half of 2009. In contrast, GDP data suggested stagnation of output and the UK was struggling to emerge from recession.

Moving through the various GDP vintages, a number of developments related to these initial observations emerge. Focussing primarily on 2002-2007, the trend in economic output for this period shows an increasing divergence from those paths indicated by the preliminary GDP estimates and the PMI survey. Indeed, at the time of the 2013 GDP vintage, 2002-2007 shows a period of rather uneven GDP growth that is barely recognisable to that indicated by the January 2010 vintage and those provided in real-time.

If 2002-2007 was characterised by moving further away from trends implied by the survey data, then the period that encapsulates the downturn and subsequent emergence from recession in 2008-2010 shows GDP revisions moving the path of the economy closer to that of the business survey data. By the start of 2013, the sharp downturn in the business cycle indicated by the PMI and the January-2013 vintage GDP series occurs in broadly similar positions (late 2007), with the business cycle turning point in Q1 2009 and the emergence from recession occurring in Q3 of the same period.

Although these estimates of turning points and emergence from recession were unchanged in the latest vintage (September 2014), data prior to the downturn in 2008 have again been revised heavily – and seemingly much further away from the trends indicated by the business surveys over the period 2002-2007.

5.2 Sources of Revisions

That GDP data are subject to revision is well known and there is a rich literature on the sources, predictability and modelling impacts of such revisions. Croushore (2011) provides an extensive survey, with historical monetary policy analysis and forecasting model evaluation all reported to be impacted by revisions. Tkacz (2010) highlights the non-trivial nature of revisions to GDP when measuring in real time the output gap, a widely used determinant of future inflation.

Brown et al. (2009) and Murphy (2009) provide some background for the sources of revisions to UK GDP. The preliminary estimate of GDP, which is produced three weeks or so after the end of a calendar quarter, contains just 40% of the data required to produce a "final" estimate. First estimates of GDP will therefore reflect a combination of hard data (usually based on sample surveys) complemented by forecasts for missing data values, particularly for the period towards the end of the quarter, when hard information are particularly scarce. As time goes by, however, forecast values are replaced by new source information and the need for forecasting diminishes. An example is the receipt of data from annual surveys or administrative sources, which provides the basis for annual benchmarking and quarterly data re-alignment.

Methodological improvements in how the ONS measures the economy can also occur. The move to annual chain-linking in 2003 is one example. Recognizing the flawed nature of reviewing "fixed" weights of GDP components only once every five years, which would mean dynamic changes within the economy would not be captured, the ONS switched to a chain-linking procedure which enabled these weights to be adjusted on an annual basis (Robojohns 2006). More recently, a wide range of changes, driven in the main by a shift to the European System of Accounts 2010 (ESA 2010) have led to changes in the interpretation (and subsequent quarterly and annual estimates) of a wide-range of macro-economic aggregates such as the measurement and treatment of spending on R&D (ONS 2014). Keeping such methodological changes in mind, quarterly GDP estimates are subsequently subject to ongoing revision and may never be considered "final".

There have been several attempts to model these revisions, although the literature is rather ambiguous on the ability of statistical models to do so with any considerable success. Cunningham et al. (2009) provides a notable attempt to predict UK GDP revisions using a signal extraction model that utilises historical observations (such as serial correlation within the revisions) augmented with data from private sector business surveys. This forms the basis of how the Bank of England (Cunningham and Jeffery 2007) deals with uncertainty around early GDP estimates. Faust, Rogers, and Wright (2005) suggest that revisions to several G7 countries, including the UK, were highly predictable over the period 1967-1998 due to their "inefficiency".

5.3 Targeting the Preliminary Estimate of GDP

Regardless of predictability and sources of revisions, the changing profile of economic history leads to a concern that relationships between GDP and explanatory data sources, such as business surveys which tend not be revised, are constantly in a state of flux. Coefficients within regression equations linking the two series may change substantially with the release of new GDP vintages. Considerable swings in model performance may result.

This provided motivation to re-assess the performance of the nowcasting models but with the real-time vintages of GDP replaced with a series that is stable and not subject to revision.

To meet this requirement a series was used that purely took the first estimates of quarter-on-quarter changes in GDP (note this series was used in figure 4 for illustration). This involved downloading the requisite spreadsheet from the excellent GDP revisions triangles and real-time database provided by the Office for National Statistics (ONS). Within this spreadsheet (presently named "Quarterly GDP at Market Prices (ABMI)"), the ONS provide a time series called "Month 1 estimate". This series measures all of the first approximations of quarter-on-quarter movements in GDP for each quarter since 1993.

The motivation for using such a series was to offer some stability to the left-hand side of the various regression equations.

Datasets used for the explanatory variables were the same, the modelling process was unchanged and the results were based on the same out-of-sample testing period of 2006Q1-to-2013Q4. However, note that the AR components were derived from the new series of preliminary GDP estimates. Moreover, given the largely non-distinguishable nature of the performance of the MIDAS and bridging equations from section 5, nowcast regressions were only produced for the latter. The results of this exercise are shown in table 3.

The pooled nowcasts of the five individual models are better at predicting preliminary estimates of GDP than those that were conducted in the real-time simulation of section 5. The improvement is in the region of 10 percentage points, with the actual RMSFE for the models that include AR components dropping from 0.50 to 0.45. However, a Diebold-Mariano test for predictive accuracy suggested that the difference was insignificant.

In contrast, statistically significant differences at the 5% level were found with the Manufacturing and Services PMI models. These both showed a considerable strengthening in accuracy over the sample period when switching to using preliminary estimates of GDP as the dependent variable. The Services PMI model (excluding the AR component) was the best performing out of all models on pure RMSFE ratio grounds, out-performing the naive real-term benchmark model by 34 percentage points (though this still remains some way off the performance of the consensus). Moreover, comparing nowcast errors against those from the equivalent real-time simulation exercise showed considerable out-performance during 2008 to early 2011, but less so in 2012 and 2013 when the real-time exercise performed on average a little better.

Nonetheless, discovering that two key and closely watched business surveys - the UK Manufacturing PMI and the UK Services PMI - provide significantly better nowcasts for preliminary GDP estimates than in the real-time modelling simulation exercise implies that either (i) the shifting nature of the GDP series does impair nowcasting performance or (ii) later-published GDP vintages offer information over and above those of the business surveys.

On the one-hand, figure 3 suggested that revisions during a period that is widely viewed to be a relatively benign economic environment moved the trends signalled by the GDP series and the PMI business surveys further away from each other. But during the deep recession, the business surveys provided information that suggested an earlier downturn and earlier emergence from recession than first indicated by the official GDP series. Subsequent revisions have brought these relationships closer into line.

So the surveys provide an early steer on first estimates during periods of relative economic calm, but later vintages add more colour. In contrast, at

Table 3: RMSFE Ratios - "Month 1" GDP Series

	Bridging Equations	
Model	Excluding AR	Including AR
Manufacturing PMI	0.75	0.73
Construction PMI	0.72	0.71
Services PMI	0.66	0.69
Industrial Production	0.87	0.74
Index of Services	0.73	0.72
Pooled Nowcast	0.67	0.68
Factor (r=1)	0.80	0.70
Factor (r=2)	0.82	0.67
Factor (r= rule)	0.85	0.66

Notes: The table shows the ratio of the Root Mean Squared Forecast Error (RMSFE) for each model to the benchmark RMSFE over the period 2006Q1-2013Q4 when making nowcasts of the quarter-on-quarter changes in GDP (the benchmark is a simple AR(1) model). A reading greater than one signals model under-performance (i.e. nowcasts are, on average, further away from first GDP estimates than the benchmark AR(1) model), while a reading lower than one indicates out-performance (i.e. the model is closer on average than the benchmark in nowcasting quarter-on-quarter changes in GDP). The dependent variable that is being nowcast is the "Month 1" GDP series as provided by the Office for National Statistics revisions triangle database.

times of rapid change, the surveys seem to offer a timely assessment of what is truly happening.

This may well reflect the nature of what the two series are measuring.

The business surveys primarily measure changes in economic performance from a perspective of breadth. The greater the level of these diffusion indices are below or above some neutral point indicates that a greater proportion of companies are experiencing similar changes in their business performance (be it growth or contraction). In a broad-based economic event - such as a financial crisis - then many companies will have shared experiences. The surveys pick up a particular turning point in the economy in a timely fashion.

In contrast, GDP data measure quantitative changes in economic output. In some respects there is no conceptual reason why diffusion-based indices would map directly with GDP series and, at times of stability, subtle changes in economic performance may not be as well captured by the surveys. But GDP data may suffer from lags at times of rapid change due to the nature of its construction (being built on forecasted elements for missing observations etc).

Knowing the differences between surveys (timely, non-revised, but lacking in detail) and GDP (extensive, broad figures, but backward-looking and likely to be revised) are vital to the interpretation and understanding of these data sources at various points in the business cycle.

6 Summary

The question posed at the start of this paper was, out of a set of competing nowcasting techniques, which one peforms best at predicting preliminary estimates of UK GDP when they convergence around a week or so before the first GDP release? While there are pros and cons with each approach from theoretical standpoints, based on a purely practical perspective, it proved hard to distinguish between their respective performances during 2006-2013.

Such a statement is common in many forecasting applications; model selection tends to be based on best global performance over some defined out-of-sample period. "Best" usually involves the use of some loss function such as the RSMFE statistic combined with a test of comparative predictive ability such as that outlined by Diebold and Mariano (1995).

This therefore leads to some questions of robustness. As hinted by figure 3 the relative nowcasting ability across the models may vary across time. However, using the global RMSFE statistics guides the nowcaster to believe there is no discernible difference between models regardless of the forecasting environment.

A future extension of the research would be to therefore consider the potential for identifying such instabilities in a more formal manner.

With this in mind, statistical tests to find various model instabilities through their out-of-sample time paths could be deployed. For example, Giacomini and Rossi (2010) propose a fluctuation test to reveal whether a model performs better than a competitor in certain periods but less so in others.

Alternatively, one may conclude that the best protection against instability in the nowcasting performance of individual models may be to use some kind of combination, such as the pooling strategy proposed by Timmermann (2006).

Nonetheless, using a replicated real-time dataset, several general findings from the nowcasting literature were found to be applicable for UK GDP. High frequency data are important in reducing nowcast uncertainty; pooling of small nowcasting regressions tend to provide greater overall accuracy than single specifications; and business survey data provide a good summary of underlying economic performance.

A key takeaway, however, is that judgement has played a positive role in nowcasting UK GDP growth during a period of considerable economic upheaval. When based on similar information, the Bloomberg consensus significantly outperformed all statistical models confirming that the Bank of England and other institutional produced nowcasts (such as the monthly estimates of GDP provided by the National Institute of Economic and Social Research), which rely to some degree on judgement, follow optimum strategies.

Several features of UK GDP data are perhaps reasons why judgement remains important. GDP has shown considerably greater variance than before the onset of the Great Recession and appears to have followed a more volatile economic path than (say for example) the Eurozone. Moreover, the GDP data have been prone to considerable revision, making relationships with indicator variables susceptible to variation over time. With this in mind, adapting a series for the dependent variable to one that provides stability with other data less prone to revision led to a net gain in the performance of a small set of pooled nowcasts.

Despite this improvement in relative accuracy, the new set of nowcasts nonetheless proved insufficient to beat the consensus forecast, while questions remain on the true underlying relationship between GDP and business survey data at different points in the economic cycle.

References

Aastveit, K A and T G Trovik (2008). “Nowcasting Norwegian GDP: The role of asset prices in a small open economy”. *Norges Bank, Working Paper* 2007/09.

Aiolfi, M, C Capistrán, and A Timmermann (2010). “Forecast Combinations”. *Banco de México Working Papers* 2010-04.

Andreou, E, E Ghysels, and A Kourtellos (2013). “Should Macroeconomic Forecasters Use Daily Financial Data and How?” *Journal of Business and Economics Statistics* 31.2, pp. 240–251.

Angelopoulou, E, H Balfoussia, and H Gibson (2013). "Building a Financial Conditions Index for the Euro Area and Selected Euro Area Countries: What Does It Tell Us About the Crisis?" *European Central Bank Working Paper Series* 1541.

Armesto, M, K Engemann, and M Owyang (2010). "Forecasting with Mixed Frequencies". *The Federal Reserve Bank of St. Louis Review* 92.6, pp. 521–536.

Baffigi, A, R Gonelli, and G Parigi (2004). "Bridge Models to Forecast the Euro area GDP". *International Journal of Forecasting* 20.3, pp. 447–460.

Bai, J and S Ng (2002). "Determining the Number of Factors in Approximate Factor Models". *Econometrica* 70.1, pp. 191–221.

Bańbura, M, D Giannone, M Modugno, and L Reichlin (2013). "Nowcasting and the real time data flow". *ECB Working Paper Series* no 1564.

Bańbura, M and M Modugno (2014). "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data". *Journal of Applied Econometrics* 29.1, pp. 133–160.

Bańbura, M and G Rünstler (2011). "A look into the factor model black box. Publication lags and the role of hard and soft data in forecasting GDP". *International Journal of Forecasting* 27.2, pp. 333–346.

Bell, V, L W Co, S Stone, and G Wallis (2014). "Nowcasting UK GDP Growth". *Bank of England Quarterly Bulletin* Q1 2014.

Boivin, J and S Ng (2006). "Are More Data Always Better for Factor Analysis?" *Journal of Econometrics* 132.1, pp. 169–194.

Bragoli, B, L Metelli, and M Modugno (2014). "The Importance of Updating: Evidence from a Brazilian Nowcasting Model". *FEDS Working Paper* No. 2014-94.

Brown, G, T Buccellato, G Chamberlain, S D Chowdhury, and R Youll (2009). "Understanding the Quality of Early Estimates of Gross Domestic Product". *Economic and Labour Market Review* 3.12.

Camacho, M, G Perez-Quiros, and P Poncela (2013). "Short-term forecasting for empirical economists: A survey of the recently proposed algorithms". *Bank of Spain Working Paper* 1318.

Croushore, D (2011). "Frontiers in real-time data analysis". *Journal of Economic Literature* 49.1, pp. 72–100.

Cunningham, A, J Eklund, C Jeffery, G Kapetanios, and V Labhard (2009). "A State Space Approach to Extracting the Signal from Uncertain Data". *University of London School of Economics and Finance Working Papers* 637.

Cunningham, A and C Jeffery (2007). "Extracting a Better Signal from Uncertain Data". *Bank of England Quarterly Bulletin* Summer 2007, pp. 364–375.

deWinter, J (2011). "Forecasting GDP Growth in Times of Crisis: Private Sector Forecasts versus Statistical models". *DNB Working Papers* 320.

Diebold, F X and R S Mariano (1995). "Comparing Predictive Accuracy". *Journal of Business and Economic Statistics* 13, pp. 253–263.

ECB (2009). "Results of a special questionnaire for participants in the ECB Survey of Professional Forecasters".

Faust, J, J H Rogers, and J Wright (2005). "News and Noise in G7 GDP Announcements". *Journal of Money, Credit and Banking* 37.3, pp. 403–419.

Forini, C and M Marcellino (2013). "A Survey of Econometric Methods for Mixed Frequency Data". *Norges Bank Working Paper* 2013-06.

Ghysels, E, P Santa-Clara, and R Valkanov (2004). "The MIDAS Touch: Mixed Data Sampling Regression Models". *CIRANO Working Papers* 2004s-20.

Ghysels, E, A Sinko, and R Valkanov (2007). "MIDAS Regressions: Further Results and New Directions". *Econometric Review* 26.1, pp. 53–90.

Giacomini, R and B Rossi (2010). "Forecast Comparisons in Unstable Environments". *Journal of Applied Econometrics* 25.4, pp. 595–620.

Giannone, D, L Reichlin, and D Small (2008). "Nowcasting: The real-time informational content of macroeconomic data". *Journal of Monetary Economics* 55.4, pp. 665–676.

Haddow, A and C Hare (2013). "Macroeconomic Uncertainty: What is It, How Can We measure It and Why Does It Matter?" *Bank of England Quarterly Bulletin* Q2 2013.

Higgins, P (2014). "GDPNow: A Model for GDP Nowcasting". *Federal Reserve Bank of Atlanta Working Paper Series* Working Paper 2014-7.

Ingenito, R and B Trehan (1996). "Using Monthly Data to Predict Quarterly Output". *Federal Reserve Bank of San Francisco Economic Review* 3, pp. 3–11.

Jansen, W J, X Jin, and M de Winter (2014). "Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts". *DNB Working Paper* No. 365.

Klein, L R and E Sojo (1989). *Combinations of High and Low Frequency Data in Macroeconometric Models*. Ed. by L R Klein and J Marquez.

Economics in Theory and Practice: An Eclectic Approach. Kluwer Academic Publishers.
Kuzin, V, M Marcellino, and C Schumacher (2011). “MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area”. *International Journal of Forecasting* 27.2, pp. 529–542.
— (2013). “Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries”. *Journal of Applied Econometrics* 28.3, pp. 392–411.
Liebermann, J (2014). “Real-Time Nowcasting of GDP: A Factor Model vs Professional Forecasters”. *Oxford Bulletin of Economics and Statistics* 76.6, pp. 783–811.
Lombardi, M J and P Maier (2011). “Forecasting economic growth in the euro area during the Great Moderation and the Great Recession”. *ECB Working Paper Series* 1379.
Marcellino, M and C Schumacher (2007). “Factor-MIDAS for now- and forecasting with ragged edge data: A model comparison for German GDP”. *Deutsche Bundesbank* Discussion Paper Series 34/2007.
Mitchell, J (2009). “Where are we now? The UK recession and nowcasting GDP growth using statistical models”. *National Institute Economic Review* 209, pp. 60–69.
Murphy, J (2009). “Revisions to Quarterly GDP Growth and Its Components”. *Economic and Labour Market Review* 3.4.
ONS (2014). “Economic Review”. *Office for National Statistics Economic Trends* September 2014.
Robojohns, J (2006). “Methodology Notes: Annual Chain-Linking”. *Office for National Statistics, Economic Trends* 630.
Rünstler, G, K Barhoumi, R Cristadoro, A D Reijer, A Jakaitiene, P Jelonek, A Rua, K Ruth, S Benk, and C V Nieuwenhuyze (2008). “Short-term forecasting of GDP using large monthly data sets: a pseudo real-time forecast evaluation exercise”. *European Central Bank, Occasional Paper Series* 84.
Schumacher, C (2014). “MIDAS and Bridge Equations”. *Deutsche Bundesbank Discussion Paper* No 26/2014.
Stock, J H and M W Watson (2002). “Macroeconomic forecasting using diffusion indexes”. *Journal of Business and Economic Statistics* 20.2, pp. 147–162.

Stock, J H and M W Watson (2011). *Dynamic factor models*. Ed. by M P Clements and D F Hendry. Chapter 2, Oxford Handbook on Economic Forecasting. Oxford University Press.

Timmermann, A (2006). *Forecast Combinations*. Handbook of Economic Forecasting. Elsevier.

Tkacz, G (2010). "An Uncertain Past: Data Revisions and Monetary Policy in Canada". *Bank of Canada Review* 2010 (Spring).

Yiu, M S and K K Chow (2011). "Nowcasting Chinese GDP: Information Context of Economic and Financial Data". *HKIMR Working Paper* No. 04/2011.

Figure 4: The Changing Profile of UK GDP History