Chasing Yesterday: Nowcasting Economic Activity with Timely Indicators

PhD Thesis submitted to the

Department of Economics at the University of Strathclyde

by

Paul Smith

March 2016

Declaration of Authenticity and Copyright

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date:

Acknowledgements

Working on this thesis has been a hard journey, with many bumps along the way. The following people have made the road to completion that little bit smoother.

First and foremost, gratitude to my supervisors Gary Koop and Rodolphe Desbordes who have not only offered expert advice and support throughout the research process, but also their wise insights on a variety of topics.

Secondly, I would like to thank all the students and staff that I have met and discussed my work with throughout the past four years, in particular Grant Allan, Julia Darby, Alex Dickson, Loe Franssen, Tobias Emonts-Holley, Stuart McIntyre, Andrew Ross and Twyeafur Rahman. During my visits to Glasgow I have always been made to feel most welcome and part of the economics department at Strathclyde.

Thirdly, acknowledgements also to Rob Dobson, Annabel Fiddes, Andrew Harker, Jack Kennedy, Oliver Kolodseike and Tim Moore; all work colleagues at Markit Economics. I appreciate the alternative perspectives on my research and the strong level of interest in my progress over the past four years.

Finally, I would never have completed this thesis without the love, support and patience from my wife Anna, to whom I am truly thankful.

Abstract

The thesis *Chasing Yesterday: Nowcasting Economic Activity with Timely Indicators* presents three separate essays rooted in the topic of nowcasting that have been written since 2013. A variety of research themes drawn from the nowcasting literature are covered, with the essays pulled together through an underlying link of the usefulness of timely economic indicators to policymakers, investors and researchers.

Following an introduction to nowcasting and the broad research themes covered in the thesis, Chapter 2 is titled "*The Importance of Being Timely*", a version of which has been recently published in the *Journal of Forecasting.* The research in the chapter is concerned with understanding the contribution quickly-released survey data make to tracking economic activity in nowcasting models. Generally speaking, policymakers want to know about real-time economy performance. However, closely watched macroeconomic time series produced by national statistics offices are published infrequently, with a time lag and are subject to revision. Such issues create uncertainty in tracking economic developments, a by-product of which is to raise the value of business and consumer surveys. Although providing less granularity than official data series, the surveys are released in a timelier manner and are generally not revised.

Using real-time data sourced from the Deutsche Bundesbank, the OECD and the Office for National Statistics, an assessment of the role that the popular and widely used Purchasing Managers' Index (PMI) play in reducing forecasting

errors in a simple "nowcasting" framework is undertaken. The empirical exercise is conducted for five developed economies and also covers the period of the Great Recession. The conclusion is clear: timing matters.

The third chapter "*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.

The final research chapter, "*Google's MIDAS Touch: Predicting UK Unemployment with Internet Search Data*", a version of which is due to be published in the *Journal of Forecasting*, changes tack somewhat by assessing the potential of internet search data as a useful source of information for policymakers when formulating decisions based on their understanding of the current economic environment. The chapter builds on earlier literature and the ideas generated in chapters 2 and 3 via a structured value assessment of the data provided by Google Trends. This is done through two empirical exercises related to the forecasting of changes in UK unemployment.

Firstly, economic intuition provides the basis for search term selection, with a resulting Google indicator tested alongside survey-based variables in a traditional forecasting environment.

Secondly, this environment is expanded into a pseudo-time nowcasting framework which provides the backdrop for assessing the timing advantage that Google data have over surveys. The framework is underpinned by a MIDAS regression which allows, for the first time, the easy incorporation of internet search data at its true sampling rate into a nowcast model for predicting unemployment.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Economists have something in common with meteorologists: both like to engage in long-term forecasting. However, being prescient about future events, which are dependent on complex natural or socially constructed systems is a tough business. And when unexpected episodes take the public by surprise – such as unpredicted weather storms or once in a generation financial crises – the public outcry over why forecasters didn't seem to "get it right" can be vociferous.

If there is a general struggle to successfully forecast what is going to happen over the medium- to long-term, then perhaps it's easier to focus on today or more pertinently *now*.

Being at the very short-end of the forecasting curve, meteorologists have come to describe such activities as *nowcasting*. Recently economists have become increasingly engaged in similar exercises.

But here is where divergences with meteorology and economics appear.

Whereas a weatherperson can look out of a window, observe current conditions and, probably with some confidence, make accurate inferences about weather patterns over the coming few hours, economists face much greater challenges. The economy is, unlike the weather, not directly observable. Moreover, data on economic performance can be slowly released and untimely. Understanding what is going on in the past, let alone today, is a major headache. Economists are it seems always *chasing yesterday*.

In this first chapter, the broad aims are to establish the general econometric challenges that nowcasters face, provide a high level overview of two main techniques that have emerged to deal with such issues, and also introduce the roles that various economic indicators play within nowcasting frameworks. This is followed by a discussion on how "big data" is beginning to shape thinking in the nowcasting space.

Throughout the following sub-sections, introductions to the work and research that constitute chapters two to four are also provided.

The thesis is, in essence, a collection of three separate essays rooted in the topic of nowcasting. While such an approach leads to an inevitable overlap in text and presentation between chapters, a variety of research themes drawn out from the nowcasting field are covered, based on a broad observation of the usefulness of timely economic indicators to policymakers. These people are charged with making important decisions on a high frequency basis, most obviously being the setting of monetary policy instruments to help aid current and future macroeconomic stability.

To support this process and make sound, optimal judgements, policymakers generally want to know where the economy is in the business cycle. For example, it makes sense if a central bank wishes to make projections about the price level (with an explicit aim of hitting some future inflation target) to have a good sense of the current starting point of the economy be it in terms of output or employment.

However, knowing "where we are now" is challenging.

The most widely used yardstick of activity, Gross Domestic Product (GDP), tends to be slowly released and is provided relatively infrequently compared to the policy-decision making process. Whereas meetings to set policy can be as often as once a month, GDP statistics are only available quarterly and inherently backward-looking. In a rather extreme situation, imagine if GDP were the only available statistic, at some points in time, policy would have to be set on the basis of having no information on economic performance.

With this in mind, it is easy to see how a central bank could therefore conceivably be on the backfoot when reacting to shocks – such as a financial crisis.

Thankfully we don't live in a one statistic world, but the example shows how policymakers are therefore likely to lean more towards other economic indicators such as business surveys or industrial production releases. These tend to be updated at least on a monthly basis and, due to powerful characteristics such as timeliness and a high correlation with changes in GDP, may be used to project or "nowcast" economic growth ahead of its release.

The research provided within this thesis is primarily concerned with the role that high frequency indicators play in the nowcasting of important macroeconomic statistics such as GDP.

Several important contributions to the literature emerge from the research.

Firstly, the thesis offers corroboration and additional weight to a number of findings that have already emerged from the field.

These include confirmation that high-frequency data are useful in nowcasting GDP, especially in periods of recession and that taking some kind of average of the nowcasts provided by many small-scale models tends to yield better results than using individual specifications.

But what distinguishes the research from others is that these findings are consistent across countries and derived from nowcasting exercises based on real-time data vintages. This is an important consideration because economic data can be revised (and substantially so) through time.

Indeed, this author considers this to be especially important when using GDP statistics and finds when comparing nowcast model performance against consensus views these results and conclusions will vary dramatically if real-time data are not used. Because of the extensive use of vintage data, the research shows that consensus-based views continue to outperform automated statistical models, implying that judgement continues to play a positive role in nowcasting GDP.

However, the central finding that holds the thesis together is that timing is an important characteristic of economic indicators.

Due to the asynchronous release of data, indicators released quickly after the end of a reference period, such as the closely-watched Purchasing Managers' Indices, play an especially important role in reducing errors early in the nowcasting cycle when other sources of information are scarce.

Such is the influence of these earlier released indicators on the nowcast error that the impact of later released data tends to be substantially reduced.

This is a departure from the traditional way that indicators are generally perceived, where predictive ability is usually seen as of primary importance. The thesis argues that there needs to be consideration of the way the indicator is to be used in practice.

Moreover, in a world where data is now getting close to being available in real-time, the next logical step is to therefore assume that these new sources of information – such as from internet search providers – will be able to perform similar roles to business surveys in helping to reduce nowcasting errors ahead of other indicators.

This theory provides the platform for a major contribution later on in the thesis: the development of a MIDAS-based nowcasting framework to predict UK unemployment. Importantly the model incorporates, for the first-time for an economics application, Google Trends data at its true weekly sampling rate. Empirical evidence provided shows that internet search data can indeed offer timelier reductions in macroeconomic nowcasting errors than ever before.

1.1 Nuts and Bolts Issues

In 1956, Chancellor of the Exchequer, Harold McMillian quipped:

> "...some of our statistics are too late to be as useful as they ought to be. We are always, as it were, looking up a train in last year's Bradshaw."

Whilst the official statistics that McMillian refers to – such as Gross Domestic Product (GDP) – still tend to be released in an untimely fashion, there has been a rapid rise in data availability over the past couple of decades or so covering both demand and supply sides of the economy. Increasing volumes of information related to financial markets have also emerged.

All of this data could be extremely useful to policymakers. Waiting for GDP data may result in slow reactions to important economic developments, especially around business cycle turning points. Such risks could be mitigated by tracking and somehow utilising higher frequency information sources.

Subsequently institutions such as central banks are likely to receive and monitor hundreds of data series each and every month. In some respects the challenge of understanding economic activity today has moved away from a data availability problem to increasingly one characterised by finding a coherent signal from ever expanding datasets.

The goal of understanding "where are we now" can be partially satisfied by various econometric methods, a number of which have emerged from the academic literature in recent years. Before drilling down into a couple of these techniques, firstly let's outline several of the econometric "nuts and bolts" that are desirable to make such tools useful to nowcasters. The general aim of nowcasting is to form timely, accurate forecasts of macroeconomic variables that are available relatively infrequently and with a lag. Naturally the focus here will leans towards Gross Domestic Product (GDP), which remains the key barometer of economic activity but is only available on a quarterly basis, is published with a lag and is subject to revision.

As noted, between GDP releases there is a constant stream of data that could be useful in understanding current economic performance. These may include very high frequency series related to the financial markets, monthly surveys of businesses and consumers, or figures provided by official statistics agencies related to sectors such as industrial production.

The desire of the nowcaster is to link these various "explanatory" indicators to GDP, which can be referred to as the "dependent" variable, and then "nowcast" its current quarter growth rate.

However, two econometric challenges emerge.

Firstly, there is a time-frequency mis-match between the dependent and explanatory variables. GDP statistics are released on a quarterly basis, but many of the indicators that could be useful in explaining changes in GDP tend to be available to a much higher frequency (e.g. daily, monthly).

This represents the first constituent feature of nowcasting: any econometric methods designed to exploit potentially useful information in high velocity indicators

Figure 1.1: The Jagged Edge

		PMI Business Surveys	Industrial Production	Retail Sales	ILO Unemployment	Index of Services	GDP
Q1 2015	April	✓	✓	✓	✓	✓	✓
	May	✓	✓	✓	✓	✓	
	June	✓	✓	✓	✓	✓	
Q2 2015	April	✓	✓	✓	✓	✓	?
	May	✓	✓	✓			
	June	✓					

need to successfully deal with mixed time frequencies.

Secondly, indicators are released at different times and vary with respect to the time periods that latest observations refer to. This results in what Giannone, Reichlin, and Small (2008) refer to as the “jagged edge” of the dataset. Nowcasters will aim to exploit such features to make real-time nowcasts. Note an important difference here to traditional forecasting applications where balanced datasets are generally presumed.

Figure 1 provides an illustration. This shows a small set of variables that could all be useful in nowcasting UK GDP for the second quarter of 2015, the preliminary estimate for which was due for release on the 28th July 2015. The figure shows the data available to someone interested in nowcasting Q2 GDP on the 7th July.

The goal is to exploit all of the above information to provide an estimate of changes in GDP some three weeks ahead of its actual release.

In this illustration, a full set of business survey data is available, but there are some missing observations for other variables. Data for industrial production and retail sales are provided for April and May, but not for June. ILO unemployment and index of services figures are available just for April.

In short, any framework designed to nowcast economic activity must be flexible enough to deal with mixed time frequencies and at the same time also deal efficiently with the observed lags in publication and the non-synchronous nature of data release.

1.2 Nowcasting Methods

As the literature related to nowcasting has expanded, several excellent surveys have emerged providing extensive overviews and technical details on various techniques. Amongst others, these include Bańbura et al. (2013), Camacho, Perez-Quiros, and Poncela (2013) and Forini and Marcellino (2013).

What is apparent from these surveys is that methods used in the nowcasting space have generally been drawn from two distinct classes: conventional regression-style approaches and more complex state-space solutions.

Regression models tend to be single equation specifications. In contrast, state-space approaches involves the solving of a system of two equations (measurement and state).

While the single-equation regression approaches may therefore be viewed as less efficient, the use of state space techniques is inevitably a little more computationally involved. This is especially the case in nowcasting applications, where the problems of mixed-time frequencies usually involves the use of Kalman filtering to extract "missing data" for a low-frequency data series. Although this has the useful benefit of providing high frequency estimates of the respective state variable, there can be lot of parameters to estimate and uncertainty about whether the model specification is a replication of the underlying system. Such issues are accentuated if the information set contains a large number of variables.

Still, several authors have nonetheless highlighted that commonalities can exist.

For instance, Mitchell et al. (2005) show how the regression approach they use to create a monthly indicator of UK GDP can be transformed into a state space framework, while Bai, Ghysels, and Wright (2013) highlight how a MIDAS regression is an exact reduced form representation of the steady state Kalman filter. Crucially, they also find little difference in the results of an empirical application for forecasting US GDP growth between MIDAS and Kalman filter specifications.

These empirical findings helped to tip the balance towards using the regression-based approaches found in the forthcoming chapters, especially when placed in the context of the attributes of flexibility and relative computational ease associated with these more conventional macroeconomic models. Put simply, as research progressed over the past three or four years, regression approaches tended to meet needs best.

Therefore, for the sake of brevity, a detailed technical overview of state space methods is somewhat out of scope, although a short discussion makes an appearance in sub-section 1.2.4. For those interested in greater details, see aforementioned surveys – especially Bańbura et al. (2013) – for various references and a technical deepdive of these nowcasting solutions.

1.2.1 Bridging Equations

Bridging equation models are an approach to nowcasting that link the lower frequency variable via some functional form to a temporal aggregation of the higher frequency indicators. This aggregation places the indicators into the lower time frequency and tends to take the form of some average or sum of indicator observations that occur between those of the lower frequency variable.

Being relatively simple and having modest technological requirements, bridging equations have been used extensively to deal with the mixed-time frequency problem, particularly when mapping monthly indicator data to quarterly GDP observations. Examples, which also provide links to similar research, include Baffigi, Gonelli, and Parigi (2004), Diron (2008) and Ingenito and Trehan (1996).

To illustrate the bridging equation set-up assume that a quarterly statistic such as GDP, and signified as Y_t, can be predicted from a single high frequency indicator, X_t^{HF}, which is sampled at $k = \{1, ..., m\}$ observations per quarter. Also assume that the aggregation of this high frequency indicator is performed using an arithmetic average.

This yields a two-step approach to the nowcast of Y_t:

$$Y_t = \alpha + \sum_{j=1}^{p} \beta_j L^j X_t + \varepsilon_t \tag{1.1}$$

where

$$X_t = \frac{1}{m} \sum_{k=1}^{m} L_{HF}^k X_t^{HF} \tag{1.2}$$

Note the use of the back-shift operator to indicate the inclusion of p lags of quarterly readings of the indicator in the forecasting equation, which can be estimated by Ordinary Least Squares (OLS). L_{HF}^k denotes the lag operator for the high-frequency variable. So if X_t^{HF} represents, for example, industrial production growth then $L_{HF} X_t^{HF}$ represents industrial production growth for the last month of the previous quarter.

A key feature of the bridging equation model is an implicit assumption that all observations of the indicator are available at the point of aggregation.

This leads to a problem if there is a desire to exploit the real-time dataflow and the jagged edge. Waiting for slowly released missing data is not viable given the costs associated with not utilising high frequency information in e.g. policy formulation.

The workaround is to predict missing observations using separate forecasting models. These are typically done via uni-variate auto-regressive equations, although Rünstler and Sédillot (2003) explicitly consider a wide variety of options including uni-variate and multi-variate specifications. They find little discernible difference to nowcasts when filling in missing elements of the individual indicators over the required horizon.

Nowcasts produced in a two-step process where indicators themselves are projected forward via some separate auxiliary time-series model could be problematic: a high quantity of parameters may need to be estimated, especially if a large number of lags are involved, while success in the overall nowcasts may rest on the quality of the underlying models feeding into the main regression equation.

A further complication with the bridging equation approach is the potential for dilution or the loss of information from individual timing innovations through aggregation. Imagine if m is very large relative to observations of Y_t^Q, which may be the case if the sampling frequency is say daily. The result is considerable difficulty in uncovering the true relationship between the indicator and the target variable.

Despite such drawbacks bridging equations nonetheless remain a popular nowcasting tool at policy institutions such as central banks. See Bell et al. (2014), Bundesbank (2014) and ECB (2008) for examples.

1.2.2 MIDAS Models

Given the reliance on auxiliary models to project ahead the high frequency indicator, which are then aggregated and plugged into a separate equation to predict the lower frequency variable, the bridge equation methodology is by construct an iterative multi-step nowcasting procedure.

There is some debate in the literature over the pros and cons when employing iterative procedures in linear forecasting models. Iterated forecasting can prove to be more efficient than a direct set-up if the auxiliary models are correctly specified. However, in the case of mis-specification, a direct forecasting procedure could prove more advantageous (Marcellino et al. 2006).

MIDAS (Mixed Data Sampling) models are an example of a direct nowcasting solution and are useful tools in alleviating a number of the problems associated with bridging equations. Rather than having to rely on aggregation of high frequency variables, these are instead directly linked to the low frequency variable via a single equation:

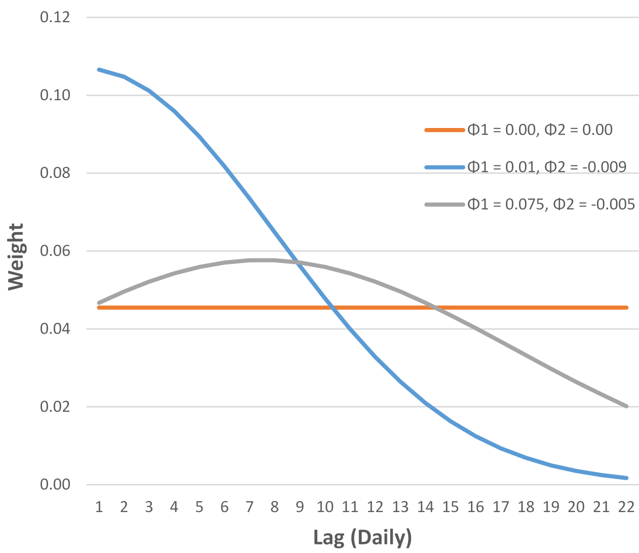
$$Y_t = \alpha + \gamma \sum_{j=1}^{p} \phi(k:\theta) L_{HF}^{k} X_{t-h}^{HF} + \varepsilon_t \tag{1.3}$$

where the regression co-efficient γ links the target low frequency indicator to a weighted sum of the indicator variable observations over the specified quarter. The function $\phi(k:\theta)$ is a polynomial that determines the weights used for the temporal aggregation of the high frequency indicator. For this reason, MIDAS models tend to be estimated by non-linear least squares (NLS).

The weighting function can take on various forms and there have been many specifications proposed. A common aim is to provide a parsimonious solution to the problem of excessive parameter proliferation when the number of high frequency indicator lags is large (see Ghysels et al. (2007) for a discussion).

One example is an exponential Almon function which uses two hyper-parameters θ_1, θ_2 to govern its shape. In this proposal, the number of lags j and sampling frequency k of the explanatory variable are explicitly taken into account:

Figure 1.2: Almon Weighting Function



$$\phi(k:\theta_1,\theta_2)=\frac{exp(\theta_1 k+\theta_2 k^2)}{\sum_{j=1}^{m} exp(\theta_1 j+\theta_2 j^2)} \tag{1.4}$$

An interesting feature of this proposal is that the practitioner has the option to determine the shape of the function by changing the values of the hyper-parameters, perhaps to match prior beliefs e.g. one may believe that the influence of the explanatory variable should monotonically decay with the number of lags included.

To re-affirm understanding, figure 1.2 provides a graphical example using daily lags of some indicator (assume 22 trading days in a month). By construct, the weights provided by the function sum to one and note when $\theta_1=\theta_2=0$ this is simple time averaging as per the bridging equation model.

Perhaps a drawback of this function is the sensitivity to different values of θ_1 and θ_2 leading to some challenge in finding the desired weighting function such as the monotonically decreasing or hump-shaped examples outlined in figure 1.2.

Another noteworthy feature of the MIDAS regression is there is no need for the extrapolation of missing observations to deal with the dataset's ragged edge: re-balancing is essentially achieved by shifting the time series of respective explanatory variables forward (or backwards) via the use of different integers for h, which reflects the difference between the forecast target period and the most recent observation of the indicator in equation 1.3. When nowcasting a quarterly statistic such as GDP, this provides the opportunity for within quarter estimations of the target variable by exploiting the timelier information provided by the high frequency indicators.

Note, however, that the model continuously has to be calculated as h varies and new data points are observed. This potentially makes practical implementation time consuming.

Since their introduction by Ghysels et al. (2004), there has been a steadily growing literature that showcases the empirical applicability of MIDAS to nowcasting. Examples can be found in Armesto, Engemann, and Owyang (2010) and Clements and Galvao (2008) for the United States. Kuzin, Marcellino, and Schumacher (2011) and Kuzin, Marcellino, and Schumacher (2013) augment the baseline MIDAS model by adding an auto-regressive term (referred to as an AR-MIDAS model) and nowcast GDP of the euro area and other industrialised countries.

1.2.3 Pooling of Nowcasts

To keep the discussion and the notation easy to follow, the equations outlined above have been expressed in terms of a single explanatory indicator.

However, as was noted in section 1.1, researchers now have access to a large and broad range of indicators that could all be useful in nowcasting.

The general approach taken by users of bridging and MIDAS models is to create a number of individual models, all of which are based on a single explanatory indicator. The determined number of models is therefore equal to the quantity of individual indicators i.e. N explanatory variables leads to N individual models.

The idea of this is largely to deal with issues of over parametrization when using many exogenous variables in one equation, while also avoiding any issues of co-linearity that can arise when using several macroeconomic time series in the same regression equation.

Moreover, there is evidence that the pooling of forecasts based on several small models can yield better predictive performance see e.g. Timmermann (2006).

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

$$\hat{Y}_t = \frac{1}{n} \sum_{i=1}^{n} \hat{Y}_{i,t} \tag{1.5}$$

Whereas equation 1.5 showcases an equal weighted average, there are many other options. Kuzin, Marcellino, and Schumacher (2013) for instance experiment with medians, weighting averages based on past nowcast performance in terms of mean-squared errors or use some kind of information criteria scheme such as the BIC to determine the weights. However, despite varying degrees of sophistication, in practice there seems to be little performance difference between the various pooling methods.

1.2.4 Factor Approaches

The bridging and MIDAS models are univariate approaches to the mixed-time frequency problem.

An alternative proposal in the literature is to use a multivariate method that casts a mixed-frequency VAR in a state-space framework. Aiming to summarize the co-movements and information held within the various series, the mixed-frequency VAR (MF-VAR) allows the joint dynamics of the indicators and the target variable to be explained.

The approace assumes that the model operates at the highest time frequency with all variables assumed generated but not necessarily observed e.g. the lower frequency data such as GDP are viewed as being a monthly data series with missing observations. The missing data and the nowcasts can be generated through the use of the Kalman filter and smoother see e.g. Kuzin, Marcellino, and Schumacher (2011) for an example.

The primary issue with a classical VAR approach is that parameter proliferation becomes a serious problem as the information set grows. Several solutions to this so-called curse of dimensionality problem could be adopted and have been proposed.

One option is to pool the nowcasts from many small systems (as outlined above with the bridging equations and MIDAS techniques).

Another option could be to employ Bayesian shrinkage techniques to avoid overfitting. In the context of nowcasting, Schorfheider and Song (2011) employ such an approach.

However, an especially popular method, and closely related to the MF-VAR models due to their state space representation, is to shrink the large information set through the use of factor techniques. The idea is to extract an unobserved state of the economy and create a coincident indicator that can then be exploited for nowcasting purposes.

Given a general observation that macroeconomic data tend to move closely together, there is a rich history in economics of using such an approach see e.g. Sargent and Sims (1977), Forni et al. (2002), Stock and Watson (2002) and, for an extensive survey, Stock and Watson (2011).

The heart of the technique is to extract from the dataset r unobservable factors which capture the bulk of the dynamics contained within the N explanatory indicators. Crucially $r << N$; the information held within a large volume of predictors is replaced by a much smaller number of estimated factors.

Using factor analysis undoubtedly offers the opportunity to analyse a large set of data and, due to the greater number of variables employed, lends some natural protection from structural breaks and individual indicators providing misleading signals.

Moreover, there is no need for any initial user judgement or opinion as can be the case when variables are pre-selected.

However, two challenges are immediately apparent with this type of modelling.

Firstly, how does the researcher go about extracting the factors and the number to retain? Secondly, how can the researcher deal with an unbalanced dataset?

As with the MF-VAR, the Kalman filter and smoother can be utilised. As the Kalman filter by construct can estimate and fill-in any missing observations via the exploitation of cross-sectional information in the dataset, the issues of the jagged edge are efficiently dealt with.

Doz, Giannone, and Reichlin (2007) and Giannone, Reichlin, and Small (2008) introduce such an approach, while recently Bańbura and Modugno (2014) suggest an EM algorithm could also be used to deal with missing observations. This idea is exploited by Bańbura et al. (2013) to nowcast US GDP, with the authors also incorporating high frequency financial variables into the framework to create a daily nowcasting factor model.

Because the extracted factors using the Kalman filter or EM algorithm are at the higher time frequency, this also offers the opportunity to use the extracted factors in a MIDAS style regression. Marcellino and Schumacher (2007) introduce such a strategy for GDP nowcasting in Germany.

Alternatively, the Kalman filter derived factors could also be aggregated into the lower time frequency and plugged into a bridge-equation as per equation 1.1. In essence this is the strategy adopted by Giannone, Reichlin, and Small (2008).

In the third chapter, the performance of bridging and MIDAS techniques – using both pooled model combinations and dataset factor shrinkage approaches – to nowcast preliminary estimates of UK GDP over the period of the depression is assessed.

While there are various differences and perceived pros and cons for each technique, ranking various approaches purely on theoretical grounds is tough, and this issue is further complicated by the commonalities they sometimes share (as highlighted earlier on page 8). The quality of a technique must surely rest on either its empirical performance i.e. its ability to successfully nowcast or perhaps how it aligns with the nowcaster's aims.

For instance, Giannone, Reichlin, and Small (2008) and Bańbura et al. (2013) write extensively about how the state space approach rooted in Kalman filtering is designed in part to allow the researcher to easily assess how the arrival of new information or data releases impacts on the nowcast through time.

Similarly, MIDAS regressions seem best suited to the application of nowcasting with very high frequency data provided by financial markets, thereby allowing the technique to showcase its value over bridging equations where aggregation to lower time frequencies may lead to the loss of information from the dilution of timing innovations.

At the time of writing, there was little in the way of comparisons between various methods in replicated real-time out-of-sample "horseraces". When conducting such a exercise, chapter 3 reveals there is in fact little to choose between methods in terms of end-of-period nowcasting accuracy. Recently, Marcellino and Schumacher (2007) come to a similar conclusion when comparing bridging and MIDAS models for the nowcasting of German GDP.

The analysis in chapter 3 also reveals that none of the aforementioned statistical models can consistently beat a consensus of professional economists in nowcasting preliminary GDP estimates.

The implication here is that, while nowcasting methods and an ability to understand current conditions has improved considerably in recent years, there needs to be some humility and recognition that pure statistical techniques can only go so far: judgement and experience continue to play a role in this arena.

For instance, the inability of statistical models to beat the consensus may reflect issues around the revisions and re-appraisal of GDP, which are especially inherent in UK national accounts statistics.

Such occurrences thereby impact on observed relationships between GDP and indicator variables and impair nowcasting performance.

With this in mind, when holding the target variable unchanged 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.

1.3 Hard and Soft Economic Indicators

Updates to nowcast estimates can be made in real time in line with the release of high frequency data and there is an agreement that utilising these indicators offers benefits for improving the accuracy of current quarter GDP nowcasts relative to some naive benchmark. For the application of GDP, numerous high frequency indicators are released throughout the nowcasting period before a first estimate is provided by official statistics agencies. These data include a vast array of sources from real economic activity and prices to indirect measures such as surveys, financial variables and money.

This leads to another common feature in the nowcasting literature: the classification of high frequency data sources into "hard" and "soft" indicators. The former refer to releases such as industrial production, which is a direct observation of an important input into GDP and is thereby likely to provide a strong short-term signal on wider economic performance. Soft indicators tend to be qualitative or indirect measures of economic conditions.

Traditionally, hard indicators were viewed as the primary source of improving nowcast accuracy with surveys containing little information beyond that already provided by the hard data see e.g. Baffigi, Gonelli, and Parigi (2004) and Rünstler and Sédillot (2003).

But as with GDP, these high quality information sources are also released with a lag, helping to create the jagged edge of the dataset that is observed when conducting real-time nowcasting exercises. Ideally an evaluation of the contribution to nowcasting accuracy of each data source needs to take into account their timeliness.

This provides an opportunity for business and consumer surveys (the so-called "soft" data) to gain in importance. Although potentially less precise – given the roles sentiment and expectations can play – these surveys are typically released around the end of the month to which they refer (equivalent industrial production data typically take around five-to-six weeks to be released). Financial variables covering exchange rates, short-term borrowing costs and stock market performance are available to even faster timescales i.e. on a daily basis.

And when explicit consideration of the lags in publication are taken into account, the earlier assertion that the soft data play little role in improving nowcast accuracy and that the hard data are the most important is reversed – timing matters.

In their study of the real time impact of macroeconomic releases in nowcasting US GDP and inflation, Giannone, Reichlin, and Small (2008) use 15 stylised "blocks" of broadly similar information or vintages to illustrate the marginal impact of different data sources on nowcast precision. Under the reasonable assumption that releases follow a broadly similar pattern through time, then those blocks that contain the earliest information on economic conditions tend to be the most effective.

Indeed, such is the reduction in uncertainty from these earlier releases that when the hard data releases enter the nowcast their impact tends to be marginal. This is an intuitive finding: given the collinear nature of macroeconomic variables the order of release should matter.

This is not to say the hard data don't contain any valuable information – on the contrary, when blocks are evaluated conditioned on timeliness, the predictive power of hard data is the strongest as per the findings using the traditional bridging models. It is only when the publication lags are accounted for that the soft indicators reveal their primary value – timing.

Providing similar evidence in this area are Angelini et al. (2008), Bańbura and Rünstler (2011), and Camacho and Perez-Quiros (2010) for the Eurozone and Matheson (2010) for New Zealand.

Note the role of very high frequency financial variables in improving nowcasting precision is somewhat less clear, however.

Aastveit and Trovik (2008) pay particular attention to the role asset prices play, suggesting that a block of information that contains timely financial data has a considerable influence in improving the nowcast of Norwegian GDP.

However, this finding is not generally supported by other studies, although in most cases financial variables are aggregated as monthly averages of daily observations i.e. unlike a MIDAS specification they are not explicitly modelled to be updated with new observations as they arrive and the assumption is availability occurs once a month (generally towards the end or the very start).

There is therefore a possibility that the importance of financial variables could be understated.

With this in mind, Alessi et al. (2014) report that simple MIDAS regressions using financial series could have improved GDP forecast accuracy both in the US and in the euro area during the financial crisis, thereby offering some evidence of the usefulness of such series around periods of economic stress.

This leads to a further observation: the majority of nowcasting methods developed and associated empirical studies tended to take place during relative economic stability i.e. during the so-called Great Moderation, a period characterised by reduced volatility in macroeconomic variables such as output and inflation. Would the various nowcasting methods perform differently during times of economic recession? Do the roles of various data sources change in importance?

Lombardi and Maier (2011) provide analysis of the backcasting, nowcasting and forecasting performance during the Great Recession of simple models based on the European Purchasing Managers' Indices (PMI) against more sophisticated dynamic factor specifications. While factor models that take a broad range of information tend to have the upper hand in most instances, the advantage of parsimonious survey-based set ups are sharpened during a period of macroeconomic volatility. This reflects their near instantaneous reaction to rapid changes in the economic environment. In contrast, sophisticated time-series based models can suffer from persistence. For these reasons, there could be justification in placing more weight on the surveys at times of economic stress or around major events e.g. 9/11 or the collapse of Lehman Brothers in September 2008.

Mitchell (2009) offers a similar conclusion when assessing various methods for nowcasting UK GDP during the onset of the Great Recession. A simple model that makes considerable use of qualitative survey data is found to provide the earliest indication of recession, but is a relatively poor performer during periods of relative economic stability compared to the more sophisticated specifications.

In chapter 2, a closer look is taken at the importance of surveys in helping policymakers gain a timely insight into real-time economy performance.

Using real-time data sourced from the Deutsche Bundesbank, the OECD and the Office for National Statistics, this chapter offers a thorough assessment of the role that the popular and widely used Purchasing Managers' Index (PMI) play in reducing nowcast errors in a simple bridging equation framework.

The empirical exercise is conducted for five developed economies and also covers the period of the Great Recession.

1.4 Big Data

Soft indicators, such as survey data, play a crucial role in refining and understanding trends in important macro-economic variables. Their advantages over richer official statistics don't lie in being able to provide an indication of magnitudes of change, rather in their timeliness and offering reliable signs of directional movement, especially around periods of economic stress. For instance, business surveys flashed brightly on policymakers' monitoring dashboards during the depths of the 2008/2009 global economic downturn, helping central banks to formulate swift responses to the considerable impact of the crisis on real economic activity.

Partly in response to the events of 2008 there has been a growing interest within policymaking circles not only in nowcasting methods, but also around data that are a) available to extremely quick timescales b) of very high-frequency and c) granular in nature and much so than traditional economic statistics (Bholat 2015).

Information that exhibits some or all of these characteristics are being classified as "big data", with many sources borne out of the increasing interaction between technology and the population. Examples include internet searching activity, debit card transactions or activity related to social media. In this section, several concepts, data issues and current research in the area of big data are introduced.

1.4.1 Statistical Traps

Some suggest that big data sources are so rich in potential they may be a game-changer for economics and how economists work (Varian 2014). The epistemological and methodological boundaries of the discipline could be stretched and pulled away from a traditional philosophical leaning of deduction (using data to rationalise theory) towards a more inductive stance (theory is instead generated by observed patterns within data).

Alternatively, others may exhibit a modicum of scepticism.

Take sampling bias, for instance. Given many of these new information sources are rooted in an engagement between people and technology, take-up and usage is likely to be different across generations so the datasets are not equivalent to $n =$ everything (as some big data proponents may suggest). These biases, which could be exaggerated by the size of the dataset, ideally need to be sought out and corrected.

Other issues are revealed in Google's very own Google Flu Trends (GFT) prediction system. This was designed to forecast changes in flu based on what people are searching for in Google. At launch, the probability of successfully forecasting changes in official health statistics seemed high due to a strong correlation with several seemingly carefully selected search terms and official US Centre for Disease Control (CDC) figures (Ginsberg et al. 2009).

But as a forecasting tool GFT has shown some unfortunate weaknesses – it stopped being very good at predicting. Forecasts were inaccurate for 100 out of 108 weeks according to Lazer et al. (2014) and was consistently beaten by a simple model set-up based just on the previous week's CDC data.

Google have been fairly secretive in which search terms they use to predict flu, so reasons why the forecasting errors are being made are, to a degree, speculative. But one persuasive reason is that, with millions of search terms being used to find patterns with a relatively short target time-series, the chances of finding spurious correlations were high if the cornerstone of the approach is in algorithms designed to find high correlations between search terms and a dependent variable. As correlation does not necessarily equate to causality, Google's early attempts at nowcasting with big data seemed to be snared by a classical statistical trap.

1.4.2 Behavioural Motivations

A somewhat less obvious issue with GFT and, for that matter, internet search statistics in predictive practices, concerns the motivation of the user to actually enter a search-term.

Think of why someone may punch into Google the search term "flu". Are they searching on the internet because they have flu symptoms and are looking for a cure? Or are they searching for flu because they have heard about some major outbreak on the other side of the world?

Lazer et al. (2014) suggest the researcher therefore needs to carefully consider social and independent searching. Is the user searching for their own purpose, or is the search more akin to some kind of herd behaviour i.e. because many others are doing the same? Such issues have considerable implications for forecasting ability and Bentley, Nyman, and Ormerod (2014) point out this could be a factor behind the persistent errors by Google in predicting the number of flu cases in recent years.

All of this leads to questions over whether a pure "let the data speak" inductive stance is feasible; there will inevitably be some work to get the dataset "fit-for-purpose". Moreover, model set-up could be dependent on the preference of the model builder themselves. Should the model be built primarily from the ground-up i.e. data-driven or is a top-down approach in the traditional, classical sense, more appropriate?

1.4.3 Big Data Practicalities

A further feature of big data research is a dependence on practicalities such as access to proprietary information held by corporates or substantial computer processing power as data feeds can be large in nature plus of very high velocity.

Institutions such as the Bank of England – who consider themselves to be an important player in the big data space, naturally so given their interest in understanding economic performance in a near as possible real-time setting – are able to leverage considerable resource to set-up data centres and recruit computer scientists to help their economists make sense of big data feeds (Bholat 2015).

Yet for independent researchers, with limited resource and restricted exposure to corporates, perhaps reliability and replication of big data research could prove to be a little too far outside of usual comfort zones (Taylor, Schroeder, and Meyer 2014).

A pertinent example is electronic payments data.

Private sector consumption data produced by national statistics agencies tends to be released slowly (quarterly and well after the reference period) and can be subject to considerable revision.

Offering a timely and cost-effective alternative, payments system data could be useful in addressing such issues with official statistics. For instance debit card transactions from major service providers such as the likes of Visa and MasterCard have the potential to offer near real-time, high-frequency (e.g. daily) broad-based information on household consumption. These companies are beginning to recognise the potential, producing monthly reports in countries such as Ireland, the UK and the US. Moreover, there are a number of research papers highlighting the positive marginal contribution such information can have in reducing GDP or private consumption nowcasting errors see e.g. Galbraith and Tkacz (2007) and Esteves (2009).

Despite its obvious potential, research with payments data has however tended to be concentrated amongst a small number of researchers, sometimes with an affiliation to central banks.

Some work has also emanated directly from these institutions e.g. Banco De Portugal (Lima 2013) and Reserve Bank of Australia (Gill, Perera, and Sunner 2012). Central banks sometimes have regulatory oversight over providers of card services, providing the opportunity to directly observe payments information. Other interested parties are unlikely to be as privileged and there is a suspicion that challenging data access has curtailed the opportunity to gauge the usefulness of payment statistics in the nowcasting space.

1.4.4 Internet Search Data

By contrast, big data opportunities for the research community are more becoming in the world of internet search data. This is largely thanks to search engine behemoth Google making internet search information freely available via their Google Trends platform.

The result has been an array of papers as researchers take the opportunity to explore the cost-effective and high-frequency nature of the data. With hundreds of millions of queries inputted into engines such as Google on a daily basis, by assessing trends in query volumes – which is provided free of charge by Google Trends on a daily and weekly basis – shifts in behavioural patterns could in theory be observed ahead of other data sources.

While there are already various examples of this data being used across a number of subjects, such as providing an advance warning of flu epidemics (Ginsberg et al. 2009), for the economics profession search data appears naturally suited as a micro-economic tool, offering rapid market intelligence to businesses, with variances in search popularity giving an understanding into how their own and competitors' goods and services are received. The data can also provide an insight into changing market behaviour and where to target advertising.

However, Google Trends also offers the macroeconomist opportunities to understand trends in important areas such as consumption, unemployment and housing markets. Being available in near real-time and ahead of the traditionally earliest indicators, search data can be used in short-term forecasting or nowcasting applications.

Since an introduction by Ettredge, Gerdes, and Karuga (2005) of the possibility of using web-search data as a predictor of macroeconomic statistics, particularly unemployment figures, results from subsequent empirical papers have been encouraging.

In two papers, Google Inc economists Hyunyoung Choi and Hal Varian (Choi and Varian 2009a; Choi and Varian 2009b) use search data to lower nowcasting errors for US retail and auto sales, new housing starts, travel destinations and initial claims for unemployment benefits, while Askitas and Zimmermann (2009) indicate that keyword searches correlate strongly with monthly German unemployment data and that a Google predictor can add value to an error prediction model. D'Amuri (2009) assesses the power of augmenting standard time series models for quarterly Italian unemployment, concluding that the data improve out-of-sample forecasting performance.

Meanwhile, McLaren and Shanbhogue (2011) perform a similar exercise for the UK labour and housing markets, comparing a simple baseline AR specification to one augmented with an internet search variable.

Building on the literature, the authors find the Google Trends data contains information above and beyond those provided by survey indicators of the UK housing and labour markets.

Chapter 4 is all about big data, and takes on the challenge of extending and offering some refinements to the emerging literature on internet search as a tool for tracking economic activity.

Having outlined a number of pros and cons of using such information, the chapter explores the thorny problem of search-term selection: faced with potentially millions of terms within a vast dataset, how does the researcher find the most relevant-term for their respective nowcasting exercises?

Two empirical exercises are then conducted in the context of predicting changes in UK unemployment.

Firstly, having used some theoretically derived keyword to kick-start a structured approach to the selection of search terms, a composite Google indicator is created. This gauge of labour market performance is then tested alongside survey-based variables in a traditional forecasting environment.

Secondly, this environment is expanded into a pseudo-time nowcasting framework which provides the backdrop for assessing the timing advantage that Google data have over surveys.

The framework is underpinned by a MIDAS regression which allows, for the first time, the easy incorporation of internet search data at its true sampling rate into a nowcast model for predicting unemployment.

1.5 Model Evaluation

Finally, a general comment on the metric or loss functions used to evaluate the performance of the models.

While admittedly the choice of error measurement in each chapter was more a growing reflection of the skills and knowledge that was built up during the research journey, rather than any explicit preference or consideration of the various pros and cons of each approach, there is nonetheless a growing production, use and sophistication of these loss functions throughout the thesis.

Chapter 2 begins with the use of the Mean Absolute Forecasting Error (MAFE) statistic. This measures the average magnitude of the nowcasting errors over a given sample but, by using the absolute value, takes into no account the direction of these errors.

The MAFE is a linear loss function – the approach in essence ensures that all the individual nowcast errors are scaled equally. For example, being 10 units away from the mean is viewed exactly twice as bad as being 5 units away.

However, it may make sense to the user to attach a greater weight to errors that are further away from the mean i.e. being 10 units away could be perceived to be more than twice as bad as 5 units away.

In this instance, a quadratic scoring rule such as the Root Mean Square Forecasting Error (RMSFE) could be a more appropriate loss function. Chapter 3 uses such a statistic.

The RMSFE could be described as the squared sample difference between the nowcast and observed outturns. Having calculated the sample mean, the square root of this number is then used.

Because the nowcast errors are squared before being averaged this means that the RMSFE statistic places a higher weight on larger errors. This of course may be important to the user. For instance, it may make sense that errors made during periods of economic shock garner greater weight as the cost of such "misses" are likely to be much greater than smaller errors recorded during periods of economic calm. Both the MAFE and the RMSFE can vary from 0 to ∞ and they are negatively oriented scores: lower values are seen as better.

However, a criticism of the MAE and RMSFE loss functions is their focus on the performance of models in providing point estimates i.e. how accurate is the model in providing the actual outturn?

While such a focus remains a dominating force in the professional forecasting field e.g. the consensus views of forecasters around important economic releases (and the high weight attached by financial markets to any differences), in recent years there has been a growing recognition of the uncertainty that comes with producing such estimates.

As Mazzi, Mitchell, and Montana (2010) discuss, rather than focussing on whether the nowcast is “right” or “wrong” in a somewhat binary fashion, it may be better to view the point estimate as being a central position in a range of uncertainty. Is it really surprising if GDP growth comes in a little bit lower or stronger than a nowcast of say 0.6%? Or in a period of heightened uncertainty, that the outturn is much higher or lower than the point estimate?

With this in mind, nowcasters may therefore wish to provide a density nowcast (or as popularized by the Bank of England, “fan charts”) and offer an estimate of the probability distribution (or uncertainty) around the nowcast.

Rather than purely focus on a loss function such as the RMSFE, the growing recognition of the uncertainty that exists with the provision of nowcasts in part motivated the use of a density function when assessing the performance of internet search data to help predict changes in unemployment for Chapter 4.

The underlying assumption was the function around the point estimate is normally distributed. There is of course debate on whether using such a function is the most appropriate approach, but such considerations were viewed to be beyond the scope of the thesis and left for future work.[1]

[1]For instance, the uncertainty could be modelled according to observed past forecast errors given these may not meet hypotheses of normality and unbiasedness see e.g. Tay and Wallis (2000). Or perhaps the forecaster wishes to get across some kind of message on whether they view risks as being to the upside or downside to some central estimate. As an example, the BoE famously uses a “two-piece” normal distribution when showcasing its projections in the Inflation Report. This choice allows for an element of asymmetry around a central modal projection, the idea being to represent the possible paths for inflation based on economic analysis and the collective subjective judgement of the Monetary Policy Committee. In many instances, these assessments of alternative outcomes to the central case are more likely to be skewed one way or the other see e.g. Britton, Fisher, and Whitley (1998).

Chapter 2

The Importance of Being Timely

This first research chapter builds on the ideas on nowcasting that were developed and outlined in the first chapter of the thesis by conducting a practical empirical exercise to showcase the value of economic indicators from the perspective of timing.

As previously noted, policymakers want to know about real-time economy performance. However, closely watched macroeconomic time series produced by national statistics offices are published infrequently, with a time lag and subject to revision. Such issues create uncertainty in tracking economic developments, a by-product of which is to raise the value of business and consumer surveys. Although providing less granularity than official data series, the surveys are released in a timelier manner and are subject to little revision.

The research extends the existing nowcasting literature by using real-time data sourced from the Deutsche Bundesbank, the OECD and the Office for National Statistics to undertake an assessment of the role that the popular and widely used Purchasing Managers' Index (PMI) play in reducing forecasting errors in a simple "nowcasting" framework. The use of real-time data is a departure from the general "pseudo-time" applications that have tended to dominate the field.

Moreover, rather than focusing on predictive ability, as has historically tended to be the case amongst researchers, it is the first time that PMIs have been assessed with the explicit aim of understanding the usefulness to policymakers of the timing properties of these surveys.

The attention on the PMIs is justified not only because of the wide references to the surveys seen within monetary policy documentation, which presumably means policymakers attach some importance to them, but also because of the opportunity to conduct the empirical exercise across countries. Again this is a rare occurrence in the nowcasting field, where empirical applications tend to be presented primarily for single countries/regions.

2.1 Introduction

Survey-based indicators of economic activity are timely and rarely revised. These are attractive characteristics, and enhance the importance of survey data to policymakers who want to gauge how the economy is performing, but are faced with a vast array of information covering key macroeconomic variables. Some of these are released infrequently, with a lag and subject to considerable changes post-release. Adding such layers of uncertainty makes optimal policy decisions just that little bit more difficult.

This chapter explores for Japan and a selection of major European countries (France, Germany, Italy, and the UK), the timing characteristic of high frequency indicators via an empirical application that focuses on very short-term forecasts or "nowcasts" of gross domestic product (GDP). The modelling exercise is conducted in line with a standard release calendar and in a way that the marginal impact on nowcasting errors of all observations for all indicators is assessed in a full and complete manner. Inspiration is in part taken from the work of Bańbura and Rünstler (2011), Bańbura, Giannone, and Reichlin (2011), and Bańbura and Modugno (2014), who explicitly model the flow of real-time information to highlight the importance of release order on an indicator's marginal predictive power for the Eurozone and the United States.[1]

Vintage data are also utilised i.e. the actual published data series available to a practitioner at each point in time during a given nowcasting period is used throughout. Improved availability and accessibility to datasets containing sequential vintages also makes real-time data applications a little easier than the past.

So, as well as conducting a cross country study, this research adds an additional layer to the literature, being a departure from the "pseudo real-time" modelling exercises found in the majority of nowcasting applications. It is a genuine replication of the data available to analysts at each point throughout the sample period (2006-2012).

The adoption of a true real-time data analysis is rooted in the observation that economic time series have a tendency to be revised, particularly "hard" data series such as GDP, industrial production and retail sales. Importantly, revisions are known to have an impact in a variety of contexts such as structural modelling, historical monetary policy analysis and, of course, nowcasting model evaluation. Croushore (2011) provides an extensive survey of this type of research, and it is clear that the revisions process could be an important aspect to consider when assessing the role of indicators in a nowcasting framework.

For example, using the Federal Reserve Bank of Philadelphia real-time database, analysis by Croushore and Stark (2001) indicates conventional forecast-error stats

[1]The term "nowcasting" is described by Bańbura, Giannone, and Reichlin (2011) as a set of forecasting methods which have a primary purpose to "predict the present, the very near future and the very near past". These methods are designed to form a coherent picture of the present but at the same time overcoming the considerable challenges that are associated with data sources that are released in a non-synchronous nature and in mixed time frequencies.

may be sensitive to the choice between latest available and real-time data. The degree of sensitivity is such that conclusions of model forecasting superiority due to lower forecasting errors than others derived in real time (such as consensus views) cannot hold.

The choice of data vintage may also affect the model-selection period so the safest method for evaluation is with real-time data. Koenig, Dolmas, and Piger (2001) add weight by finding out-of-sample forecasting performance for US GDP is substantially improved if, on both right- and left-hand sides of equations, estimations are with real-time rather than end-of-sample vintage data.

However, in contrast, Diron (2008) finds only a limited impact when analysing the role of revisions in bridging models for the Eurozone.

Particular use is made in the empirical application of the popular and widely used Purchasing Managers' Index (PMI) survey data. These indices are often referenced in the monthly minutes of Bank of England monetary policy meetings or the European Central Bank monthly bulletins, implying PMI data contain useful information to aid the monetary policy process. With this in mind, several studies that extract common factors from datasets containing many macroeconomic data series find a first factor tends to correlate highly with PMI data. This suggests that PMIs do a good job in summarising several different sources of information and offer a timely update about the underlying performance of economies see e.g. Andreou, Ghysels, and Kourtellos (2013) and Lombardi and Maier (2011). However, generally speaking, studies of the role of PMI data have focused on predictive power rather than explicitly accounting for timeliness within the flow of information that characterises policymaking.

PMIs are diffusion indices, based on the individual qualitative responses of companies that contribute to a monthly survey panel, which has been selected to reflect the underlying structure of a sector/economy according to official Gross

Valued Added (GVA) and company size information. The indices are calculated by taking the proportion of companies reporting 'up' and adding a half of those that report 'no-change' (hence if all companies reported 'no-change' the index would be equal to 50.0 with, by extension, the index bounded between 0 and 100).

Throughout the empirical applications below, the PMI data are used as 'given'. This contrasts with studies that have made use of the underlying panel responses such as Mitchell, Smith, and Weale (2013), who recently suggested that, if the aim is to forecast or nowcast economic activity, then it may be more efficient to disaggregate the panel by linking individual firms to official growth rates via their qualitative business survey reporting records. On the contrary, with the academic literature on panel quantification of qualitative panel surveys in mind, the PMI data used in this application is more closely related to the traditional "aggregate" approaches (see Pesaran and Weale (2006) for a survey). And while it is clear that accessing and utilising individual responses to the PMI panels (especially given the global reach of the PMI company database) would be an attractive extension of the research, at this juncture such an approach was considered to be beyond the scope of this chapter, not least of all due to the aim of showcasing the value of PMIs using data that are widely available to the public.

As a prelude, the emphasis placed on PMI and survey data, particularly in European policy-making circles, is well placed. In an out-of-sample recursive nowcasting exercise, the surveys have a clear positive impact in reducing forecasting errors associated with estimating current quarter changes in GDP. Value is directly derived from the timing of release relative to other popular macroeconomic data indicators.

The chapter proceeds as follows. Section 2.2 outlines the nowcasting econometric framework used to explore the timing aspect of a small set of macroeconomic indicators, while section 2.3 describes the dataset. Section 2.4 provides findings from the empirical exercise before section 2.5 compares and contrasts the findings with recent nowcasting applications. Section 2.6 summarises.

2.2 Methodology

The primary aim of the chapter is to explore the timing characteristics of PMI data against a small number of popular macroeconomic time series (both soft and hard) in nowcasting quarter-on-quarter changes in GDP. While there is a rapidly growing literature and debate of the best way to deal with the mixed time frequency problem, in this instance a parsimonious and easy to implement two-step bridging equation framework suffices (an extended discussion on the methodology is provided in section 2.5). The bridging equation 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), Grasmann and Keereman (2001), Ingenito and Trehan (1996), Mourougane and Roma (2002), Parigi and Schlitzer (1995), Mourougane and Roma (2002), Parigi and Schlitzer (1995), and Rünstler et al. (2008).

2.2.1 The Nowcast Regression

The first step in performing the nowcast is to convert some high frequency variable X_t^{HF} (e.g. a PMI) into the same time domain as the low-frequency variable Y_t (i.e. quarterly GDP). The simplest transformation would be to calculate a simple average of the m readings of X_t^{HF} observed between readings of Y_t (e.g. between Y_{t-1} and Y_t). The transformation equation is therefore:

$$X_t = \frac{1}{m} \sum_{k=1}^{m} L_{HF}^k X_t^{HF} \tag{2.1}$$

L_{HF}^k denotes the lag operator for the high-frequency variable. So if X_t^{HF} represents, for example, industrial production growth then $L_{HF}X_t^{HF}$ is industrial production growth for the last month of the previous quarter.

This first step is only performed for those variables that have not been transformed into three-month on three-month rate of changes. In the case of this research, equation one would principally be used for the PMI and similar survey based data. For others, readings for the final month of a calendar quarter (e.g. observations for Mar, Jun, Sep, Dec) are equivalent to quarterly GDP growth and no averaging of the three monthly readings for each quarter is required (more on this in section 2.3).

The second step is then to link the explanatory and dependent variables through some kind of defined regression approach, such as an autoregressive distributed lag (ADL) specification:

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i L^i Y_t + \sum_{j=1}^{q} \beta_j L^j X_t + \varepsilon_t \tag{2.2}$$

Note the use of the lower frequency lag operator on X_t; that is q lags of the time averaged X_t^{HF}'s are used wherever applicable, with the number of lags determined by the Akaike Information Criterion (AIC) up to a maximum of four. P lags of Y_t are also utilised, also determined by the AIC to a maximum of four. Given the contemporaneous nature of the indicators involved, zero or at most one lag is found to be the best specification.

Nonetheless, a flexible approach is taken with regard the structure of the regression equations, which are first estimated at the start of 2006 based on the sample to that date but are then reviewed and adapted accordingly at the start of each calendar year in case structures are found to vary as the sample window expands. In the empirical work, a full ADL specification works best in some instances (for example, a single lag of GDP is found to be useful in most instances when nowcasting current quarter UK or French GDP), whereas a distributed lag regression may work in other cases, or even a regression that contains zero lags of both the dependent and explanatory variables.

The bridging equation utilises one explanatory variable to explain changes in GDP. 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) to leverage the idea seen throughout much of the literature on bridging equations, and more recently with alternative approaches such as MIDAS regressions, that the pooling of forecasts based on several small forecasting (or nowcast) functions can yield better forecasting performance. See Aiolfi, Capistrán, and Timmermann (2010) for a discussion.

Therefore, with N explanatory variables, an equivalent sized set of nowcasts will be available at any point in time. Let $\hat{Y}_{i,t}$ represent a current quarter nowcast of GDP, with the overall nowcast, $\hat{Y}_t$, the time varying weighted average of these:

$$\hat{Y}_t = \sum_{i=1}^{N} \omega_{i,t} \hat{Y}_{i,t} \tag{2.3}$$

where $\omega_{i,t}$ represents a set of normalised weights that sum to one at time t i.e. $\sum_{i=1}^{N} \omega_{i,t} = 1$. Individual weights are based on the correlation of each explanatory indicator's relationship with quarterly changes in GDP (as determined by the R^2 statistic). Those with a higher R^2 are subsequently given a greater weight in producing the overall nowcast, thereby allowing a greater influence on the nowcast of stronger indicators of GDP. However, rather than fixing these weights across the whole sample, they are re-applied (based on data available at the time) at the start of each calendar quarter. This helps to account for any changes in the realtionships between GDP and respective indicators.

Table 2.1: Typical Monthly Order of Release and Vintage (January 2013)

Order	1	2	3	4	5	6	7	8
France	MPMI/IPMI	CPMI	SPMI	TRA	IP	ESI	RS	
	(Dec-12)	(Dec-12)	(Dec-12)	(Nov-12)	(Nov-12)	(Jan-13)	(Dec-12)	
Germany	MPMI/IPMI	SPMI	CPMI	FacOr	IP	TRA	ESI	RS
	(Dec-12)	(Dec-12)	(Dec-12)	(Nov-12)	(Nov-12)	(Nov-12)	(Jan-13)	(Dec-12)
Italy	MPMI/IPMI	CPMI	SPMI	IP	TRA	RS	ESI	
	(Dec-12)	(Dec-12)	(Dec-12)	(Nov-12)	(Nov-12)	(Dec-12)	(Jan-13)	
Japan	EWS	CC	TRA	RS	IP	MPMI/IPMI		
	(Dec-12)	(Dec-12)	(Dec-12)	(Dec-12)	(Dec-12)	(Jan-13)		
UK	MPMI/IPMI	CPMI	SPMI	IP	TRA	RS	IoS	CC
	(Dec-12)	(Dec-12)	(Dec-12)	(Nov-12)	(Nov-12)	(Dec-12)	(Oct-12)	(Jan-13)

CC = Consumer Confidence; EWS = Economy Watchers' Survey; ESI = Economic Sentiment Index; RS = Retail Sales; MPMI = Manufacturing PMI; TRA = Trade Balance; IPMI = Investment PMI; IP = Industrial Production; CPMI = Construction PMI; IoS = Index of Services; SPMI = Services PMI; FacOr = Factory Orders
Schedule based on average release day in 2012, not the release dates that occurred in January 2013.
Source: Financial Times Online Economic Calendar

2.2.2 The Jagged Edge

A feature of nowcasting is being able to use high frequency data to produce intra-period nowcasts. A current quarter GDP nowcast, $\hat{Y}_t$, should be improved and refined via the utilisation of the information that is observed in the high frequency data between Y_{t-1} and Y_t. In this research, a primary objective is to judge how the current quarter nowcast evolves with the arrival of new information. Regressions are therefore ran on the first working day of the calendar quarter (e.g. Jan 1st, Apr 1st, Jul 1st, Oct 1st) and then updated as fresh observations for each of the explanatory variables are released, with the order being in typical chronological order (this order is determined from the online Financial Times Economic Calendar. See table 2.1 for the indicators used and a general order of release).

However, the standard bridging equation framework implies that all readings for a full calendar quarter are available when conducting a nowcast. This allows the monthly series to be transformed into quarterly time-series as per equation 2.1. But due to lags in data publication and the non-synchronous nature of releases, data for the monthly indicators are available at different times. The result is what Giannone, Reichlin, and Small (2008) refer to as the dataset's "jagged" edge and, to correct this problem and conduct nowcasts on a continuous basis as a quarter evolves, the econometrician needs to "fill in the gaps".

The simplest solution would be to use some kind of auxiliary modelling to forecast the missing observations of respective time-series. There are several options available. Rünstler et al. (2008) use an autoregressive (AR) approach, while Klein and Sojo (1989) deploy ARMA modelling. After experimentation, Ingenito and Trehan (1996) settle on BVARs. Rünstler and Sédillot (2003) explicitly consider a wide variety of options including univariate and multivariate specifications. Little discernible difference for current quarter forecasts is found, although multivariate models hold the edge over a longer time-frame.

Given the lack of consensus, univariate AR models with a maximum lag of 12 are used in this research to forecast missing monthly readings of the individual indicators over the required forecasting horizon (up to 5 or 6 months ahead dependent on the position in any given quarter the nowcast is conducted). Lag lengths are again determined by the AIC. The results of the auxiliary forecasts are then combined with observations already available. The high frequency variables are subsequently transformed into a quarterly time series to match the time frequency of GDP observations.

2.2.3 Assessment

If the explanatory variables contain useful information, the accuracy of the nowcast should improve as imputed forecasts are replaced by actual readings. The order in which the data are released could also be important.

To assess whether these hypothesizes hold, individual country models are run in a recursive fashion beginning January 2006 (for Q1 2006), with weighted nowcasts produced as each new point is released until just prior to the first estimates of Q2 2012 GDP were published. The exercise is conducted recursively for a total sample of 26 quarters (to provide an idea of the total number of times the exercise is performed, this equates to over 600 nowcasts for some countries).

The actual nowcast is compared with first estimates of quarter-on-quarter changes in GDP, any difference being the forecasting (or nowcasting) error. Differences between quarterly changes in GDP and a simple benchmark model are also recorded. This benchmark is an AR(1) of quarter-on-quarter changes in GDP.

Being based on equivalent datasets and information available to forecasters at a particular point in time, nowcasts are a replication of a real-time data setting i.e. the modelling process produces the same results as would have been the case if an analyst was nowcasting with this approach for e.g. Q2 2008 GDP on 8th May 2008 or e.g. Q3 2011 GDP on 2nd October 2011. As analysts and policymakers primarily forecast first published estimates of GDP, the results are also compared against a real-time GDP series.

2.3 Data

Vintage time-series data are sourced for the dependent variable, GDP (Y_t), and a small set of explanatory variables $X = i, ..., N$ for each country from a variety of sources.

Technological advances in recent years mean macro-economists now have at their fingertips a vast array of high frequency information, covering both demand and supply sides of economies, so there is seemingly a large pool of potential indicators to choose from.

However, when consideration is made of data vintage availability (which is crucial for conducting a genuine “real-time” nowcasting exercise), few real-time databases exist. The Deutsche Bundesbank provides vintage data for Germany and information is obtainable from the Office for National Statistics (ONS) and the Bank of England for the UK. In addition, the OECD offers historical vintages for 21 economic variables that appear on a monthly basis in their Monthly Economic Indicators publication.

Now thinking about the desirable characteristics of the explanatory variables, these could include timely release, few historical revisions and, ideally, a good correlation with GDP and/or components. Moreover, several indicators are persistently reported by the financial press and closely watched by analysts. Examples here would be industrial production, retail sales and trade statistics. As these tend to be used by statistical agencies in providing early estimates of GDP they are, as such, viewed as high quality indicators of economic activity.

Vintage data for all of these indicators is readily available from the aforementioned OECD database. Where vintage data is available directly for individual nations e.g. Germany and the UK, there is scope for using additional “hard” indicators, which are defined here as measures of economic activity released by official statistics agencies. Although high quality, and sometimes used as direct inputs into GDP calculations, they tend to be subject to revision and published with a lag (i.e. some time after the period they refer to).

“Soft” indicators, in contrast, refer to timely, generally unrevised, measures of activity such as business and consumer surveys. As noted, extensive use is made of the Purchasing Managers’ Indices (PMI) surveys produced by Markit Economics. These closely-watched surveys capture developments in the manufacturing, services and, in some instances, construction sectors by asking purchasing managers or company executives a variety of questions regarding output, orders, inventories, employment and prices. The nature of the surveys invokes a qualitative response from respondents as companies provide an “up”, “down”, or “same” response to statements such as "Please compare your production/output this month with the situation one month ago". The PMI surveys are therefore not to be viewed as a sentiment index per se: the wording of the question leads to a more factual response when compared to surveys which focus on expected trends, or assessments of current conditions based on the opinion of respondents e.g is your assessment of current business conditions "satisfactory", "good", "poor", or "unfavorable"?

There are other advantages of using the PMIs. These include rare revisions to historical data series and a consistent methodology across countries (which, in a cross country study, where sources and standards of data can vary, is rare). Released at the start of a calendar month (with data referencing the previous month) they are published well ahead of equivalent data. Moreover, the response rates for the monthly surveys are typically high (Markit Economics reports a monthly response rate of around 75%-80% across its panels), while there is importance attached to panel stability. This could be an important aspect in delivering a consistent time series and enhancing nowcasting ability.

In the empirical application, the headline indices from the manufacturing, investment goods, construction and service PMIs are used. For manufacturing and investment goods, the PMI takes a weighted average of the indices for five component series (output, new orders, employment, suppliers' delivery times and stocks of purchases), but for services and construction, these are based on a single question covering the one-month change in business activity.[2]

[2]Markit Economics produce "flash" PMI results for the Eurozone, France and Germany i.e. advance indications of "final" readings. These early estimates are based on around 85% of sample sizes and are offered around a week before "final" releases at the start of a calendar month. In this study, "final" readings are used, which are then rarely revised. A notable exception is the UK Manufacturing PMI. Although early estimates are not provided, the historical time series can be revised on a monthly basis. This is due to seasonally adjusted data being re-calculated each month due to a reliance on the X-12 seasonal adjustment program. Vintage datasets are not available for this series. A similar approach is taken for the Economic Sentiment Indicators

Table 2.2: Data Sources, Vintages and Transformations By Country

	France	Germany	Italy	Japan	UK	Transformation
Gross Domestic Product (GDP) (a)	Q1-90	Q1-91	Q1-91	Q1-94	Q1-90	Q%C
Index of Production (a)	Jan-90	Jan-91	Jan-90	Jan-94	Jan-90	3m/3m%C
Manufacturing PMI (b)	Apr-98	Apr-96	Jun-97	Oct-01	Jan-92	n/a
Factory Orders (c)	n/a	Jan-91	n/a	n/a	n/a	3m/3m%C
Construction PMI (b)	Sep-00	Sep 99	Jul 99	n/a	Apr-97	n/a
Index of Services (d)	n/a	n/a	n/a	n/a	Jan-95	3m/3m%C
Services PMI (b)	May-98	Jun-97	Jan-98		Jul-96	n/a
Retail Sales Volumes (a)	Jan-90	Jan-91	Jan-91	Jan-94	Jan-90	3m/3m%C
Consumer Confidence (e)	n/a	n/a	n/a	Jan-94	Jan-90	n/a
Exports (a)	Jan-90	Jan-91	Jan-91	Jan-94	Jan-60	3m/3m%C
Manufacturing PMI: Capital Goods (b)	Apr-98	Apr-96	Jun-97	Oct-01	Jan-92	n/a
Economic Sentiment Index (e)	Jan-90	Jan-91	Jan-91	n/a	n/a	n/a
Economy Watchers' Survey (f)	n/a	n/a	n/a	Jan-00	n/a	n/a

(a) Source: OECD Main Economic Indicators; (b) Source: Markit Economics; (c) Source: Deutsche Bundesbank;
(d) Source: Office for National Statistics; (e) Source: European Commission; (f) Source: Cabinet Office;
Q%C = Quarterly Percent Change
3m/3m%C = Three Monthly Percent Change

With data collected, transformations of the data are produced where applicable to ensure stationarity. Variables such as retail sales, industrial production, trade and similar indicators are transformed into three-month on three-month percent changes in line with the preferred measure of GDP (quarter-on-quarter percent changes). For the survey data, no transformations are required. Table 2.2 provides details on data sources, transformations and respective time series histories for each country.

2.4 Nowcasting Results

To assess the impact of new information on the current quarter's weighted nowcast, the evolution of the ratio of the mean absolute forecasting error (MAFE) to the benchmark AR(1) MAFE is tracked. Readings higher than one (below one) indicate under (above) performance of the bridging equation model relative to the benchmark (in terms of forecasting current quarter GDP). Any instances

readings for Europe. Rather than flash readings, "final" readings are used.

where one model significantly outperforms the other at the 5% level according to the Diebold and Mariano (1995) test of predictive accuracy are marked with an astrix (note such tests were conducted across the whole sample, not for the sub-samples). Figures 2.1-2.5 show how the ratio typically evolves for each country over the whole sample (2006-2012Q2) and various sub-samples including the period of severe recession during 2008-2009. Table 2.3 offers a selection of MAFE ratios alongside the actual MAFE for the naive benchmark.

The first point to note is the high frequency data contain valuable information and help to improve nowcast accuracy as the quarter evolves (see data columns three and four). From the release of the manufacturing and investment PMI data on the first working day of a quarter ("day one") until the final piece of high frequency data is released just prior to the first estimate of GDP around four-to-five months later (the "final" day), there is a decrease in MAFE's and an improvement in nowcast accuracy relative to the benchmark.

This is consistent with expectations and previous research. High frequency data contain valuable information which helps to improve intra-period forecasting of GDP. As more information is accumulated leading up to the first estimate of GDP and monthly forecasts replaced (and those remaining improved) by actual observations, nowcast accuracy rises. In this particular application, the gains reach around 30% for the sample as a whole, with the exception of the UK where the improvement is considerably lower at 19.4%.[3]

Remaining on a general theme, there are noticeable deteriorations in absolute nowcasting performance over the period that encapsulates the depths of the fi-

[3]This may reflect in part earlier releases of GDP estimates in the UK, which can be as quick as three weeks after a quarter's end compared to typically six in Europe and Japan. Quarterly GDP changes in the UK were also found to be highly correlated with their first lag, leading to a reasonable performance of the benchmark model, which therefore made it more difficult to "beat".

Table 2.3: MAFE Ratios By Country

	2006-2012Q2				2006-2007				2008-2009)				2010-2012Q2			
	AR (MAFE)		Model (Ratio)		AR (MAFE)		Model (Ratio)		AR (MAFE)		Model (Ratio)		AR (MAFE)		Model (Ratio)	
	Day 1	Final Day	Day 1	Final Day	Day 1	Final Day	Day 1	Final Day	Day 1	Final Day	Day 1	Final Day	Day 1	Final Day	Day 1	Final Day
France	0.37	0.39	0.91	0.67*	0.25	0.33	0.89	0.63	0.64	0.57	0.88	0.63	0.25	0.30	1.00	0.77
Germany	0.69	0.74	1.03	0.69*	0.26	0.26	0.87	0.86	1.38	1.48	1.09	0.65	0.48	0.55	0.97	0.69
Japan	0.85	0.85	0.92	0.69*	0.34	0.35	0.91	0.98	1.61	1.51	0.90	0.56	0.66	0.73	0.88	0.75
Italy	0.54	0.53	0.92	0.72*	0.30	0.35	1.00	0.74	1.02	0.97	0.88	0.70	0.41	0.34	0.80	0.75
UK	0.51	0.39	0.91*	0.81*	0.08	0.07	0.95	0.78	1.00	0.68	0.95	0.81	0.47	0.41	0.85	0.80

[a] The model nowcasts are compared with first estimates of quarter-on-quarter changes in GDP, any difference being the forecasting (or nowcasting) error. These nowcasting errors are collated over the sample period (2006-2012Q2) and noted sub-samples to create the Mean Absolute Forecasting Error (MAFE) statistic. Differences between quarterly changes in GDP and a simple benchmark model are also recorded. This benchmark is an AR(1) of quarter-on-quarter changes in GDP. Model ratio is the MAFE of respective nowcasting models relative to the MAFE of the benchmark model.

[b] "Day 1" is the first day of the calendar quarter. "Final Day" refers to the receipt of final high frequency information before the first release of current quarter GDP.

[c] An astrix means that one model significantly outperforms the other at the 5% level according to a Diebold-Mariano test (these were conducted over the whole sample only).

nancial crisis (2008-2009). This can be seen in the sharp increases in the MAFEs of the naive AR(1) benchmark models between the 2006-2007 and 2008-2009 sub-samples across all countries (compare data columns one and two, with nine and ten). The ability to nowcast GDP was clearly much more difficult during this period, but it seems the high frequency data played a stronger role. Reductions in nowcasting MAFEs compared to the benchmark are generally greater over the sub-sample period of 2008-2009 for all countries, with the exception of the UK. For Japan and Germany, the out-performances are particularly marked. This could be a reflection of the especially sharp deterioration in the performance of the benchmark models for these two countries (note the widening of the benchmark's model MAFE in 2008-2009 compared to 2006-2007). The structure of the financial crisis, whose most immediate effect was to impact on world trade, hit countries with a greater reliance on exports as a source of economic growth especially hard.

Turning to the role of individual data releases in the evolution of forecast errors, tables 2.4-2.6 lists the biggest reducers of the MAFE ratios for each country again over the whole sample, and various sub-samples. In a real-time data setting, some common themes emerge.

Firstly, given the ADL specification of the nowcasting regression, the release of GDP data are naturally important in reducing nowcast errors.

Secondly, early releases of industrial production data tend to be the most important high frequency indicators in reducing errors when nowcasting quarterly changes in GDP. This is not surprising; statistics offices rely on supply-side indicators such as industrial production data when formulating early estimates of GDP. They would be expected to have a considerable positive influence on nowcasting model performance.

One may suppose hard indicators are the most important releases to watch. How-

Table 2.4: Reductions in MAFE Ratios: France and Germany Top 10 Data Releases

	2006-2012Q2		2006-2007		2008-2009		2010-2012Q2	
France								
1	GDP (Q-1, 1st Est)	-9.5%	GDP (Q-1, 1st Est)	-27.7%	TRA (M2, Q0)	-5.5%	GDP (Q-1, 1st Est)	-19.0%
2	IP (M2, Q0)	-3.9%	IP (M2, Q0)	-6.4%	IP (M2, Q0)	-4.9%	SPMI (M2, Q0)	-6.3%
3	CPMI (M1, Q0)	-3.2%	TRA (M2, Q0)	-4.7%	MPMI (M1, Q0)	-4.3%	CPMI (M2, Q0)	-4.5%
4	TRA (M3, Q0)	-2.4%	IP (M3, Q0)	-3.8%	CPMI (M1, Q0)	-4.2%	SPMI (M1, Q0)	-3.8%
5	TRA (M2, Q0)	-2.4%	CPMI (M1, Q0)	-2.3%	MPMI (M2, Q0)	-3.4%	IP (M1, Q0)	-3.8%
6	ESI (M1, Q0)	-2.0%	MPMI (M2, Q0)	-2.0%	TRA (M3, Q0)	-2.9%	MPMI (M1, Q+1)	-3.1%
7	MPMI (M2, Q0)	-2.0%	TRA (M1, Q0)	-1.9%	IP (M3, Q-1)	-2.6%	TRA (M3, Q0)	-3.0%
8	SPMI (M2, Q0)	-2.0%	SPMI (M3, Q0)	-1.9%	SPMI (M3, Q0)	-2.2%	TRA (M3, Q-1)	-2.7%
9	SPMI (M3, Q0)	-1.5%	GDP (Q-1, 2nd Est)	-1.6%	ESI (M1, Q0)	-2.2%	ESI (M1, Q0)	-2.4%
10	SPMI (M1, Q0)	-1.1%	ESI (M1, Q0)	-1.1%	ESI (M1, Q+1)	-1.9%	CPMI (M1, Q0)	-2.3%
Germany								
1	IP (M1, Q0)	-5.4%	Tra (M3, Q-1)	-12.5%	IP (M1, Q0)	-7.8%	FO (M3, Q-1)	-9.7%
2	IP (M2, Q0)	-5.2%	MPMI (M3, Q0)	-5.9%	IP (M2, Q0)	-6.2%	TRA (M2, Q0)	-8.3%
3	GDP (Q-1, 1st Est)	-5.1%	CPMI(M2, Q0)	-5.8%	GDP (Q-1, 1st Est)	-6.1%	IP (M3, Q-1)	-8.0%
4	FO (M3, Q-1)	-4.2%	SPMI (M3, Q-1)	-5.2%	FO (M1, Q0)	-5.1%	CPMI(M1, Q0)	-6.4%
5	CPMI(M1, Q0)	-3.6%	CPMI(M1, Q0)	-4.2%	FO (M2, Q0)	-4.4%	IP (M2, Q0)	-5.1%
6	TRA (M2, Q0)	-2.7%	SPMI (M2, Q0)	-2.5%	MPMI (M1, Q0)	-2.8%	GDP (Q-1, 1st Est)	-4.9%
7	FO (M2, Q-1)	-2.0%	FO (M2, Q-1)	-2.4%	FO (M2, Q-1)	-2.6%	ESI (M1, Q0)	-4.8%
8	IP (M2, Q-1)	-1.7%	FO (M3, Q-1)	-2.0%	IP (M2, Q-1)	-2.6%	Tra (M1, Q0)	-2.9%
9	Tra (M1, Q0)	-1.6%	MPMI (M1, Q0)	-1.9%	CPMI(M1, Q0)	-2.4%	IP (M1, Q0)	-2.7%
10	FO (M1, Q0)	-1.6%	Tra (M1, Q0)	-1.3%	FO (M3, Q-1)	-2.3%	SPMI (M3, Q-1)	-1.7%

The table shows the average percent reduction in the MAFE statistic over the sample (2006-2012Q2) and sub-samples for specific data releases. The nowcasts are compared with first estimates of Q%C in GDP, any difference being the forecasting (or nowcasting) error. These are collated over the sample period and sub-samples to create the Mean Absolute Forecasting Error (MAFE). Differences between Q%C in GDP and a benchmark are also recorded. This benchmark is an AR(1) of Q%C in GDP. Model ratio is the MAFE of nowcasting models relative to the benchmark MAFE;

M1 = Month 1; M2 = Month 2; M3 = Month 3; Q-1 = Previous Quarter; Q0 = Current Quarter; Q+1 = Following Quarter

Table 2.5: Reductions in MAFE Ratios: Italy and Japan Top 10 Data Releases

	2006-2012Q2		2006-2007		2008-2009		2010-2012Q2	
Italy								
1	IP (M1, Q0)	-5.9%	IP (M1, Q0)	-6.6%	IP (M2, Q0)	-6.7%	IP (M1, Q0)	-9.9%
2	Man/Inv PMI (M1, Q0)	-4.1%	GDP (Q-1, 1st Est)	-6.3%	IP (M3, Q-1)	-4.6%	SPMI (M3, Q-1)	-7.6%
3	SPMI (M3, Q0)	-3.1%	GDP (Q-1, 2nd Est)	-5.4%	MPMI (M1, Q0)	-4.6%	MPMI(M1, Q0)	-4.5%
4	IP (M3, Q0)	-2.7%	IP (M3, Q0)	-4.8%	IP (M1, Q0)	-3.8%	ESI (M1, Q0)	-3.1%
5	IP (M3, Q-1)	-2.6%	SPMI (M3, Q0)	-3.0%	SPMI (M3, Q0)	-3.5%	TRA (M2, Q0)	-3.0%
6	IP (M2, Q0)	-2.2%	MPMI (M1, Q0)	-2.3%	TRA (M1, Q0)	-3.3%	IP (M3, Q-1)	-2.3%
7	GDP (Q-1, 2nd Est)	-2.1%	CPMI (M2, Q0)	-1.9%	IP (M3, Q0)	-2.1%	IP (M3, Q0)	-2.2%
8	ESI (M1, Q0)	-1.8%	SPMI (M3, Q-1)	-1.4%	GDP (Q-1, 2nd Est)	-2.0%	SPMI (M3, Q0)	-2.2%
9	SPMI (M3, Q-1)	-1.8%	ESI (M1, Q0)	-1.4%	CPMI (M3, Q-1)	-1.8%	ESI (M1, Q+1)	-2.1%
10	TRA (M1, Q0)	-1.4%	IP (M2, Q-1)	-1.2%	MPMI (M2, Q0)	-1.7%	SPMI (M1, Q+1)	-2.0%
Japan								
1	IP (M3, Q-1)	-6.2%	MPMI (M2, Q0)	-6.3%	IP (M1, Q0)	-14.5%	IP (M3, Q0)	-9.7%
2	IP (M1, Q0)	-4.6%	GDP (Q-1, 2nd Est)	-5.7%	IP (M3, Q-1)	-7.0%	GDP (Q-1, 1st Est)	-6.4%
3	IP (M3, Q0)	-4.0%	CC (M1, Q0)	-3.5%	Tra (M2, Q0)	-5.9%	EWS (M3, Q-1)	-5.3%
4	Tra (M2, Q0)	-2.3%	IP (M3, Q-1)	-1.6%	MPMI (M1, Q0)	-5.4%	IP (M3, Q-1)	-3.5%
5	EWS (M3, Q-1)	-2.1%	MPMI (M3, Q0)	-1.4%	IP (M2, Q0)	-4.2%	EWS(M3, Q0)	-3.0%
6	GDP (Q-1, 1st Est)	-1.4%	RS (M3, Q0)	-1.1%	EWS(M1, Q0)	-3.5%	MPMI (M3, Q0)	-2.2%
7	Man/Inv PMI (M3, Q0)	-1.3%	Tra (M1, Q0)	-0.4%	RS (M3, Q-1)	-1.6%	RS (M2, Q0)	-1.8%
8	RS (M3, Q-1)	-1.2%	RS (M1, Q0)	-0.3%	Tra (M1, Q0)	-1.5%	MPMI (M2, Q0)	-1.5%
9	IP (M2, Q0)	-1.2%	IP (M3, Q0)	-0.2%	MPMI (M3, Q0)	-1.3%	CC (M2, Q0)	-1.5%
10	EWS(M3, Q0)	-1.1%	CC (M3, Q0)	-0.2%	RS (M1, Q0)	-1.0%	GDP (Q-1, 2nd Estimate)	-1.1%

The table shows the average percent reduction in the MAFE statistic over the sample (2006-2012Q2) and sub-samples for specific data releases. The nowcasts are compared with first estimates of Q%C in GDP, any difference being the forecasting (or nowcasting) error. These are collated over the sample period and sub-samples to create the Mean Absolute Forecasting Error (MAFE). Differences between Q%C in GDP and a benchmark are also recorded. This benchmark is an AR(1) of Q%C in GDP. Model ratio is the MAFE of nowcasting models relative to the benchmark MAFE;
M1 = Month 1; M2 = Month 2; M3 = Month 3; Q-1 = Previous Quarter; Q0 = Current Quarter; Q+1 = Following Quarter

Table 2.6: Reductions in MAFE Ratios: UK Top 10 Data Releases

UK	2006-2012Q2		2006-2007		2008-2009		2010-2012Q2	
1	IP (M1, Q0)	-6.1%	IP (M2, Q0)	-12.7%	IP (M1, Q0)	-6.1%	IP (M1, Q0)	-8.0%
2	IoS (M1, Q0)/GDP (Q-1, 3rd Est)	-3.7%	SPMI (M1, Q0)	-9.0%	IP (M2, Q0)	-4.5%	IoS (M1, Q0)/GDP (Q-1, 3rd Est)	-7.5%
3	CPMI(M1, Q0)	-2.9%	CPMI(M3, Q0)	-3.0%	IP (M3, Q-1)	-2.7%	CPMI(M1, Q0)	-5.8%
4	IP (M3, Q-1)	-2.6%	CPMI (M3, Q-1)	-2.4%	CPMI(M2, Q0)	-1.8%	IP (M3, Q-1)	-2.5%
5	SPMI (M1, Q0)	-1.8%	MPMI (M1, Q0)	-2.4%	MPMI (M1, Q0)	-1.7%	CC (M3, Q0)	-1.8%
6	MPMI (M1, Q0)	-1.1%	IP (M3, Q-1)	-2.2%	IoS (M1, Q0)/GDP (Q-1, 3rd Est)	-1.6%	SPMI (M1, Q0)	-1.4%
7	IP (M2, Q0)	-1.1%	MPMI (M3, Q0)	-2.0%	SPMI (M1, Q0)	-1.3%	RS (M1, Q0)	-1.2%
8	CPMI(M2, Q0)	-0.6%	SPMI (M2, Q0)	-1.4%	Man/Inv PMI (M2, Q0)	-1.1%	CC (M2, Q0)	-1.1%
9	CPMI(M3, Q0)	-0.5%	CPMI(M1, Q0)	-1.3%	CPMI(M1, Q0)	-1.0%	SPMI(M3, Q0)	-0.8%
10	RS (M1, Q0)	-0.5%	RS (M3, Q0)	-0.9%	IP (M2, Q-1)	-0.9%	RS (M3, Q0)	-0.6%

The table shows the average percent reduction in the MAFE statistic over the sample (2006-2012Q2) and sub-samples for specific data releases. The nowcasts are compared with first estimates of Q%C in GDP, any difference being the forecasting (or nowcasting) error. These are collated over the sample period and sub-samples to create the Mean Absolute Forecasting Error (MAFE). Differences between Q%C in GDP and a benchmark are also recorded. This benchmark is an AR(1) of Q%C in GDP. Model ratio is the MAFE of nowcasting models relative to the benchmark MAFE;

M1 = Month 1; M2 = Month 2; M3 = Month 3; Q-1 = Previous Quarter; Q0 = Current Quarter; Q+1 = Following Quarter

ever, consideration must also be made of timing and order of release. Preceding official data publications there is a wealth of information for equivalent time periods available. And this is where the value of survey data is derived.

They generally improve nowcasts by providing a steer on the likely direction of quarterly changes in GDP ahead of official data releases. The PMI data seem particularly important in this regard, substantially reducing MAFEs. This is especially the case early in the quarter when little information is available on current quarter economic activity.

Figures 2.1-2.5 highlight this visually, but what is perhaps of some surprise is the PMIs (and for that matter other surveys such as the Economy Watchers' Survey for Japan and Economic Sentiment Indicators in Europe) make a number of appearances in the top 10 data releases for reducing nowcast errors, with the observations of the high frequency data relating to the third month of the previous quarter or the first two months of the quarter being forecast. These indicators are generally not only adding more value than hard indicators such as trade and retail sales data with regards timing, but also net reductions in MAFEs.

The value of the survey data and the role of timing in nowcasting models was also enhanced during the Great Recession.

Throughout tables 2.4-2.6, for the sub-sample period of 2008-2009, PMI data are again widely referenced. The highest PMI impacts are from the manufacturing surveys for the first month of the quarter. This is the one of the earliest high frequency indicators released and has a big positive impact on reducing MAFEs. The suggestion is that PMIs provided key early indications of the severe stress that was being experienced in the real economy during the midst of the financial crisis. Official data releases perhaps added the quantitative colour, but not until several weeks later. Policymakers didn't or don't usually have the luxury of waiting for that confirmation when reacting to fast moving events. The surveys played in a timely manner an invaluable role in highlighting the impact on economic activity of the chaos engulfing financial markets.

2.5 Nowcasting: Alternative Methods and Comparisons

To place this chapter's empirical findings into some context, this section compares and contrasts against recent developments in the nowcasting literature. Some discussion of the methodological approach taken is also included.

As noted in the introduction, nowcasting principally involves solving two econometric problems: mixed time frequencies and a dataset's ragged edge in "real-time".

Bridging equations deal with these challenges in simple and easy to implement ways: temporal aggregation is performed by using straight averages to convert the high frequency data to the lower frequency sampling rate, while any "missing observations" tend to be estimated via univariate forecasting techniques.

Attributes such as simplicity and ease of implementation deserve recognition when considering the practical application of modelling techniques. And the ongoing use of bridging equations in central banks as a means of tracking economic developments suggests practitioners do attach weight to such attributes (see e.g. the Bank of England (2013) for a recent example, where the authors use bridge equations to nowcast changes in world GDP growth with a small number of explanatory variables, including PMI data).

Conceptually, bridging equations are, however, seemingly sub-optimal in dealing with the two problems they are designed to address.

On temporal aggregation, the averaging of indicator variables (say over a quarter) provides the simplest method to convert higher frequency data into a lower sampling rate but at 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 a regression equation the explanatory variables at their original sampling rate. Under this approach, the econometrician is taking the view that all observations have unique coefficients. However, as the sampling rate rises then the problem of parameter proliferation becomes increasingly acute. As an example, if daily data on interest rates were used as an explanatory variable for nowcasting GDP, over a quarter there would be 66 separate observations (based on 22 trading days each month). The number of co-efficients needed for estimation starts to become unfeasibly large (even more so if lags of the explanatory variables were included).

As shown in the introduction and the previous chapter, Ghysels, Santa-Clara, and Valkanov (2004) propose a solution that finds some middle ground by introducing the so-called MIDAS (Mixed Data Sampling) regression. The proposal is to take a weighted sum of the high frequency variables that are linked to a low frequency variable through a single coefficient in a single regression equation. This avoids both the problems of parameter proliferation and, in the world of nowcasting, avoids the need for the forecasting of missing observations as per the bridging equation frameworks.

Specific applications of MIDAS to nowcasting GDP are found in Armesto, Engemann, and Owyang (2010) and Clements and Galvao (2008) for the United States. Kuzin, Marcellino, and Schumacher (2011) and Kuzin, Marcellino, and Schumacher (2013) augment the approach by adding an AR term (referred to as an AR-MIDAS model) and nowcast euro area GDP and other industrialised countries.

All support findings from applications of bridging equations: higher frequency indicator data provide important within period information in improving near-term nowcasts of economic output. Moreover, gains in nowcasting performance also tend to be realised through the pooling of results for individual MIDAS specifications as the regressions can realistically only handle a small number of variables.

The research conducted by Kuzin, Marcellino, and Schumacher (2013), which is performed as a pseudo real-time exercise due to the large number of variables, also breaks out the period of severe recession (2008-2009) and compares to the more benign economic environment of preceding years. In line with the empirical results of this paper, there is a sharp deterioration in absolute nowcasting performance during 2008-2009, but nonetheless considerable out-performance of a naive benchmark model. This adds further weight to the view that high frequency data were especially important during this period. Moreover, an interesting by-product is that the usage of latest published series, rather than vintage data, does not seem to have a significant impact on overall conclusions (in line with the author's assertions).

Despite the theoretical persuasiveness of MIDAS regressions direct comparisons between bridging equations and MIDAS models (which seem relatively rare in the literature) show little discernible difference in terms of nowcast accuracy.

While this is explored a little further in the following chapter, a hunch is that this could be a function of the use of monthly data, where only three readings are observed each quarter, limited lag lengths are employed and the univariate forecasting of missing values is conducted over a relatively short period.

The conceptual advantages of MIDAS modelling are therefore expected to come to the fore when sampling frequencies between explanatory and dependent variables are greater. An obvious example here is in the use of high frequency financial market data, where MIDAS regressions can be an effective tool at extracting previously hard to reach information. With this in mind, applications by Andreou, Ghysels, and Kourtellos (2013) for US GDP and Libero and Moretti (2013) for euro area inflation provide interesting, recent and relevant examples.

Bridging and MIDAS methods are single equation frameworks. Forecasts within period and future observations of the target variable are based on direct links between current (and lagged) values of indicator variables. However, macroeconomists are now faced with literally hundreds of data series to measure current economic conditions in real-time. If one was to use bridging equations or MIDAS models for all available series, then these approaches would rapidly become cumbersome. Practitioners are subsequently tempted to track only a handful of data series to maintain a manageable modelling framework. This was ultimately the approach taken in this chapter's empirical work, although datasets were naturally hamstrung given the desire to use vintage data series and their limited availability across countries.

However, it is recognised overall model performance ultimately rests on the performance of this handful of macroeconomic variables: there is a vulnerability to a breakdown in their relationship with GDP. And using a restricted dataset can also come at the cost of potentially important information being discarded.

So, rather than pre-selecting indicators, the information held within a large volume of predictors is replaced by a much smaller number of estimated factors.

Incorporating factors into a nowcasting framework should subsequently provide a solid foundation for understanding the current profile of economic output well ahead of less-timely GDP releases.

In an application for the US economy, Giannone, Reichlin, and Small (2008) introduce such a statistical framework that can deal simultaneously with the problems inherent in both nowcasting and large dataset dimensionality. Central to the GRS approach is that the high frequency variables have a factor structure that follows a vector autoregressive (VAR) process:

Crucial within this set up is that the Kalman filter and smoother can be used to extract the factors. Adopting such an approach means any missing observations are efficiently dealt within the framework. There is an added benefit of exploiting cross-sectional information from within the dataset.

Achieving a nowcast in this manner is similar to the second step in the bridging equation methodology (using the nowcasting parlance, the GRS approach is essentially "bridging with factors"). Alternatively, rather than some kind of temporal aggregation, the high frequency factor could be plugged into a MIDAS regression. Marcellino and Schumacher (2007) adopt such an approach to nowcast German GDP (with similar results to the integrated state-space framework).

There are empirical applications of the GRS-type model across many regions and countries, with a number suggestive that bridging equations are inferior nowcasting tools, with sizeable gains in nowcast accuracy to be secured from adopting a multivariate modelling strategy.

What is less clear, however, is the comparative performance of the two methods during 2008-2009. Although the literature has provided plenty of instances where both techniques have offered gains in nowcasting accuracy relative to simple benchmark specifications, results from many experiments are conducted within the period of the so-called Great Moderation. And the empirical results of this paper showed a clear deterioration in absolute nowcasting performance of bridging equations during 2008-2009, but a marked relative improvement to a simple benchmark.

With this in mind, Lombardi and Maier (2011), in an application for the euro area and its various member countries, as well as Mitchell (2009) for the UK, compare and contrast bridging and factor nowcasting specifications during the crisis period.

Benchmarks are more easily beat by both techniques in a relative sense compared to the final years of the Great Moderation. In line with the conclusions of this chapter's cross-country empirical application, timely information during high volatility periods is again shown to be especially important.

However, single equation nowcasting specifications based on qualitative data have an advantage over large-scale factor models as they can adapt relatively more quickly to changes in the economic environment: simplicity is their strength. While those models based on large datasets in the most part yield greater nowcast accuracy, they can suffer from problems such as persistence.

2.6 Chapter 2 Summary

In this chapter, the timing aspect for a small set of closely watched macroeconomic indicators was explored across five countries. Exploration was achieved through a simple nowcasting model, which successfully deals with the two principal econometric problems associated with this type of research; mixed time frequencies and the dataset's "jagged" edge. The approach uses a two-step bridging equation framework, and follows standard regression techniques which made the implementation of using sequential vintage series for several high frequency macro-economic indicators easier. Although the process proved to be data intensive and, at times, cumbersome, the flexibility of the framework allowed for a pure nowcasting application with a small number of explanatory indicators. An easy assessment of these various macroeconomic indicators from both perspectives of timing and impact on model accuracy was provided.

Of the high frequency indicators, the biggest reducer of MAFE ratios was industrial production data, not surprising given it's substantial role in the construction of preliminary GDP estimates. This was a common feature across the five countries.

However, when the timing aspect of a macroeconomic indicator is considered, the strength of survey data such as PMIs comes to the fore. These indicators provide guidance on the underlying direction of economic output, especially early in a quarter when little other information is known. The surveys can also have a bigger impact on improving nowcast accuracy than later released statistical releases such as retail sales.

High frequency data also provide invaluable information at times of severe economic stress. For policymakers, who are required to make decisions at more frequent intervals than some official data releases are available, the characteristic of timeliness and accuracy in understanding current macro-economic conditions is invaluable. Using only a small number of macro-economic variables revealed around 30%-40% reductions in MAFEs compared to a naive benchmark model for the period 2008-2009.

In the final section, the bridging equation methodology and the results was discussed in the context of two popular alternative methods from the literature: MIDAS regressions and complete state-space system solutions.

While these approaches offer conceptual advantages and persuasive theoretical improvements, the period of severe economic stress seen during 2008-2009 suggests there remains room to use variants of the bridging equation methodology. Indeed, a rounded, holistic, and watchful approach when formulating GDP nowcast expectations seems a sound strategy to pursue. Single-equation approaches based on PMIs and other timely survey data are likely to remain firmly within the policymaker's nowcasting toolbox.

Figure 2.1: Nowcast Model Evolution (France)

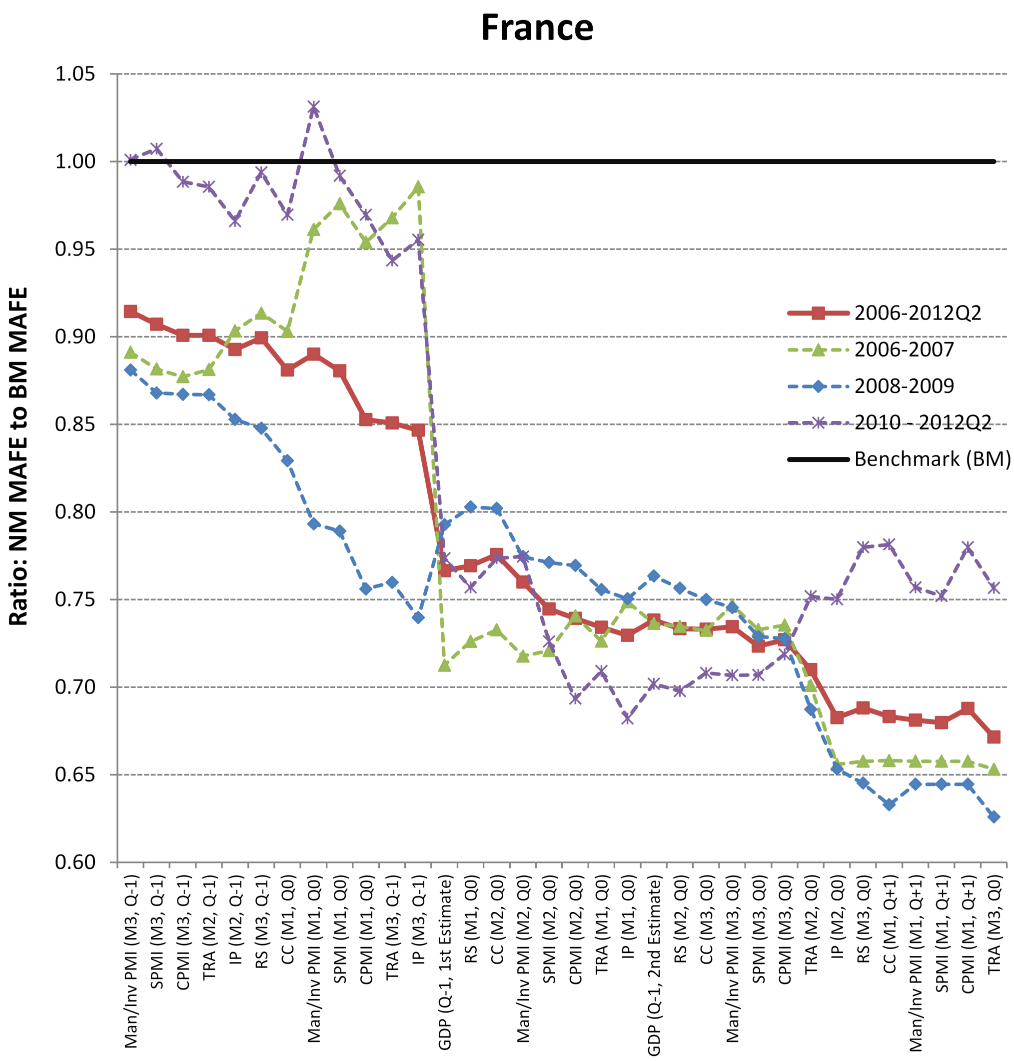


Figure 2.2: Nowcast Model Evolution (Germany)

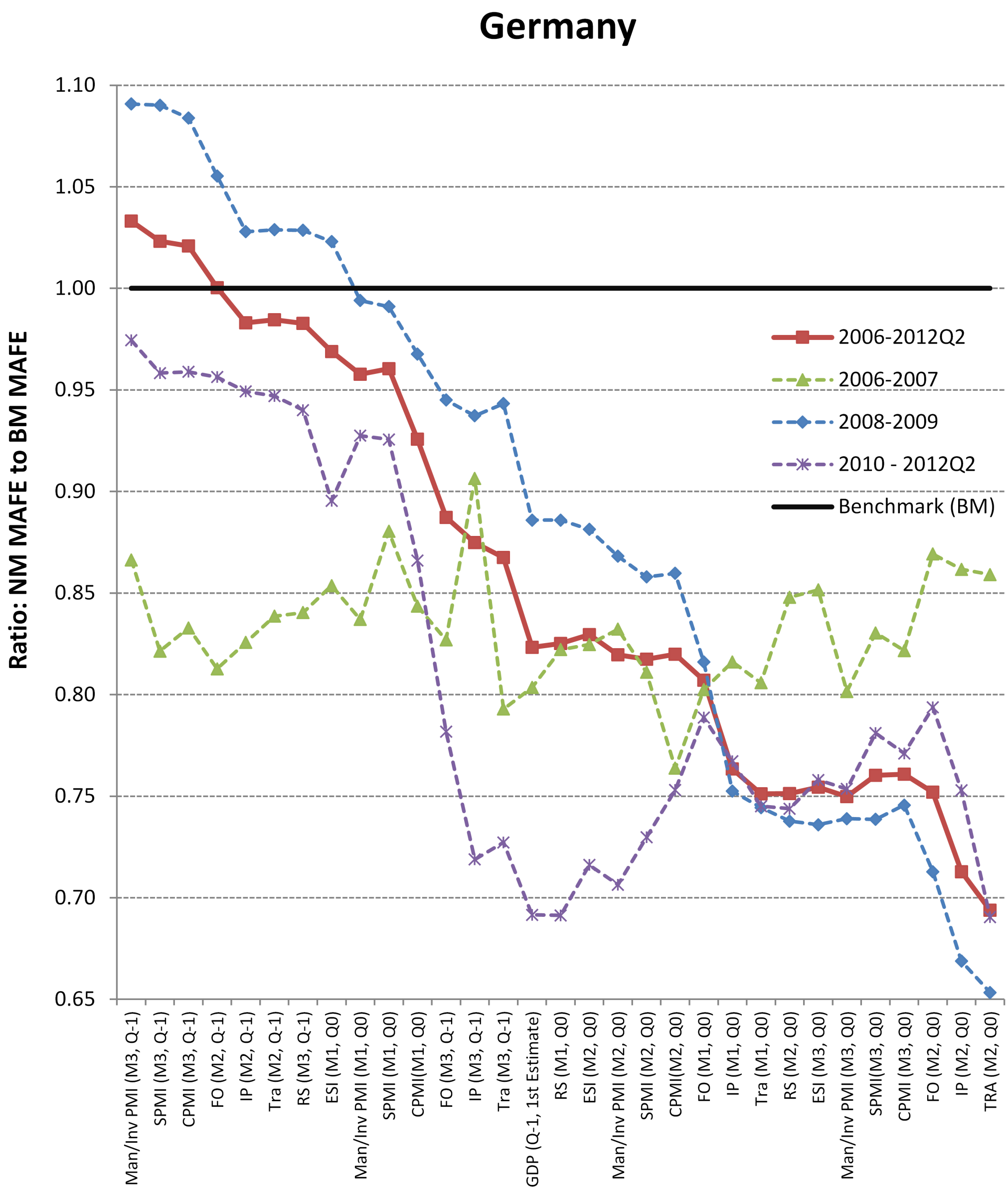


Figure 2.3: Nowcast Model Evolution (Italy)

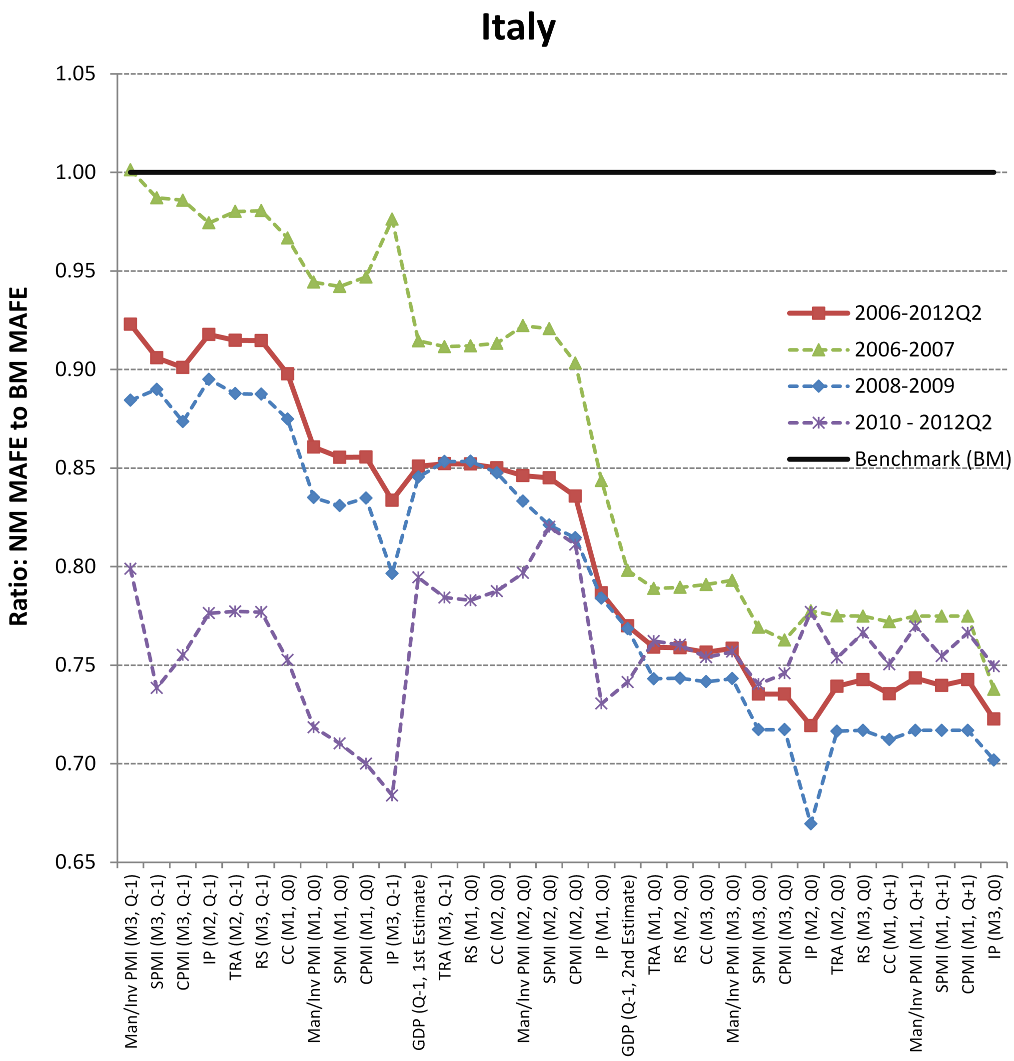


Figure 2.4: Nowcast Model Evolution (Japan)

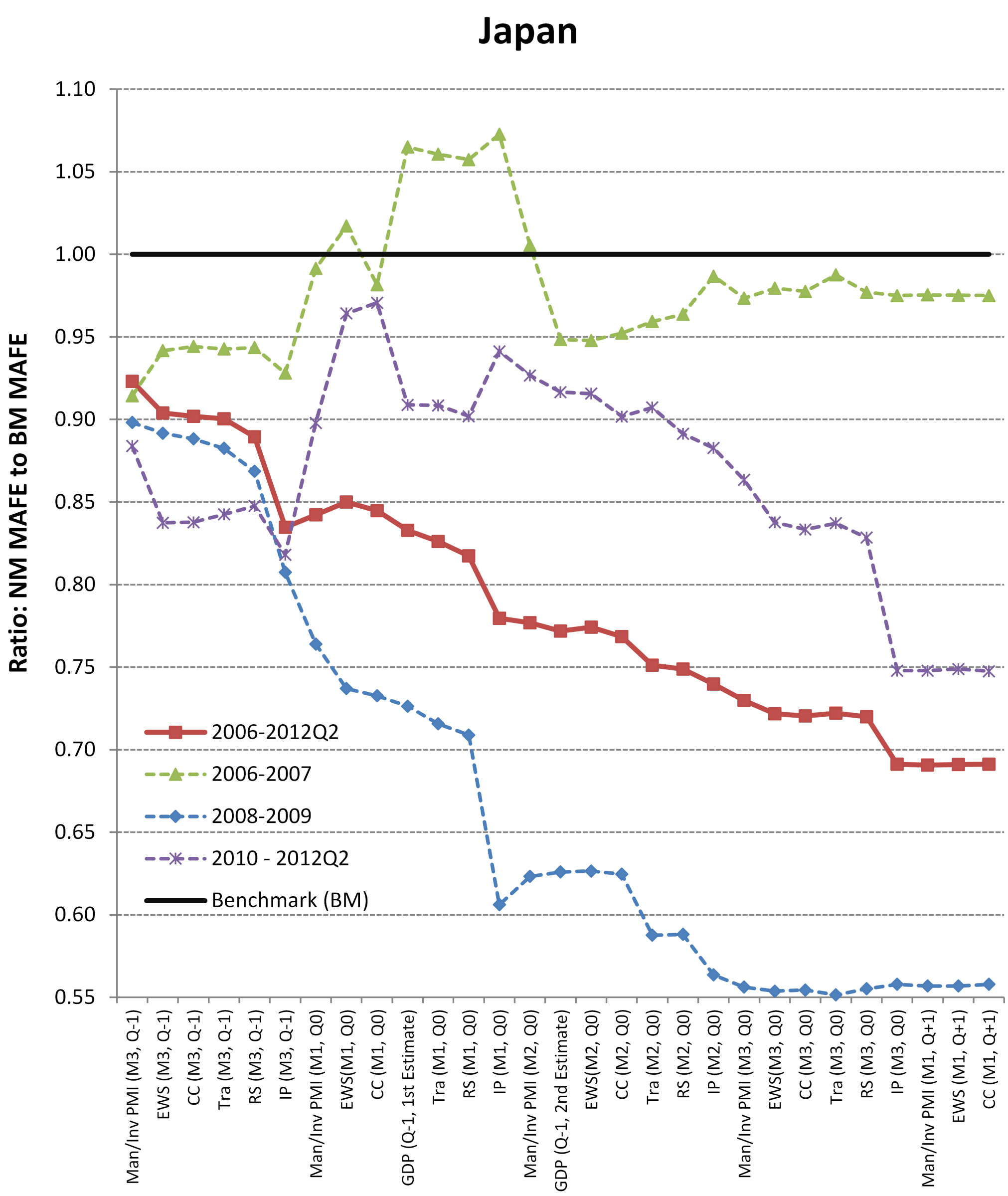
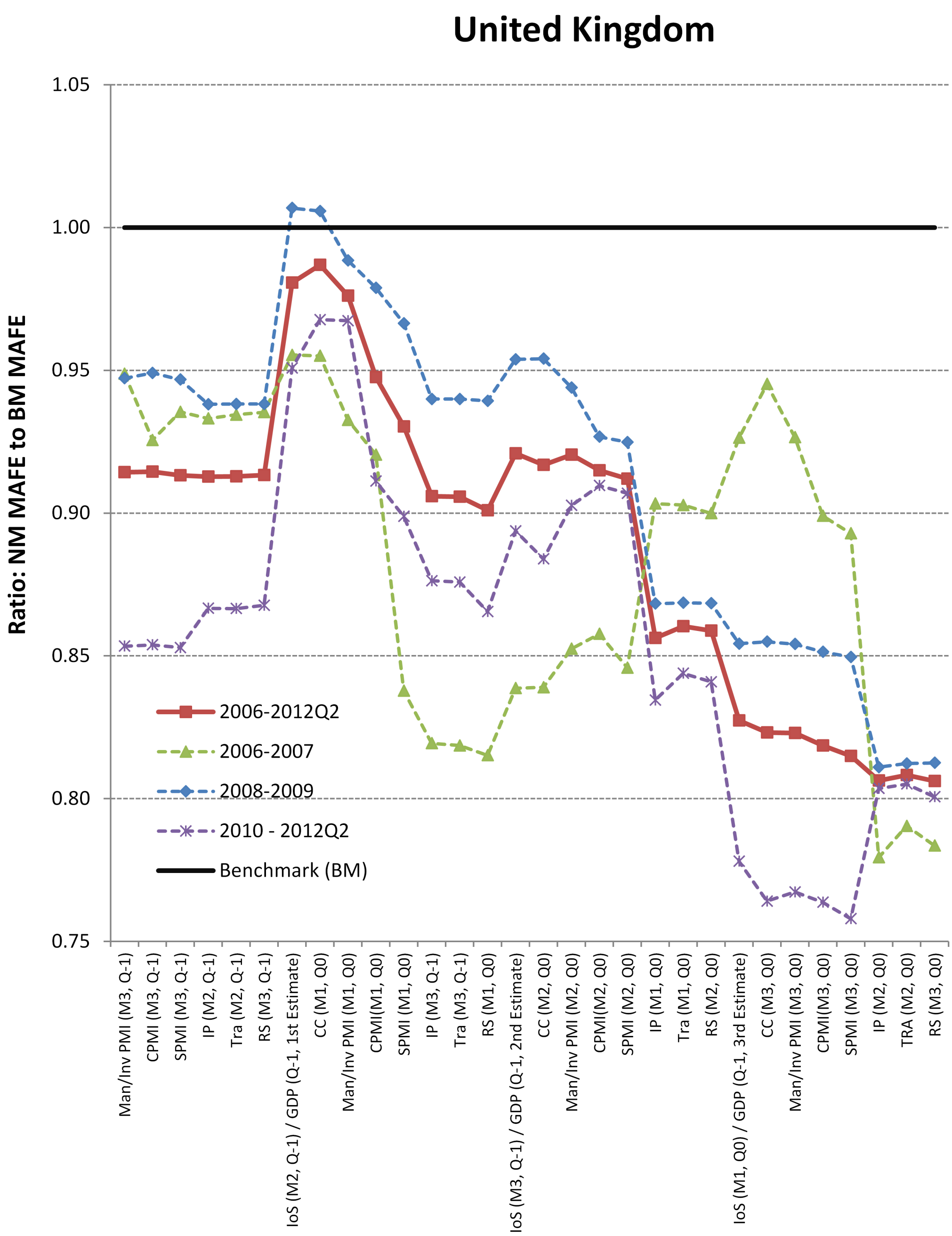


Figure 2.5: Nowcast Model Evolution (United Kingdom)



Chapter 3

Nowcasting UK GDP during the Depression

Chapter 3 takes a somewhat different tack to chapter 2 by reviewing 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.

This chapter 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 chapter 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 chapter 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.

3.1 Introduction

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?

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, which is the principal focus of this chapter. 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. So the contribution of the research is to 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:

[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).

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 3.2-3.4, 3.5 provides further empirical results of a re-running of the recursive out-of-sample nowcasting exercises. The difference, however, 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.

3.2 Model Frameworks

The aim of section 3.2 is to provide an overview of the various statistical models used in this chapter 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.

3.2.1 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:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \varepsilon_t \tag{3.1}$$

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:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \gamma X_t + \varepsilon_t \tag{3.2}$$

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

$$X_t = \frac{1}{m} \sum_{k=1}^{m} L_{HF}^{k} X_t^{HF} \tag{3.3}$$

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 3.2, 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:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \gamma \sum_{j=1}^{p} \phi(k : \theta) L_{HF}^{k} X_{t-h}^{HF} + \varepsilon_t \tag{3.4}$$

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 chapter'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:

$$\hat{Y}_t = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_{i,t} \tag{3.5}$$

3.2.2 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 chapter, 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:

$$X_t^{HF} = \Lambda f_t^{HF} + \varepsilon_t \tag{3.6}$$

where $f_t^{HF} = (f'_{1,t}, f'_{2,t} ... f'_{r,t})$ 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:

$$\hat{Y}_t = \alpha + \beta_1 Y_{t-1} + \gamma f_t + \varepsilon_t \tag{3.7}$$

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 3.3.

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

$$Y_t = \alpha + \beta_1 Y_{t-1} + \gamma \sum_{j=1}^{p} \phi(j : \theta) L_{HF}^{k} f_{t-h}^{HF} + \varepsilon_t \tag{3.8}$$

Finally, a note on the 24 series used to create f_t^{HF} (descriptions of which are provided in table 3.1).

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.

3.2.3 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 3.1). 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

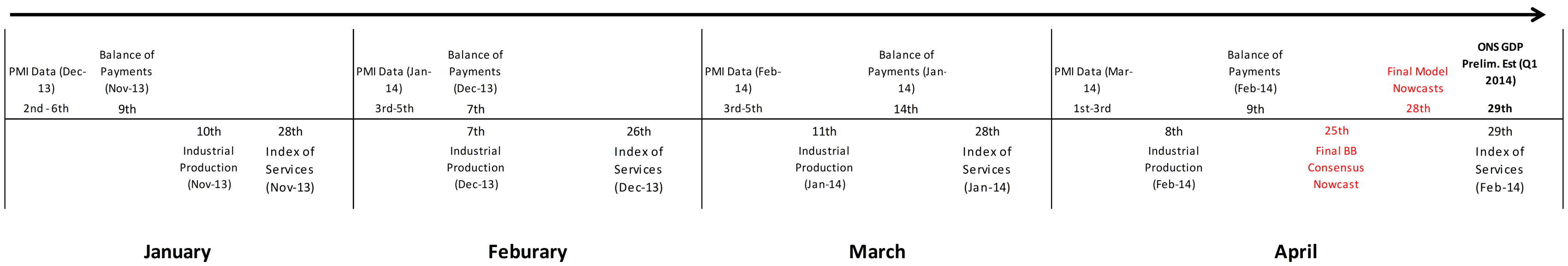
be available in some consensus form via Bloomberg's monthly surveys, the timing of the final poll and the release of its findings means this is the summary of the final best guesses of UK GDP growth for a specific quarter. The statistical models are run as if nowcasting UK GDP around a week prior to the first estimate so these should be viewed as being broadly comparable to the consensus.

3.3 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 3.3. 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".

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 prac-

Figure 3.1: Nowcasting UK GDP Timeline for Q1 2014



titioners 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 3.2. 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 3.4 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 timeframe (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 3.1.

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 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.

Table 3.1: 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

3.4 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 3.4.1 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 3.4.2 looks specifically at the relative performances of the bridging equations and MIDAS-based models. The section concludes with an examination of model performances relative to the market consensus view.

3.4.1 Modelling Accuracies

Table 3.1 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.

Table 3.2: 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.

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. 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 3.1, figure 3.2 provides a visualisation of selected model nowcasts relative to ONS preliminary estimates.

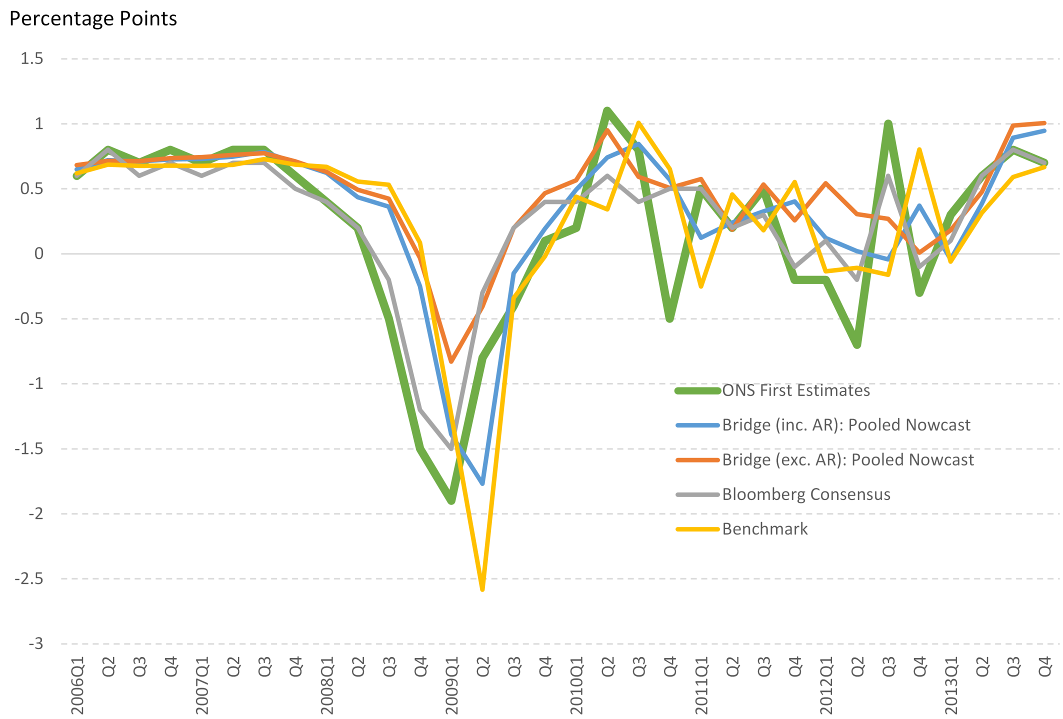
The weakness of the AR(1) benchmark is (not surprisingly) especially evident around turning points such as the trough of the severe recession in Q1 2009 and at other times of economic volatility (e.g. 2010-2012). These problems are also evident when an AR(1) component is incorporated into respective nowcasting models (see e.g blue line in figure 3.2).

When the AR(1) component is turned off, turning points are captured more easily and issues of lag dissipate. But magnitudes of change during 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 3.2). 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,

Figure 3.2: Selected Model Nowcasts Relative to ONS Preliminary GDP Estimates (Quarter-on-Quarter Growth)

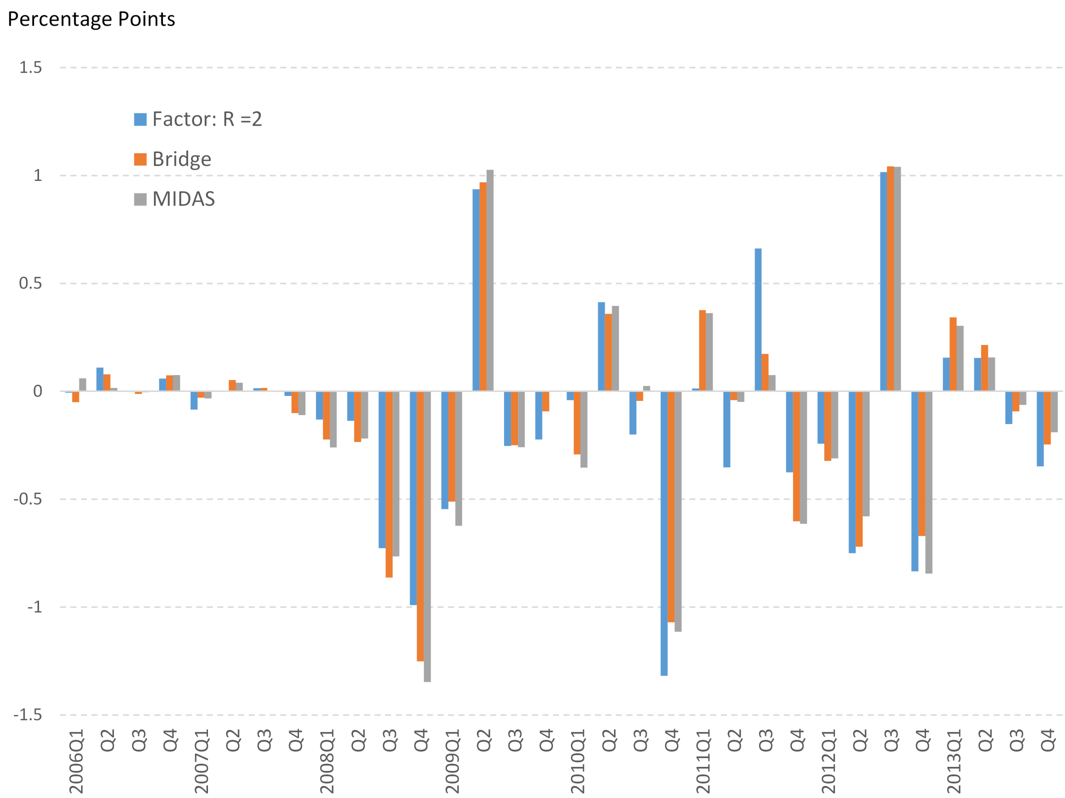


which tends to yield better performance on average. This again resonates with the literature.

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 mar-

Figure 3.3: Model Nowcast Errors



ket 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 3.3 for an illustration.

3.4.2 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 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).

3.4.3 The Consensus Nowcast

A standout result from table 3.1 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 chapter 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 3.4, but note if the consensus performance was compared against the latest GDP vintage (rather than real-time information) the nowcast accuracy quoted in table 3.2 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.

3.5 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.

3.5.1 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 3.4. 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.

3.5.2 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 revision 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”.

3.5.3 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 3.4 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 3.4, nowcast regressions were only produced for the latter. The results of this exercise are shown in table 3.3.

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 3.4. 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

Table 3.3: 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.

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 3.4 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 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.

3.6 Chapter 3 Summary

The question posed at the start of this chapter 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 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 3.3 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) and used extensively throughout this thesis.

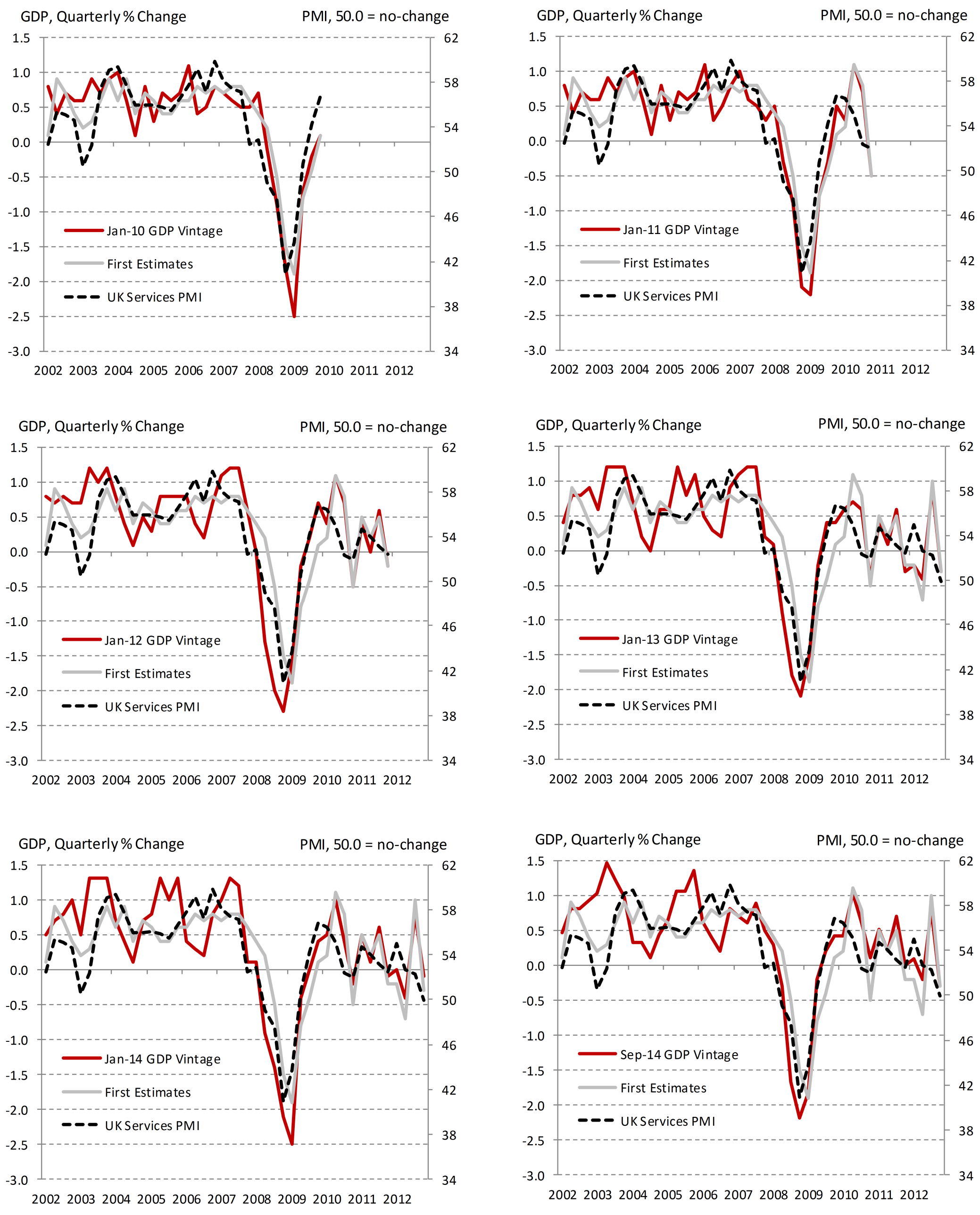
Nonetheless, using a replicated real-time dataset, several general findings from the nowcasting literature were found to be applicable for UK GDP. High frequency data are important in reducing nowcast uncertainty; pooling of small nowcasting regressions tend to provide greater overall accuracy than single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

Figure 3.4: The Changing Profile of UK GDP History



Chapter 4

Google's MIDAS Touch: Predicting UK Unemployment With Internet Search Data

In this final research chapter, attention returns to the subject of timing.

Chapter 2 showed that the timing of release was an important characteristic to consider when assessing the usefulness to researchers of an economic statistic in nowcasting applications.

The finding that timeliness is important has implications for a world where data is becoming available to quicker timescales than ever before. Based primarily on the traces of electronic information that people increasingly leave behind when interacting with others, so-called "big data" sources could provide the opportunity for economists to better understand "where are we now".

With this in mind, chapter 4 focuses on the potential of internet search data as a data pool for policymakers when formulating decisions based on their understanding of the current economic environment.

Earlier literature is built upon via a structured value assessment of the data provided by Google Trends. This is done through two empirical exercises related to the nowcasting of changes in UK unemployment.

Firstly, economic intuition provides the basis for search term selection, with a resulting Google indicator tested alongside survey-based variables in a traditional forecasting environment.

Secondly, this environment is expanded into a pseudo-time nowcasting framework which provides the backdrop for assessing the timing advantage that Google data have over surveys.

The framework is underpinned by a MIDAS regression which allows, for the first time, the easy incorporation of internet search data at its true sampling rate into a nowcast model for predicting unemployment.

4.1 Introduction

In 2013, 36 million or 73 percent of UK adults accessed the internet every day, some 20 million more than just seven years previously. With this increased penetration has come an associated rise in day-to-day usage for activities such as finding information about goods and services (up to 66 percent from 58 percent), or as a tool to find a new job (in 2013, 67 percent of unemployed adults looked online for a job or submitted a job application).[1]

Many of these users, it would seem, use a search engine as their portal into the online world. These services essentially act as an intermediary by bringing together web users via terms entered into a query box. Google Inc, which is the dominant force in search engine provisions, processes hundreds of millions of such terms and queries on a daily basis. Such is the popularity of the eponymous search engine, the term "Google" now enters the Oxford Dictionary as a verb ("to Google").

The data associated with online search activity offers a number of opportunities

[1] Office for National Statistics: Internet Access - Households and Individuals (2013)

for the researcher. If an increased number of people are, for example, searching online for flat screen TVs, could this provide an indication that purchasing of such goods will soon rise and offer an early insight into consumer spending activities? Or if there is an increase in searches for benefits associated with unemployment, could this give an early indication of rises in joblessness?

And if such shifts in behaviour are observed ahead of more traditional sources such as surveys or backward looking official data, then timely internet based information could be used to make better and more optimal decisions in areas such as monetary policy or investment.

4.2 Internet Search Data In the Literature

Investigations of search data have been applied across a number of fields such as predicting changes in tourism numbers (Bangwayo-Skeete and Skeete 2015; Yang et al. 2015), offering an advance warning of flu epidemics (Ginsberg et al. 2009) or predicting exchange rate volatility (Smith 2012).

In economics, Ettredge, Gerdes, and Karuga (2005) provide one of the first examples of using web-search data as a predictor of macroeconomic statistics, particularly unemployment figures, but it was the release by Google of its freely available service “Google Insights for Search” in 2008, later to be usurped by “Google Trends”, that gifted researchers a readily available platform to analyse search data.

Google’s own economists – Hyunyoung Choi and Hal Varian (Choi and Varian 2009a; Choi and Varian 2009b) – provided early illustrations of how Google Trends data can be used to give an advance indication of US retail and auto sales, new housing starts, travel destinations and initial claims for unemployment benefits. When comparing one step-ahead forecasting errors, adding search data as a regressor in simple seasonal auto-regressive and fixed effects models tends to out-perform those that exclude this variable. While in some cases the gains are only a few percent, for auto sales the improvement was substantial at 18 percent and for new housing starts the gain was 12 percent. For initial claims data, the gain in forecasting accuracy was as high as 16 percent.

Choi and Varian's papers spawned a number of related studies, with most applications generally based on predicting some kind of variable that can be linked to the behaviour of households, such as the consumption of goods or activity in the labour and housing markets (although recently Koop and Onorante (2014) explore the possibility of using Google search data to help improve nowcasts of macroeconomic variables using dynamic model selection methods).

Askitas and Zimmermann (2009) show how keyword searches correlate strongly with monthly German unemployment data and how a Google predictor can add value to an error prediction model, while Fondeur and Karame (2013) look at the usefulness of Google data in predicting youth unemployment in France. D'Amuri (2009) assesses the power of augmenting standard time-series models for quarterly Italian unemployment and concludes that the data improve out-of-sample forecasting performance. McLaren and Shanbhogue (2011) perform similar exercises for the UK labour and housing markets, comparing simple baseline AR specifications to those augmented with internet search variables. The authors go as far as suggesting that Google Trends data may contain information above and beyond those provided by survey indicators. Consumption based applications can be found in Kholodilin, Podstawski, and Silverstovs (2010) and Schmidt and Vosen (2009).

While there is a general consensus that the data are useful in various short-term forecasting (or "nowcasting") applications, equally there are number of challenges and pitfalls to overcome, particularly around the selection of search terms. Which terms are most relevant to predicting a target variable? What is the motivation of the user to enter a search term? Lazer et al. (2014) highlight that there needs to be careful consideration of social and independent searching. Is the user searching for their own purpose, or is the search more akin to some kind of herd behaviour (i.e. because many others are doing the same)? Such issues have considerable implications for forecasting ability and have been suggested as a key reason behind the persistent over-estimation by Google search data in predicting the number of flu cases in recent years (Bentley, Nyman, and Ormerod 2014).

4.2.1 Chapter Aims

In this chapter, the aim is to contribute to the debate through an empirical application that primarily considers the respective abilities of Google and competing survey-based models to forecast changes in unemployment.

The subject has been touched upon by McLaren and Shanbhogue (2011), but several refinements need to be applied to their approach, plus exploration of other avenues to gain a greater understanding of the role Google search data can play in macro-economic forecasting.

Firstly, the selection of search terms. McLaren and Shanbhogue (2011) used a single term "JSA" (Job Seekers' Allowance). This approach has a number of flaws. For example, the reliance on a single search-term seems a dangerous strategy especially as this acronym relates to a specific UK unemployment benefit, the name of which has been subject to various changes over time. With this in mind, some alternative strategies to term selection are proposed in section 4.3, based on economic intuition and a hybrid of ideas found within similar literature.

Analysis of the quality of Google indicators relative to survey-based variables occurs in section 4.4 through traditional linear regression equations that link the target variable (unemployment) with these explanatory variables. Statistics and tests around model specification, coefficient stability, and out-of-sample forecast performances are conducted.

A larger extension of the literature on the applicability of Google trends data in economics occurs in sections 4.5 and 4.6.

Google and survey indicators are analysed within a MIDAS regression framework designed to "nowcast" unemployment on a weekly basis over an eight-week "nowcasting" period.

Greater clarity on the purposes of nowcasting is provided later, but for now let nowcasting be defined simply as an effort to understand what has happened in the very near past, what is happening today or what is happening in the very near future. Such aims are generally achieved by linking a dependent variable to a dataset containing various soft (e.g. qualitative survey data) and hard explanatory variables (e.g. quantitative data) via some kind of econometric forecasting model. All of these indicators are assumed to be useful in predicting the dependent variable and, as new information on these predictors is released, then econometric models can be updated.

Moreover, Google information is available to quicker timescales than survey variables. This could be an important advantage. As shown in the previous chapter, the marginal predictive power of an indicator can be linked to its release schedule making it an important consideration when trying to understand its role and value to the economic forecaster.

Through the incorporation of MIDAS regressions into a pseudo-time nowcasting framework, there is an opportunity to understand how weekly data and its associated timing advantage over other variables can reduce nowcast uncertainty through time. MIDAS regressions are used because they offer a neat solution to the problem of mixed time frequencies that exist within the Google model frameworks: Google data are available weekly, yet unemployment data is released monthly. As far as is known, this is the first time that MIDAS regressions have been used with Google data for an economics specific application.

4.3 Internet Search Data: Term Selection

It's easy to be persuaded by the theoretical benefits of using internet search data as a potential monitor of economic behaviour. Compared to traditional survey-based sampling, it is quicker, timelier and cheaper. It is arguably a "purer" form of monitoring behaviour: aforementioned surveys tend to be based on questionnaires and are reliant on the coercion of a response, but data entered into a search engine provide a more honest appraisal of consumer preferences and behaviour; there is no bargaining or strategic game being played. The digital traces left by households offer a true reflection of their intentions.

Taking this further, there is a suggestion that the constituents of the so-called "big data" revolution, such as internet search and similar sources of electronic information, have such potential that the old-sampling techniques could, in the future, be rendered obsolete. Given the perabytes of data to be made available, maybe there will be no need for scientific theory and associated modelling.

This rather apocalyptic view is, of course, somewhat extreme. Many of the lessons from decades of acquired knowledge on statistics need to be kept in mind when working with the new data sources. An example is sampling bias. Although gaps are narrowing, internet usage still tends to be selective, linked to factors such as age, with younger people more likely to go online than older people.

Moreover, extracting a meaningful signal from such vast datasets could prove problematic. This is a key challenge faced with using search data and the selection of search terms. There are literally hundreds of millions of terms that users can input into the search box. The potential for excessive noise in the dataset is obvious. In the following sub-sections, the various methods and a proposed approach to term selection are outlined.

4.3.1 The View from the Literature

Concentrating on changes in unemployment, which is the primary focus in this chapter, there are several examples within the emerging literature on internet search data on term selection.

Choi and Varian (2009b) begin their selection process by asking "What would you search for if you thought you might lose your job?", and suggest terms such as "vacancies" or "jobs" may be relevant. As Google Trends places many of these terms into pre-arranged categories such as "Jobs and Education", the user can immediately download groups of these and similar terms for analysis. Suhoy (2009) takes a similar "off-the-shelf" approach (and with some success) when predicting changes in unemployment (and other variables) in Israel.

In contrast, Askitas and Zimmermann (2009) adopt a more minimalist tactic, using a restricted choice of keyword sets based on theoretical grounds of economic intuition such as connections with unemployment offices, possible reactions to high skilled workers to the fear of losing jobs or popular job-site searches. This approach therefore addresses the task of nowcasting joblessness from two angles: the flow out of unemployment and the flow into unemployment.

Alternatively Koop and Onorante (2014) begin the selection process by using what is essentially a "root" term which is the target variable (i.e. "unemployment"). Having downloaded the corresponding search volume, similar data can be sourced for related terms which the Google Trends interface provides (these are the most popular related terms to the root). This process continues with any obviously unrelated and repeated terms deleted. The resulting data are subsequently aggregated into a composite Google index.

4.3.2 Search Terms and UK Unemployment

Term selection in this chapter is a hybrid of the aforementioned approaches. Using a carefully considered "root" term, related terms are also downloaded and then aggregated to form a single composite Google index for unemployment.

As per Choi and Varian (2009a) the process begins by considering what someone would search for if they perceived that they had just been, or were about to be, made unemployed. It seems sensible to suggest that initial searches are likely to be focused on two areas: available benefits to the unemployed and vacancies in a particular field that the worker specialises in.

From initial data exploration, focusing on vacancies (or the flow out of unemployment) proved to be a considerable challenge: for example, job openings across (and within) the public, construction, manufacturing and service sectors seem to require disparate searching terms.

Another concern with using terms related to specific job searching activity is the difficulty in separating those related to joblessness and those related to "on-the-job" activities i.e. searching for a new position while already actually in a job. The latter, one would assume, is pro-cyclical in nature i.e. increased searching "on-the-job" is likely to be influenced by a positive economic environment. This could counteract searching related to actual joblessness.

Following careful consideration, it was felt that greater success may be found in focusing on searches related to the flow into unemployment. This seems to be the approach taken by McLaren and Shanbhogue (2011) through their decision to use a single search term "JSA" (which is an acronym for Job Seekers' Allowance, a form of benefit paid in the UK to people who are unemployed and actively seeking work).

However, there are a number of immediate concerns. JSA is subject to eligibility criteria, may not be claimed by all those unemployed and may therefore not fully capture developments in the labour market. Moreover, using a single-term to predict changes in unemployment seems dangerous and subject to a relationship breakdown with a target dependent variable, especially given the term is specific and vulnerable to change. For instance, the UK government has recently proposed to merge JSA with other benefits, such as those related to housing, to create a system known as universal credit. The term JSA may in time become obsolete.

The root search term that was chosen to reflect the flow into unemployment was therefore "redundancy". It was felt that this was a little more generic in nature and less susceptible to the aforementioned issues of the term JSA.

4.3.3 Creating a Search-Term Unemployment Indicator

The "redundancy" term was placed into the Google Trends interface and the corresponding search volume data downloaded. Google Trends also provides data on related terms (these are inputs most commonly entered before and after the original term). Volumes for these related terms were also subsequently downloaded, and the process repeated for each of these until a natural end arrived. Duplications were deleted and terms that were obviously unrelated to either the "root" or the target variable (unemployment) were also removed (which inevitably required an element of judgement).

Following this process of reduction, redundancy and 20 associated (either directly or indirectly) search variables remained, a list of which is provided in the appendix.

On inspection of the data, however, there was a concern over the underlying quality of some of these search terms. Many had lengthy strings of zeros at the beginning of their respective time series (which begin in January 2004).

The presence of these zero readings primarily reflects the way that Google Trends provides the numbers. The time series data represent how the number of searches for a particular term relative to the total number of Google searches changes over time. Google then normalise the data and presents them on a scale of 0-100. Zeroes are used when Google have insufficient data for that term (the term's search popularity is so low relative to the total number of searches the derived number is no greater than zero). In contrast, 100 represents the high watermark for that particular search term i.e. the point in time when it was at its most popular in a relative sense.

This leads to a number of issues to be wary of:

- As the data don't actually represent absolute search volumes, rather relative popularity on Google, declines in any downloaded time series on a week-to-week basis may not actually represent a fall in the raw number of searches but lower than average changes in volume.
- The data are also based on sub-samples, so back history can change, inducing excessive instability for less popular searches.
- If there is a desire to transform respective time series using log or per-cent deviations then the presence of (or the potential for) zeroes within time series adds a layer of complexity.

The third point subsequently led to some reluctance in doing any manipulations to the dataset such as taking differences or making transformations. With the lead also being taken from the existing literature on search data, individual Google search terms were subsequently used "as provided".

How the 21 search terms were used in forecasting equations was, however, for additional consideration.

For instance, there seemed increasing instability in the data series outside of the main root "redundancy" search term. As noted above, some data series have a number of zeroes within the series. Presumably this reflects in part the increasingly thin volumes of search activity (i.e. lower popularity for more nuanced search terms). The feeling was at the individual level these data series would struggle to provide reasonable forecasting ability, especially when the data series are truncated for sample.

With processing ease in mind, and with the presence of excessively low search volumes for a number of individual search terms, the creation of 21 separate individual forecasting regressions was somewhat undesirable. Instead, some kind of composite Google index which captured the information contained within the dataset seemed a cleaner and easier way forward.

Therefore a dynamic weighting system was created, which changed on a weekly basis i.e. with the addition of each weekly data point. This involved taking a sum of the search readings for p Google search terms (GST) for a particular week, and then creating individual weights for each search term based on their own individual contributions for that week (as per the formula below):

$$W_{i,t} = \frac{GST_{i,t}}{\sum_{i=0}^{p} GST_{i,t}} \tag{4.1}$$

This means those searches that have achieved a relatively greater increase in popularity attract a greater weight when creating the resulting composite Google Redundancy Index (GRI), which by definition is the sum of the weighted individual search term volumes for each week during the sample period.

As the Google data are provided on a weekly basis, for initial testing the series were converted into monthly time periods. This was achieved by breaking down the weekly search volumes into separate days (assigning the weekly value to each day) before aggregating these into months. For those weeks which overlap two calendar periods, the search volume is essentially weighted according to the number of days each month has in that week. For additional information, a numerical example is provided in the appendix.

Finally, visual inspection of the monthly series suggested that Google data exhibited characteristics of seasonality. This was removed by running the respective derived monthly time series through the X-12 ARIMA seasonal adjustment programme, which is freely available from the US Census Bureau.

4.4 Google and Survey Data: First Look

Regression models were created to assess the relationships between unemployment, Google search and survey data. Models are all based around the following specification:

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \gamma X_t + \varepsilon_t \tag{4.2}$$

The target variable Y_t is the three-month change in the ILO unemployment measure, which is line with McLaren and Shanbhogue (2011). It is therefore assumed that policymakers at the Bank of England attach some importance to this metric.

The explanatory variable, X_t, represents a single variable that is part of a larger dataset that contains a number of time series that may be useful in predicting changes in unemployment (all of which are assumed to exhibit the characteristic of stationarity). This includes the GRI and two other Google variables, namely search volumes for JSA and equivalent data for the GRI root term “redundancy”.

Within the dataset there are also eight variables drawn from surveys of business and consumers. These include data from the KPMG/REC Report on Jobs, which is based on survey information provided by recruitment agencies. Respondents answer questions covering topics such as the demand from business for staff, vacancies at firms and labour availability. The data are provided in the form of diffusion indices, in a similar vein to the closely-watched Purchasing Managers' Indices (PMI), where a reading of 50.0 represents no-change on the previous month.

Business employment expectations data are sourced from the Bank of England's Agents Survey, while there is also a measure of monthly changes in employment at businesses from the UK PMI surveys (covering manufacturing, services and construction sectors). The expectations of consumers regarding employment is provided by DG EcFin. More detailed descriptions of these variables are provided in the appendix.

As well as N single variable models, two small-scale factor models were created. As macro-economic time series tend to move in broadly similar fashions, this procedure essentially summarizes the information provided by a cohort of seemingly disparate indicators into a single variable. This can then be used directly in regression equations such as 4.2. Factor techniques have become a popular option for this kind of macro-economic modelling in recent years (see e.g. Stock and Watson (2011) for a discussion on factor methods).

The derived factors in this case were created from the first principal component from a static principal component analysis exercise similar to that used in the previous chapter. The first one (marked as F1 in the tables below) includes the eight survey-based variables, while the second factor (F2 in the tables below), includes the survey variables plus times series for the GRI and the Google "root" term, redundancy. This provides the platform to observe the marginal benefit (or cost) of adding Google search terms to the summarised information provided by the surveys.

Finally, a combination of the Akaike Information Criteria (AIC) and in-sample significance in a linear regression was used to determine an optimum number of lags of Y_t to use in regression equation 4.2. An autoregressive specification of order 2 was found to work well, and also subsequently offered as an additional model.

4.4.1 In-Sample Regression Statistics

In-sample regression statistics are provided in table 4.1. The sample period covers January 2004 to November 2014.

Generally, the various explanatory variables are statistically significant at the 5% level, with signs as expected. Note that several survey variables would be expected to have negative coefficients as these are designed to measure business demand or expectations for employee growth. This implies an inverse relationship with these indicators and unemployment should exist.

With specific reference to the Google indicators, the GRI performs extremely well and broadly on par with the survey variables in terms of coefficient statistical significance (p-value <1%) and R^2 performance. The Google "root" term variable also perfoms well, albeit not quite to the same degree as the GRI. This suggests that the additional Google information included in the GRI may add some value to that provided by the "root" term.

In contrast, the JSA Google term is not so good. The statistical significance of JSA in the regression equation is less than 10%, but negative. This implies a rise in searching for JSA by internet users is associated with a fall in unemployment, a counter-intuitive result.

Table 4.1: In Sample Model Regression Statistics: Jan-2004 to Nov-2014

	AR(2)	GoogJSA	GoogRed	GRI	RoJPP	RoJTB	RoJVC	RoJSA	PMIEM	ConExp	BoESR	BoEMN	F1	F2
α	1.12	8.17*	-46.26*	-57.25***	101.49***	109.83***	130.11***	-68.74***	225.19***	-20.25**	4.76	-1.42	4.99*	4.78
P-Value	0.70	0.09	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.15	0.63	0.09	0.10
St. Error	2.94	4.84	23.60	17.05	27.28	31.05	29.82	20.18	57.29	8.45	3.25	2.91	2.90	2.86
Y_{t-1}	1.12***	1.09***	1.09***	1.04***	1.02***	1.05***	0.98***	1.04***	1.01***	1.08***	1.08***	1.03***	1.00***	0.99***
P-Value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
St. Error	0.09	0.09	0.09	0.09	0.09	0.08	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
Y_{t-2}	-0.22**	-0.23**	-0.25***	-0.27***	-0.23**	-0.27***	-0.26***	-0.30***	-0.28***	-0.29***	-0.26***	-0.26***	-0.29***	-0.30***
P-Value	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
St. Error	0.09	0.09	0.09	0.08	0.08	0.08	0.08	0.09	0.08	0.09	0.09	0.08	0.08	0.08
X_t	n/a	-0.22*	1.09**	1.40***	-1.81***	-1.97***	-2.25***	1.45***	-4.39***	0.70**	-7.71**	-12.83***	20.61***	22.24***
P-Value	n/a	0.07	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00
St. Error	n/a	0.12	0.54	0.40	0.49	0.56	0.52	0.41	1.12	0.26	3.19	3.63	4.80	4.95
F-stat	335.47	228.84	230.42	246.98	250.37	247.61	261.15	247.34	253.80	236.93	234.09	247.93	260.24	263.85
Sig	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$AdjR^2$	0.837	0.840	0.841	0.850	0.852	0.851	0.857	0.850	0.854	0.845	0.843	0.851	0.857	0.858

Notes: The table reports the regression statistics of equation 4.2 using various Google and survey indicators (see appendix for a full description of the individual terms).

*10% significance; **5% significance; ***1% significance

4.4.2 Coefficient Stability

Although in-sample significance is found across all explanatory variables, it's also interesting to consider the stability of regression coefficients across rolling sub-samples. With this in mind, the overall sample is reduced to cover the period January 2004 to December 2006. Equation 4.2 is then continually re-estimated on a three-year rolling regression basis for each model until the end of the sample period i.e. the second regression estimation would be February 2004 to January 2007, the third March 2004 to February 2007 and so on and so forth.

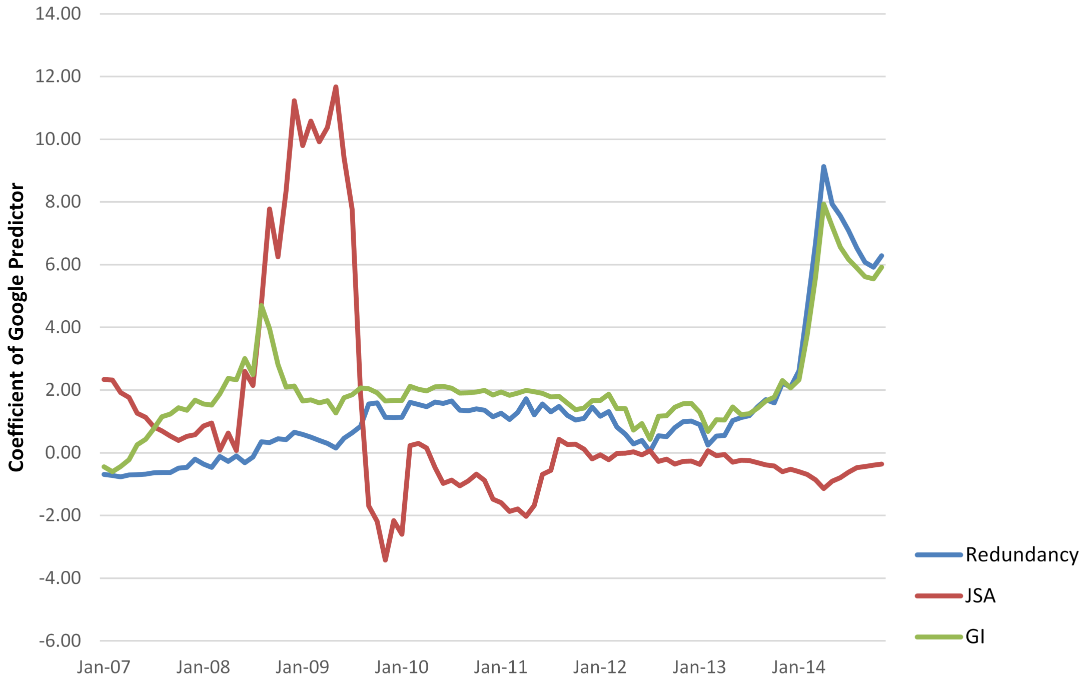
Throughout the process a note of the coefficient (θ) for the respective explanatory variable is taken, with the results plotted to build an insight into the stability of the relationship with the dependent variable over time. Ideally, the coefficients shouldn't differ too much over the sample period. Figures 4.1 to 4.3 show coefficients on Google indicators, selected survey indicators and the F2 factor model (note results were broadly similar for all survey and factor model coefficients).

Starting with Google indicators, the coefficient on JSA exhibits a large degree of instability during the period 2007 to 2010 and switches from a positive to a negative relationship over the sample period. As noted in section 4.3.1, this counter-intuitively implies that a rise in searching for JSA is associated with a fall in unemployment.

Turning to the two other Google indicators, with the exception of a brief period at the start of the sample period, these both maintain their expected signs. Moreover, there seems good stability in the respective coefficients up to 2014, with both moving in narrow ranges. However, through 2014 there is a considerable increase in coefficient values.

Moving on to the survey indicators, these typically move in narrow ranges up to 2012, and show expected signs. However, around 2012 there is a considerable

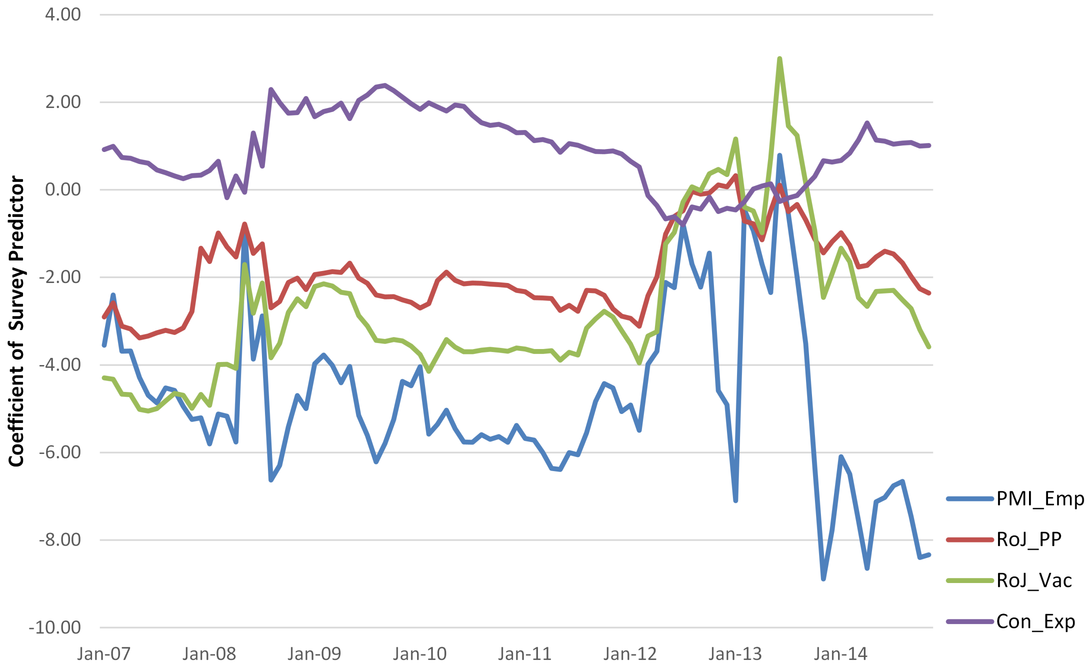
Figure 4.1: Time Varying Coefficients for Google Predictors



shift in the relationship between the target and survey variables with observed switches in the coefficient signs.

These observations are perhaps best captured in figure 4.3, which summarizes the coefficient stability on the F2 factor variable plus associated uncertainty around the coefficient readings (through 95 percent confidence bands). Note the considerable stability up to early 2012 in this coefficient, before a sharp deterioration and a noticeable widening of the 95 percent interval band. This instability lasts until early 2014 before a return to a range similar to that seen pre-2012 (there were similar findings in terms of confidence bands across the survey variables, but for brevity these are not reported).

Figure 4.2: Time Varying Coefficients for Selected Survey Predictors

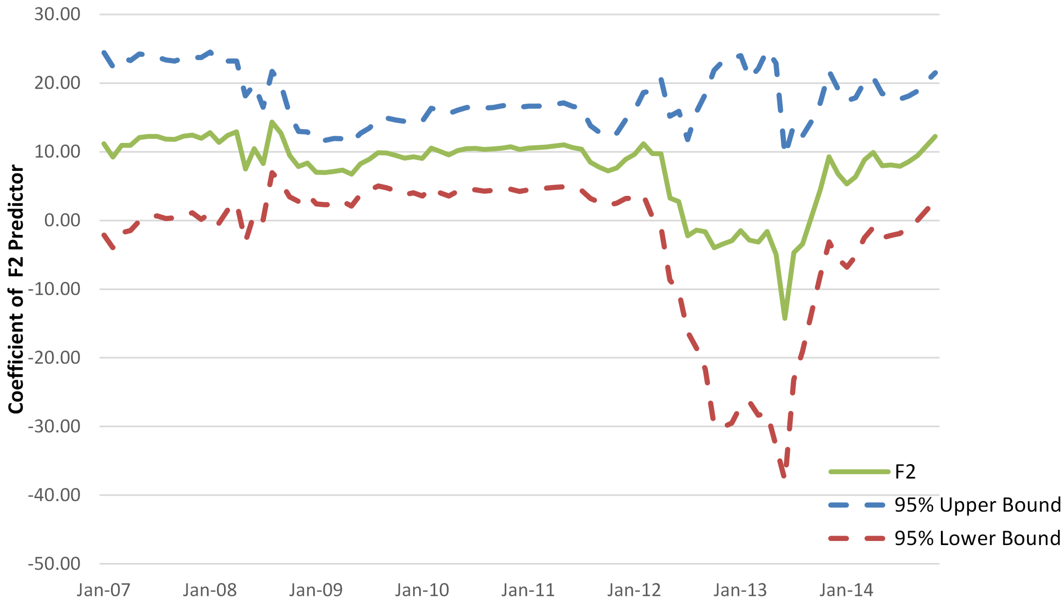


4.4.3 Out-of-Sample Performance

A primary usefulness of the various Google and survey-based data would be to help provide advance guidance of changes in unemployment. If forecasts are accurate then they could be used to gain insight of the likely movement in official data ahead of the actual release.

One-step ahead forecasting performances of the various models were subsequently compared against a benchmark, which is defined as a "no-change" forecast i.e. unemployment is projected to fall (rise) at the same rate as the previous observation. Note that the regressions are recalculated on an "expanding" window basis i.e. the first out-of-sample forecast (January 2009) is based on the estimated regression equation (4.2) for data covering January 2004 to December 2008, the second forecast (February 2009) on the sample January 2004 to January 2009 etc.

Figure 4.3: Time Varying Coefficient for F2 Predictor



Note for figures 4.1-4.3: Illustrations of the time-varying coefficients for the Google and selected survey indicators (expanding regression window). The lines represents the coefficients of the various explanatory indicators from the expanding regressions of equation 4.2 with the three-month change in the ILO unemployment rate used as the dependent variable.

Relative accuracy of the individual and factor models are assessed through the root mean square forecasting error (RMSFE), while a classic Diebold and Mariano (1995) test for comparative predictive accuracy is employed to assess the statistical significance of any improvement (or deterioration) in the RMSFE relative to the benchmark model.

Table 4.2 provides the results of the full out-of-sample (January 2009 to November 2014) simulations with the Diebold-Mariano (DM) and associated p-value statistics.

In general there are modest gains over the benchmark model (BM) in the one-step ahead forecasting comparisons, with the best performing models (GRI and

Table 4.2: In Sample Model Regression Statistics: Jan-2004 to Nov-2014

Model	BM	AR(2)	GoogJSA	GoogRed	GRI	RoJPP	RoJTB	RoJVC	RoJDS	PMIEM	ConExp	BoESR	BoEMN	F1	F2
RMSFE	38.5	37.4	38.0	36.7	35.6	36.2	36.2	35.5	36.3	35.7	37.3	38.5	36.9	35.7	35.3
RMSFE Ratio to BM	1.00	0.97	0.99	0.95	0.92	0.94	0.94	0.92	0.94	0.93	0.97	1.00	0.96	0.93	0.92
DM Statistic	n/a	0.99	0.53	1.60	1.71	1.19	1.08	1.57	1.26	1.56	0.87	0.00	0.74	1.41	1.61
P-Value	n/a	0.24	0.34	0.11	0.09	0.20	0.22	0.12	0.18	0.12	0.27	0.40	0.30	0.15	0.11

Notes: Each model is first estimated for the period up to December 2008 as per equation 4.2

The root mean-squared forecast error (RMSFE) for one month ahead forecasts are compared against actual unemployment outturns over the out-of-sample period for each individual model;

See Appendix A3 for further details on individual model terms

RoJVac) providing an 8% outperformance relative to the benchmark. Note again the relative weakness of the JSA-based model (GoogJSA).

Moreover, using a factor model based on the common signal provided by both Google and survey data is generally better than using individual models.

The DM statistics suggests that any "outperformance" over the benchmark is statistically insignificant.

But having previously noted instability between independent and target variables around 2012, the robustness of the full out-of-sample results is questionable.

So a split of the results into two broadly equal sub-samples (January 2009 to December 2011 and January 2012 to November 2014) was produced. These results are provided in tables 4.3 and 4.4.

Gains over the benchmark arrive primarily in the early years of the out-of-sample simulations with outperformance during 2009-2011 as high as 21% for the best performing model (RoJTB).

Other KPMG/REC Report on Jobs variables also performed well over this period such as RoJPP and RoJVac. Also note the strong performance of the F1 and F2 factor models. All of these one step-ahead forecast improvements over the benchmark model are statistically significant at the 5% level according to the DM test statistic.

Turning to the Google indicators, GRI offers a 12% gain over the benchmark, again a statistically significant result. However, note the disappointing performance of GoogJSA. While the general improvements in forecasting accuracy seen for individual models matches up with McLaren and Shanbhogue (2011), the poor performance of GoogJSA sits in contrast. Using a search term such as JSA is not a robust strategy.

Table 4.3: Out-of-Sample Nowcast Testing: Jan-2009 to Dec-2011

Model	BM	AR(2)	GoogJSA	GoogRed	GRI	RoJPP	RoJTB	RoJVC	RoJDS	PMIEM	ConExp	BoESR	BoEMN	F1	F2
RMSFE	40.5	37.7	39.3	37.5	35.5	33.7	32.1	33.1	35.7	35.8	38.5	37.5	39.0	33.9	34.1
RMSFE Ratio to BM	1.00	0.93	0.97	0.93	0.88	0.83	0.79	0.82	0.88	0.89	0.95	0.93	0.96	0.84	0.84
DM Statistic	n/a	2.08	0.89	1.95	2.04	2.46	2.85	2.77	1.93	1.68	0.82	1.02	0.40	2.24	2.08
P-Value	n/a	0.05	0.27	0.06	0.05	0.02	0.01	0.01	0.06	0.10	0.28	0.23	0.36	0.04	0.05

Notes: Each model is first estimated for the period up to December 2008 as per equation 4.2;

The root mean-squared forecast error (RMSFE) for one month ahead forecasts are compared against actual unemployment outturns over the out-of-sample period for each individual model;

See Appendix A3 for further details on individual model terms

Table 4.4: Out-of-Sample Nowcast Testing: Jan-2012 to Nov-2014

Model	BM	AR(2)	GoogJSA	GoogRed	GRI	RoJPP	RoJTB	RoJVC	RoJDS	PMIEM	ConExp	BoESR	BoEMN	F1	F2
RMSFE	36.4	37.1	36.6	35.7	35.7	38.5	40.0	37.8	36.9	35.5	35.9	39.5	34.7	37.4	36.4
RMSFE Ratio to BM	1.00	1.02	1.00	0.98	0.98	1.06	1.10	1.04	1.01	0.97	0.98	1.09	0.95	1.03	1.00
DM Statistic	n/a	-0.32	-0.10	0.37	0.30	-0.76	-1.25	-0.49	-0.20	0.39	0.32	-1.11	0.86	-0.36	0.01
P-Value	n/a	0.38	0.39	0.37	0.38	0.30	0.18	0.35	0.39	0.37	0.38	0.21	0.27	0.37	0.40

Notes: Each model is first estimated for the period up to December 2011 as per equation 4.2;

The root mean-squared forecast error (RMSFE) for one month ahead forecasts are compared against actual unemployment outturns over the out-of-sample period for each individual model;

See Appendix A3 for further details on individual model terms

From 2012 onwards, there has been a deterioration in the forecasting accuracy of many models when it comes to benchmark outperformance particularly RoJTB. This seems to reflect a strong improvement in the performance of the benchmark, which has proven hard to beat over the period 2012-2014.

Notwithstanding the stylized fact that in-sample results don't necessarily always translate into similar out-of-sample results, it's fairly easy to follow how the issues of instability identified in section 4.4 could spillover into actual real-world nowcasting performance.

Indeed these concerns were the motivation for the relatively informal use of sub-samples from the full out-of sample nowcasting test period of 2009-2014, which showed some variance in nowcasting capability of the indicator models.

Such time variant instability in the predictive content of models is a well-known phenomenon in macroeconomics.

With this in mind, in an extensive review of the literature, Pesaran and Weale (2013) discusses a number of tools available to better evaluate the predictive ability of models. These include the one-time reversal and fluctuation tests proposed by Giacomini and Rossi (2010) which can be useful to identify a one-time break in forecasting ability or to reveal which models are performing better (or worse) than competitors at various points in time.

While such formal tests are left for future work, note that Pesaran and Weale (2013) point to empirical evidence that forecast combination can sometimes provide the best natural protection against model instabilities arising from e.g. misspecification.

This evidence sits well with much of the research presented throughout this thesis where the most consistent results have tended to be associated with the pooling of individual nowcasts.

4.5 A Weekly Model for Predicting Unemployment

With some careful term selection, Google indicators offer similar characteristics to monthly survey data when predicting changes in unemployment.

However, to this point, the assessment of the usefulness of the Google data has been somewhat unrealistic.

As highlighted in chapter 2, consideration of the true value of an explanatory variable used to form a forecast of some dependent variable should also include an assessment of that variable's timeliness relative to other predictors and how the data are to be used in practice. What is the release schedule of all explanatory variables? Does an earlier release of an individual explanatory variable provide information that helps reduce some uncertainty around a prediction that is subsequently confirmed (or perhaps built upon) by later released data?

When thinking about the prediction of UK unemployment, Google data are a) available weekly and b) generally provided to quicker timescales than surveys. Could these features mean that the information provided by Google, being the first released, offer a steer on changes in unemployment that is then confirmed by later-released survey data (and potentially bolstering the value of the former)?

To investigate, a pseudo "nowcasting" framework is created. Within this system, the underlying release structure of Google, survey and official data sources is mimicked. The rationale is to provide insight into how these interact with the uncertainty that surrounds a nowcast of UK unemployment during and after the month that is being nowcast.

Before explaining the framework in greater depth, first a reminder on what is meant by nowcasting.

Nowcasting is essentially concerned with the process of making some prediction of the present, the very near future or the very recent past via the linking of some macro-economic variable (usually GDP) to the flow through time of information being produced by a dataset of heterogeneous variables.

That is, due to an asynchronous release schedule for the explanatory variables, nowcasts for a particular reference period tend to be updated in line with releases of new data (or information).

To make this clear, consider how information flows if one wished to nowcast unemployment.

Following the close of a calendar month, there is typically a lag of six-to-seven weeks until the release of official data pertaining to the target unemployment metric. During the reference period (i.e. the nowcast month of interest), and then between the end of the reference period and the official data release, information on the performance on how the labour market performed can be gleaned from surveys and other sources of data (such as that offered by Google Trends).

As is typical with this type of application, the nowcast tends to be updated as new information arrives. With Google data available weekly and also to quicker timescales than monthly surveys, this advantage could help to improve a nowcasting model of unemployment in a timely fashion; predictions can be updated on a weekly basis both during and after the close of the reference period. This is the focus for the rest of the chapter.

4.5.1 Nowcasting Methods: A Very Brief Overview

Updating an unemployment nowcasting model on a weekly basis throws up a couple of challenges from an econometric perspective.

Firstly, the dataset's "jagged edge" has to be accounted for. This refers to the fact that a set of explanatory variables can have different release dates. The result can be a number of missing observations at the end of a sample when producing a nowcast.

A second challenge is the time frequency mismatch between dependent and explanatory variables. Unemployment data is available on a monthly basis, yet Google trends data are available weekly.

Various proposals exist within the nowcasting space on how to deal with these two challenges, and several surveys have emerged recently to offer a summary and extensive details of the various approaches. Examples include Bańbura et al. (2013), Camacho, Perez-Quiros, and Poncela (2013), and Forini and Marcellino (2013).

Whilst readers with a desire for greater detail are invited to look to these survey papers, very briefly the literature points to three competing methods as the leading techniques to deal with the two primary nowcasting challenges.

Perhaps the most popular is the relatively simple bridging equation, which explicitly deals with the mixed-time frequency issue by taking an average of the higher-frequency observations and linking the results to the lower frequency variable via a linear regression. When missing observations exist due to the "jagged" edge then these can be predicted through the use of e.g. autoregressive forecasting techniques. An example of such an approach was of course found in chapter 3.

However, a well-known criticism of the bridging equation framework is that the averaging of high frequency data is performed with the potential loss of information from individual innovations.

Giannone, Reichlin, and Small (2008) go some way to addressing this through state space solutions that use factor extraction to summarise the data held within the explanatory variables. Rather than averaging observations, the factors can be used at their higher frequency. Such techniques also make use of Kalman filtering, which can easily deal with missing observations and ensure that the dual nowcasting challenges are addressed within a single model framework.

An alternative approach is that of the MIDAS regression. While a fuller description is provided below, the features of an explicit treatment of the mixed-time frequency problem with no need to forecast missing observations made this technique a favoured choice. Further confidence in the viability of MIDAS as a nowcasting tool is provided by Bai, Ghysels, and Wright (2013). They show that MIDAS style regressions offer very similar forecasting results to the methods advocated by Giannone, while at the same time offering a much less computationally demanding estimation procedure than those based on Kalman filtering applications.

On these grounds, a MIDAS model seems an excellent choice for producing weekly unemployment nowcasts.

4.5.2 MIDAS Regression

Unemployment data are available monthly, whereas the Google predictors are provided on a weekly basis. MIDAS methods provide an easy and intuitive way to deal with such a situation.

Reconsider the regression specification outlined earlier, which linked estimated values of the dependent variable to p lags of itself and some explanatory variable.

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \gamma X_t + \varepsilon_t \tag{4.3}$$

In this instance, both variables are implicitly in the same time domain. But consider when this isn't the case. Perhaps the explanatory data are released weekly or daily, whereas information for the dependent variable are only available monthly. More formally, let us say there are m readings of the high frequency explanatory variable X_t^{HF} observed between each release of the dependent variable. An arithmetic average of the high frequency observations X_t^{HF} can be used to create X_t:

$$X_t = \frac{1}{m} \sum_{k=1}^{m} L_{HF}^{k} X_t^{HF} \tag{4.4}$$

The use of such a transformation equation is commonly seen in bridging equation methods. It is a parsimonious, popular and easy to implement framework which deals with the challenge of mixed time frequencies.

But given the potential for information loss from using an arithmetic mean, an alternative solution would be to include, on the right-hand side of equation 4.2, the explanatory variables at their 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 $m = 22$ if one assumes 22 trading days per calendar month).

To address some of the criticisms of the bridging equation approach related to information loss, but avoiding excessive parameter proliferation, some middle ground can be provided by using a regression in the following form:

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \gamma \sum_{j=1}^{p} \phi(k : \theta) L_{HF}^{k} X_{t-h}^{HF} + \varepsilon_t \tag{4.5}$$

where the regression co-efficient γ links the target low frequency indicator to a weighted sum of the indicator variable observations over the specified quarter. The function $\phi(k : \theta)$ is a polynomial that determines the weights used for the temporal aggregation of the high frequency indicator.

The weighting functions used in MIDAS regressions can take on many functional forms and various specifications have been considered, including non-linear versions. The form could be dependent on the user's own preferences such as placing greater weight on the more recent values. Ghysels, Sinko, and Valkanov (2007) discuss these, but note many of the non-linear forms have been focused on financial applications where m is generally large. Faced with a relatively parsimonious case ($m = 4$) the nowcasting model in this chapter works with a form of MIDAS where the weights are estimated without restriction, as per Marcellino and Schumacher (2007).

Another noteworthy feature of the MIDAS regression is there is no need for the extrapolation of missing observations to deal with the dataset's ragged edge: re-balancing is essentially achieved by shifting the time series of respective explanatory variables forward (or backwards) via the use of different integers for h, which reflects the difference between the forecast target period and the most recent observation of the indicator. When nowcasting a quarterly statistic such as GDP this provides the opportunity for within quarter estimations of the target variable by exploiting the timelier information provided by the high frequency indicators.

There are several examples of using MIDAS techniques to forecast low frequency macroeconomic times series such as GDP. For the interested reader, Armesto, Engemann, and Owyang (2010) provide an intuitive and easy-to-follow introduction to the topic. See also Kuzin, Marcellino, and Schumacher (2011) and Kuzin, Marcellino, and Schumacher (2013) for MIDAS nowcasting the GDP of the eurozone and various industrialised countries. Recently Allan et al. (2014) employ a MIDAS model to nowcast GVA for Scotland.

4.5.3 Weekly Google Data

Before turning to the systematic application of the MIDAS model when nowcasting unemployment, first a note on the practical problem when mapping weekly and monthly observations. When using weekly data, a problem is an inconsistent number of weeks each calendar month.

An adaption of the approach proposed by Hamilton and Wu (2014) to generate a balanced weekly data set where each month consists of four "weeks" is subsequently used.

This process involves generating the first week of data for a month by calculating a number running from the 1st to the 7th of the month.

Week 2 then subsequently covers the 8th-14th and week 3 covers the period 15th-21st. Week 4 is generally a little longer as it always runs from the 22nd to the final calendar day (e.g. usually 30th or 31st).

Moreover, as Google define a week as running Sunday through Saturday a secondary issue is the first day of the month doesn't necessarily occur on a Sunday. Therefore, as per the monthly data transformation in section 4.3, the weekly numbers were assumed to be the same across all days of the week to create a daily series covering the full sample period.

The four weekly readings for each calendar month are created by taking respective averages of the newly created daily series. For further clarity, a numerical example is provided in the appendix.

4.5.4 Nowcasting Timelines

As the Google data are available throughout the reference calendar month, which is referred to as month M0, the nowcasting exercise begins at the end of the first week of that month. Assuming that the data are available immediately without delay from the Google Trends search engine, the first "prediction" occurs on the 8th day of month M0, with updates then provided on the 15th and 22nd. The final "weekly" update that refers to the reference period is provided on the 1st day of the next calendar month (M1).

This process continues through month M1, which is where the exercise ends and leaves around a two-to-three week gap until the release of the UK labour market statistics for the reference period (which typically occurs around week 3 of month M2).

Of course throughout the process, which consists of 8 separate nowcasts, monthly data from the business surveys and official labour statistics also become available. Survey data for M0 is typically available by the end of the first week of month M1 offering an opportunity to hopefully confirm (or enhance) the signal provided by the earlier available Google data.

Moreover, there are also updates to labour market statistics throughout the forecasting period related to M-2 and M-1 (i.e. months immediately preceding M0). These updates will need to be incorporated into the nowcasting framework due to impacts on benchmark estimates and respective nowcasts. Note that these are essentially unavailable for the first six weeks of the nowcast formulation. The regression equations are adapted accordingly to reflect this (e.g. coefficients β_1 and β_2 can be set to zero during weeks 1 and 2 and then β_1 is set to zero for nowcasts running through weeks 3-6).

A diagrammatic example (figure 4.4) provides further understanding of the timelines and dataset structures typically faced when conducting the respective nowcasts.

4.5.5 Model Averaging

The nowcasting regressions developed are single-variable approaches: the dependent variable is linked to just one explanatory variable. With a mixture of Google and survey predictors, the nowcast framework needs to be extended.

Rather than including all of the explanatory variables in a single regression equation, there are two approaches common in the nowcasting literature. One is a factor-based approach, another is to take a bunch of individual models and use some kind of average. That is with N explanatory variables, an equivalent sized set of nowcasts will be available. This chapter takes the averaging approach, supported in part by evidence showing forecast pooling can yield good predictive performance. See for example Timmermann (2006).

Individual weights, ω_i, are applied at time t for each model $Y_{i,t}$ where $i = 1, ..., N$, with two weighting schemes considered:

Firstly, equal-weights:

$$\omega_{i,t} = \frac{1}{N} \sum_{i=1}^{n} Y_{i,t} \tag{4.6}$$

Secondly, weights based on a Mean Square Forecasting Error (MSFE):

$$\omega_{i,t} = \frac{MSFE_{i,t}^{-1}}{\sum_{t=1}^{n} MSFE_{i,t}^{-1}} \tag{4.7}$$

Under this specification, those models that have historically provided the better forecasting performance secure a bigger weight in respective nowcasts.

4.6 Weekly Nowcasting Results

Two metrics are used to assess the performance of the model over the eight-week nowcasting period. Individual weeks are highlighted in tables 4.5 and 4.6 as W1, W2, W3, W4, W5, W6, W7 and W8.

RMSFE statistics are again provided to assess the quality of the point nowcast, with the lower values in table 4.5 suggestive of greater accuracy in predicting changes in unemployment. Comparison of individual model performance and the two different weighting schemes is made against a no-change nowcast. This "naive" benchmark nowcast is equal to the most recently observed rate of change in unemployment. The ratios in the table are designed to show more easily whether the benchmark has been beaten; taking the RMSFE of an individual model as a numerator and the benchmark model RMSFE as a denominator means a reading below one is viewed as positive.

While point estimates continue to dominate the professional forecasting field see e.g. consensus views around important economic releases or the HM Treasury monthly poll of independent forecasters, it is worth recognising that such estimates provide no indication of the inherent uncertainty associated with producing them.

To provide some idea of this uncertainty, the sums of the log predictive likelihoods is provided in table 4.6. These help to provide an assessment of the predictive distribution (commonly referred to as a density forecast) associated at each point a respective nowcast is made. Note a higher number implies a better nowcasting performance.

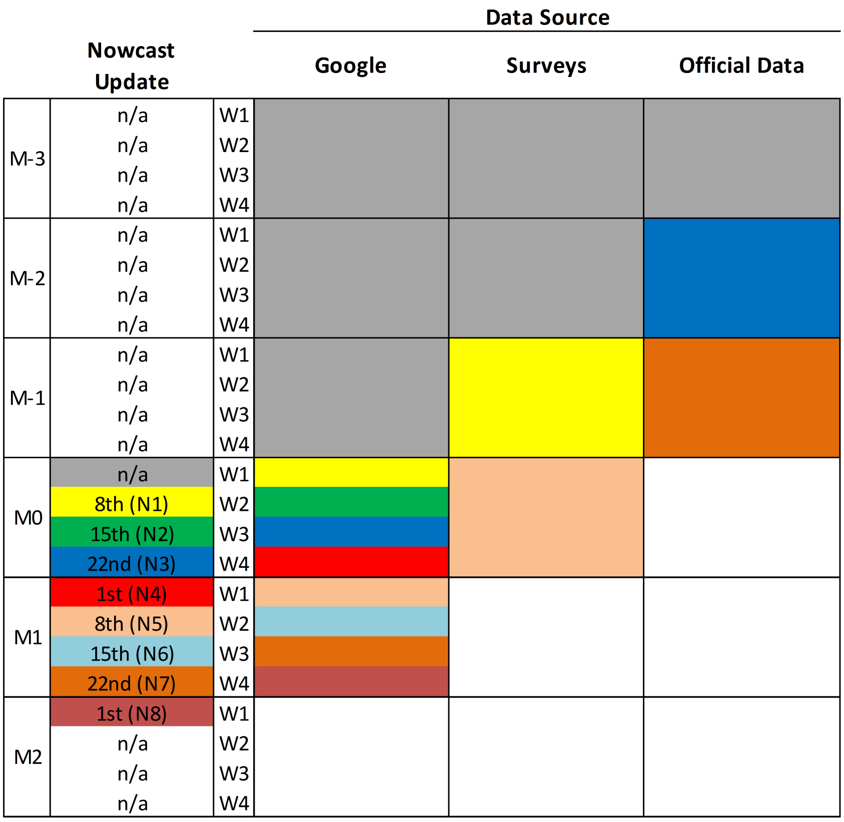
Moreover, to better illustrate any differences between GoogRed and GRI against other models, figure 4.5 shows the evolution of the sum of these log predictive likelihoods through the eight-week nowcasting cycle.

Table 4.5: Sums of Logs Predictive Likelihoods for Weekly Nowcasts

	W1	W2	W3	W4	W5	W6	W7	W8
MSFE Weights	-399.21	-399.15	-397.15	-396.53	-385.09	-385.01	-353.71	-353.49
Equal Weights	-403.20	-403.21	-398.99	-398.72	-386.38	-386.27	-354.24	-354.00
GoogRed	-406.63	-406.77	-390.92	-387.71	-389.51	-389.13	-355.80	-354.66
GRI	-415.43	-416.04	-390.23	-388.88	-390.16	-389.86	-354.95	-353.85
RoJPP	-407.76	-407.76	-401.31	-401.31	-385.96	-385.96	-355.29	-355.29
RoJTB	-410.71	-410.71	-401.98	-401.98	-388.17	-388.17	-355.31	-355.31
RoJVC	-397.80	-397.80	-398.81	-398.81	-383.83	-383.83	-354.39	-354.39
RoJDS	-406.10	-406.10	-404.78	-404.78	-389.00	-389.00	-356.40	-356.40
PMIEM	-397.54	-397.54	-400.74	-400.74	-386.15	-386.15	-354.94	-354.94
ConExp	-413.99	-413.99	-408.78	-408.78	-393.52	-393.52	-358.68	-358.68
BoESR	-424.85	-424.85	-411.21	-411.21	-397.34	-397.34	-361.96	-361.96
BoEMN	-412.32	-412.32	-404.84	-404.84	-391.71	-391.71	-358.05	-358.05

Notes: The table shows the sums of the log predictive likelihoods for each individual model. These are designed to provide an assessment of the predictive distribution (density forecast) associated with each point during the nowcasting cycle. Note a higher number implies a better nowcasting performance;
See Appendix A3 for further details on individual model terms

Figure 4.4: Typical Nowcast Time and Data Availabilities



Notes: The colours in the chart highlight the changing structure of the dataset as the nowcast of the three-month on three-month change in unemployment is updated on a weekly basis. The first nowcast (N1), which is conducted on the 8th day of the reference month (M0), indicates that the first week of Google data relating to M0 are now available and can be used (see yellow fill under the "Google" column). The data are also freshly available for the surveys for the preceding month (M-1), highlighted by the yellow colour fill covering weeks 1-4 (W1,W2, W3 and W4) in the column named "Surveys". Nowcast two (N2), to be typically made on the 15th day of M0, shows that the second week of Google data relating to the reference period has become available, but there are no updates to surveys of official data during this week (highlighted by just one green fill in the Google column). Nowcast three (N3), typically made on the 22nd day of the month, indicates that the third week of Google data for M0 plus official data relating to month M-2 are now available to update the nowcast. This is highlighted by the use of the blue box fills in the chart. The update process for nowcasts four through eight can easily be seen by following similar logic to that described for nowcasts one through three.

Figure 4.5: Sums of Log Predictive Likelihoods for Weekly Nowcasts

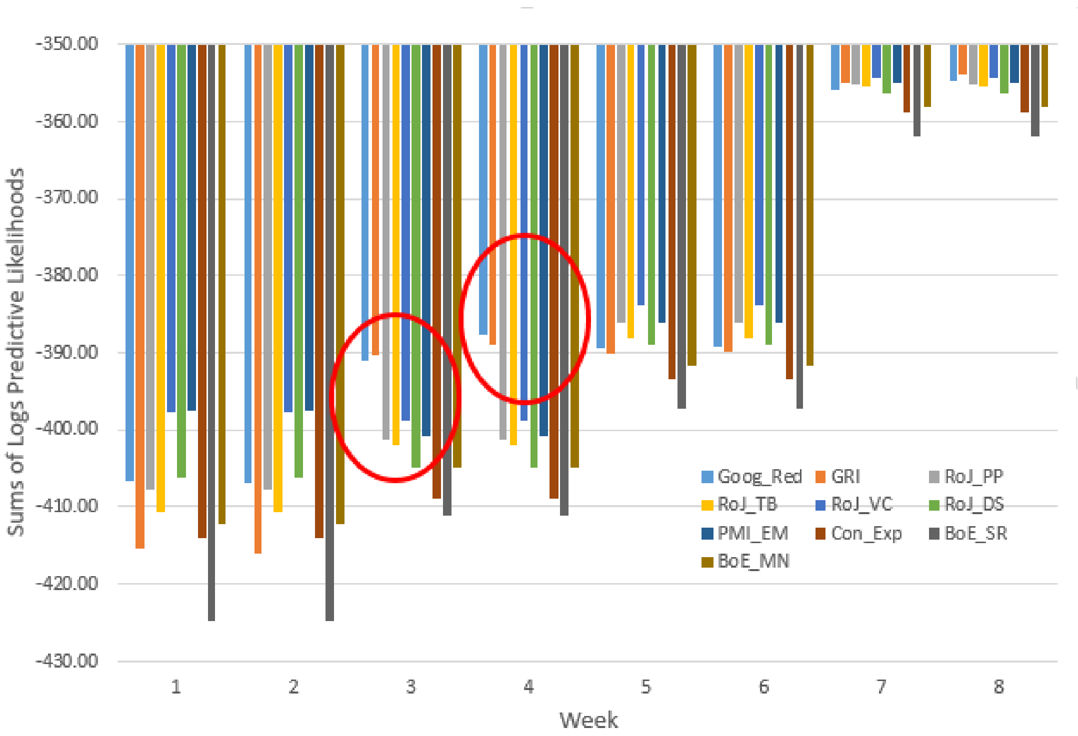


Table 4.6: RMSFEs for Weekly Nowcasts

	W1		W2		W3		W4		W5		W6		W7		W8	
Model	RMSFE	Ratio	RMSFE	Ratio	RMSFE	Ratio	RMSFE	Ratio	RMSFE	Ratio	RMSFE	Ratio	RMSFE	Ratio	RMSFE	Ratio
Benchmark	77.79	1.00	77.79	1.00	60.73	1.00	60.73	1.00	60.73	1.00	60.73	1.00	38.53	1.00	38.53	1.00
MSFE Weights	62.91	0.81	62.93	0.81	60.32	0.99	59.79	0.98	52.68	0.87	52.65	0.87	35.16	0.91	35.06	0.91
Equal Weights	65.73	0.84	65.80	0.85	61.51	1.01	61.18	1.01	53.42	0.88	53.37	0.88	35.38	0.92	35.28	0.92
GoogRed	69.16	0.89	69.40	0.89	56.56	0.93	54.66	0.90	55.16	0.91	54.98	0.91	36.06	0.94	35.58	0.92
GRI	78.33	1.01	78.87	1.01	57.29	0.94	55.65	0.92	56.38	0.93	56.24	0.93	35.73	0.93	35.25	0.91
RoJPP	71.65	0.92	71.65	0.92	63.99	1.05	63.99	1.05	54.05	0.89	54.05	0.89	36.16	0.94	36.16	0.94
RoJTB	74.06	0.95	74.06	0.95	65.20	1.07	65.20	1.07	55.79	0.92	55.79	0.92	36.21	0.94	36.21	0.94
RoJVC	62.33	0.80	62.33	0.80	61.60	1.01	61.60	1.01	52.04	0.86	52.04	0.86	35.48	0.92	35.48	0.92
RoJDS	68.04	0.87	68.04	0.87	65.93	1.09	65.93	1.09	55.34	0.91	55.34	0.91	36.30	0.94	36.30	0.94
PMIEM	62.34	0.80	62.34	0.80	62.93	1.04	62.93	1.04	53.53	0.88	53.53	0.88	35.67	0.93	35.67	0.93
ConExp	74.83	0.96	74.83	0.96	68.93	1.13	68.93	1.13	58.34	0.96	58.34	0.96	37.25	0.97	37.25	0.97
BoESR	81.73	1.05	81.73	1.05	69.47	1.14	69.47	1.14	59.86	0.99	59.86	0.99	38.54	1.00	38.54	1.00
BoEMN	71.23	0.92	71.23	0.92	65.07	1.07	65.07	1.07	56.19	0.93	56.19	0.93	36.94	0.96	36.94	0.96

Notes: The table shows the Root Mean Square Forecasting Error (RMSFE) of the point nowcast for the various models over the sample period. The statistic is based on the errors of these models in predicting a change in unemployment. The benchmark model is a no-change nowcast (the assumption is that the most recently observed rate of change in unemployment is carried over to the next period); The ratios show whether the benchmark model has been beaten by respective individual models by comparing respective RMSFEs. Any reading below one points to outperformance of the benchmark; See Appendix A3 for further details on individual model terms

While there is a rich literature on the functional form that the predictive density forecast should take (see section 1.5 for a discussion), in this instance it is assumed that past forecasting errors are normally distributed and unbiased. The point forecast is therefore the predictive mean with the predictive variance equal to that of the errors made in past forecasts.

The nowcasts are based around the MIDAS regression outlined in equation 4.5. Based on earlier findings, in its general form, the regression makes use of two lags of the dependent variable and a single explanatory variable, be it one of two Google or eight survey predictors used in section 4.3 (the JSA model is removed from the analysis).

This means there are 12 models to assess if the two weighted nowcasts are included. As highlighted in section 4.4.4, the way the data flows over the eight-week nowcasting period means for the first two updates (W1 and W2) dependent variable data relating to the first two lags utilised in the regression model are not available. As these can not be included there is subsequently sole reliance on one of the Google or survey predictors to make a nowcast. For weeks W3 to W6 data for the first lag is still unavailable and the regression equation is adapted accordingly.

For the survey variables, during weeks W1-W4, which can be referred to as the current month nowcasts (i.e. during the actual month that the nowcast is related to), the most recent values are used (i.e. data relating to month M-1). This changes with the release of reference month data for surveys and are subsequently used in the nowcasts during the following month (W5-W8).

Evaluation is over a six-year period spanning January 2009 to November 2014. The data outlined in tables 4.5 and 4.6 lead to the following observations.

Starting at a very general level, the RSMFE and log-predictive likelihood metrics both indicate that the benchmark model can be beaten with the help of explanatory data provided by the Google and survey indicators.

Moreover, the data indicate that model averaging helps improve nowcast predictive accuracy. Both predictive analysis metrics indicate that such approaches tend to yield more positive results than individual model specifications. This is in tune with much of the nowcasting (and forecasting) literature. The data also suggest that using a MSFE weighting scheme could be better than deploying equal-weights, although the differences are negligible.

Turning attention to the performance of particular variables, there is a great deal of importance attached to the release of the hard dependent variable data (which is used as lags in the regression equations). These data enter the regression equations in weeks 3 and 7, with their arrival associated with sharp drops in RMSFEs and noticeable improvements in the sum of the predictive log likelihoods across the vast majority of models.

The Google-based models (GoogRed and GRI), whose performance is of primary interest here, operate to a similar standard to the survey-based models in the earliest nowcast periods (weeks 1 and 2). They show no sign of providing any useful information over and above that provided by the survey data (which recall at this stage of the nowcast evolution is based on data referring to month M-1, rather than the reference month). Note that the KPMG/REC Report on Jobs Vacancies Index (RoJVC) and the Markit/CIPS UK Employment Index (PMIEM) are the best individual predictors in the first two nowcast weeks. Perhaps the strong performance of RoJVC is unsurprising as one would intuitively expect vacancies data to offer some leading properties over changes in unemployment.

The situation changes somewhat during weeks 3 and 4 as GoogRed and GRI show a distinctly better performance than survey-based models. Moreover, GoogRed and GRI are the only two models to beat the benchmark in this period. These also show distinctly higher log predictive likelihood readings.

The data suggest that the double hit of Google and official data in week 3 is a pivotal moment in reducing nowcast uncertainty, which is consolidated with the release of the final week of Google data for the reference month (week 4).

The signals provided by the Google-based models are broadly confirmed by the survey data during week 5 in the sense that RMSFEs and predictive densities are now similar to those provided by GoogRed and GRI for weeks 3 and 4. These also show distinctly higher log predictive likelihood readings (and is highlighted in figure 4.5 by the two red circles over the log predictive likelihoods for weeks 3 and 4).

This offers a sense that the Google data have, due to their timeliness, benefits over and above the surveys in predicting changes in unemployment.

However, later-released Google data (i.e. for the following month) provide little additional predictive benefit, with RMSFEs and predictive densities showing little change over weeks W5-W8. They still nonetheless perform comparably well to the survey-based models over this period. Indeed, the GRI model is the strongest performer (albeit just) in week 8 and GoogRed is joint second with RoJVC, just ahead of PMIEmp.

Finally, the charts outlined in figure 4.6 showcase how the nowcast for the favoured MSFE-weighted model evolves over the eight-week period. Using a symmetric density forecast based on the normal distribution also offers the opportunity to provide interval bands around the point nowcasts. Bands equal to one standard deviation are subsequently provided, which show the performance evolution of the MSFE-weighted nowcast over the eight weeks that updated nowcasts are produced.

Nowcast performance is naturally weaker and uncertainty much greater during the earlier nowcast periods, the latter signalled by relatively wide intervals around the point nowcasts. Performance improves greatly and the model uncertainty diminishes markedly as information from Google, surveys and official data accumulates. By weeks 7 and 8 the performance of the model is highly satisfactory. Actual outturns are tracked well, albeit with one criticism perhaps that turning points are a little slow to be picked-up. This probably reflects the reliance of auto-regressive components in the regression equation.

Another takeaway is that uncertainty is generally greater around the point nowcasts in the second-half of the sample period, as indicated by the wider interval margins over 2012-14 compared to 2009-11. This is in line with the findings of section 4.4, suggesting that unemployment nowcasting has being harder during the last three years relative to the previous three years.

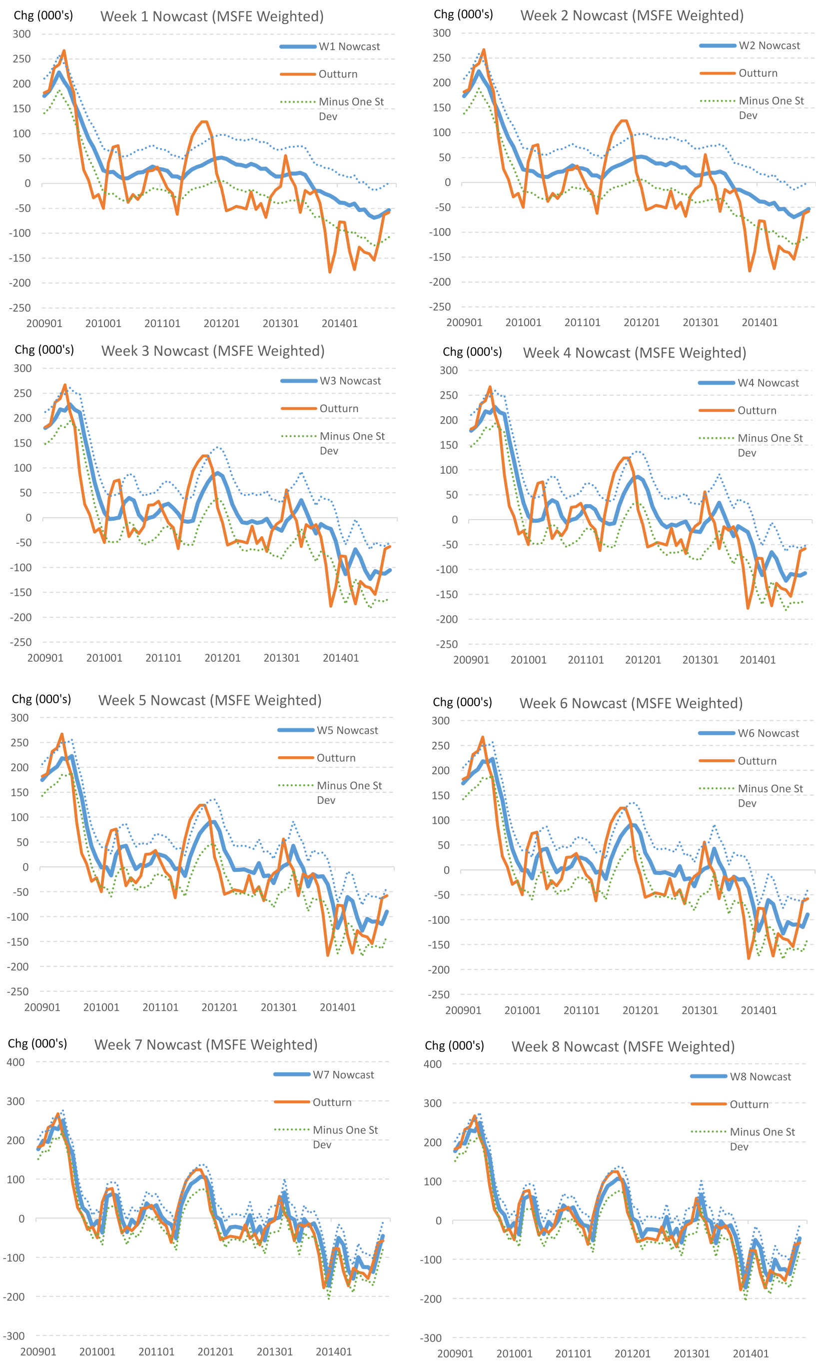
These observations match-up with the recent performance of the labour market (from a macroeconomic perspective). Figure 4.6 indicates that the model tended to underestimate the extent to which unemployment was falling in recent years. Such upside surprises seem to tally with the so-called UK productivity puzzle, debated by many commentators caught out by strong reductions in unemployment at a time when overall UK economic performance was underwhelming.

4.7 Chapter 4 Summary

Given the recent interest in internet search data as a predictor of economic variables, this chapter's aim was to provide an assessment of the usefulness of information from Google Trends for such forecasting purposes. Internet search data have a number of desirable features to policymakers and investors, particularly around their timeliness which helps to bolster their ability to offer guidance to changes in slower released official measures of unemployment.

Moreover, Google data are available to even quicker timescales than surveys, are provided to a greater time frequency and, for the time being, available freely,

Figure 4.6: Evolution of MSFE Weighted Nowcasts



making them a cost-effective alternative source of qualitative economic information.

A drawback of Google data (or internet search data) is that finding suitable search terms is challenging. There needs to be careful consideration of where to start. Economic intuition is one such place and, in line with previous literature, has been shown to assist with the selection of a "root" term that underpins a composite index based on associated search terms.

Some success was found in using a "redundancy" root-term in helping to predict changes in UK unemployment over a longer-time period than similar research conducted by the Bank of England. Starting from a traditional regression-based one-step ahead out-of-sample forecasting exercise, derived Google Trends data offered similar (and at times better) performance in terms of forecast accuracy to survey-based counterparts. The performance of both Google and survey-based data was especially strong during 2009-2012, a period which includes the tail-end of a very deep period of economic recession. However, beating a naive benchmark has subsequently proved more challenging, with all models finding it difficult to outperform a simple no-change forecast since 2012.

As has been argued by several authors, including key players in the emerging field of nowcasting, a true assessment of the quality of any predictor should also include a look at its ability to reduce nowcast uncertainty in real-time. Google data have a timing advantage and the value of this should be explored. Sections 4.5 and 4.6 set-up a small-scale pseudo-time nowcasting framework for such a purpose.

The results were encouraging, with derived Google indicators tending to record, relative to competing survey variables, reduced nowcast errors in a lower range of uncertainty in weeks 3 and 4 of an eight-week unemployment nowcasting window. These signals were subsequently confirmed by the release of the weekly survey-based data in week 5. This timing advantage could be useful to policymakers in offering an early steer on likely directional changes in unemployment and seem to offer a viable supplementary source of information to be used in the policy-decision making process.

Note the word supplementary is carefully selected here; there were many challenges to overcome when deriving robust Google indicators and there is vast evidence of the benefits of using multiple variables in nowcasting to guard against possible breakdowns between dependent variables and individual indicators. A sensible strategy seems to use the data as a complement rather than a substitute to surveys.

While a relatively simple framework, deliberately so to compare the performances of Google and survey-based variables, the weekly nowcasting model actually did a good job at predicting three-month changes in unemployment and offers further positive "proof of concept" evidence of search-term data usefulness in the forecasting arena. While admittedly the choice of a root-term "redundancy" was designed primarily to offer a viable alternative to colloquial terms for unemployment benefits that are designed to measure the flow into unemployment, it seems a natural progression to extend the model by looking at terms that also measure the flow out of unemployment. Given the segmentation of the labour market, this will be a challenge. Perhaps individual models measuring specific industries/areas could be created and then combined to create an overall indicator.

Moreover, the nowcasting framework presented in this chapter can be easily extended to other areas, such as household consumption, an area not well covered by official data. There is already tentative evidence of value in Google data in predicting retail trends (see e.g. Chamberlain (2010)) which can be built upon. Other high frequency electronic sources of data such as those provided by debit card providers can also be incorporated into such a framework. MIDAS regressions make such ideas easy to implement.

Finally, a word of caution. Google has shown a penchant for closing down services with little notice, so there is a risk to researchers that sources of interesting information are withdrawn. For instance, Google Correlate, which was originally considered as an alternative source of data for this chapter and advocated by Varian (2014) as a search-term selection source for predicting new housing sales in the United States is, rather frustratingly, no longer updated by Google.

Chapter 5

Closing Remarks and Future Developments

The research that forms *Chasing Yesterday: Nowcasting Economic Activity with Timely Indicators* has led to a number of contributions to the field of nowcasting, some adding weight to earlier assertions from elsewhere in the literature, and others offering new insights for researchers.

From an overall perspective, two general observations can be made regarding the usefulness of timely economic indicators.

Firstly, high frequency data provided by e.g. business surveys are extremely useful in helping to track economic activity and improving the accuracy of intra-period nowcasting of macroeconomic data.

Secondly, for both GDP and unemployment, the usefulness of timely data sources is especially prominent during periods of economic crisis. The deep global recession of 2008/2009 offers such a period to observe how nowcasting models that incorporate timely indicators performed. The accuracy of these models was greatly enhanced during the financial crisis relative to benchmarks such as AR(1) specifications. During the so-called Great Moderation these benchmarks were shown to have been difficult to beat.

There was already a growing consensus amongst the research community around these two points, but the research within the thesis stretched and extended the literature through the utilisation of real-time data vintages. As databasing of vintage data becomes better and easier to work with, a greater emphasis seems likely to be placed on using such sources going forward. Studies that are based on "pseudo-time" information will become an exception rather than the norm.

A corollary of this, and a point that was demonstrated in chapter 3, is that comparisons of model generated nowcasts against consensus-based nowcasts must utilise vintage data to ensure their validity. By using real-time data, model based nowcasts were considerably weaker than those provided by a consensus of analysts (who it is assumed use some element of judgement in their assessments).

Any generalisation of such findings is of course somewhat dangerous as the comparison was made for a single country (the UK), where it was noted the GDP data generating process may have unique properties. While the latter led to a new potential path for researchers interested in predicting UK GDP to focus on i.e. the relationships between preliminary estimates of GDP and business survey data, chapter 3 nonetheless raises questions on the validity of existing claims by pseudo-time studies of consensus-busting model performance.

In a second broad departure from the general norm of focusing on single countries/regions, in chapter 2 the two key findings mentioned above were demonstrated via a rare cross-country study.

However, the larger contribution from this chapter was to confirm that the closely-watched and well used Purchasing Managers' Indices (PMIs) are not only helpful in tracking economic activity, but can help to reduce errors early in the intra-period nowcasting of quarterly GDP.

Such is their usefulness in terms of timing, PMIs and similar survey data can actually diminish the value of later released "hard" indicators which have traditionally been viewed as the most important information sources in the data release cycle. This adds credence to the observation that PMIs are widely used around the world within central banks and other such policy-making institutions: their timeliness is of up-most importance to those operating in a world where decision-making has to be made in real-time.

Armed with the confirmation that timing is an important characteristic of an economic indicator, chapter 4 took this mantra and applied it to a relatively new information source: internet search data.

Using an extract of Google Trends data, based on a new idea of combining economic intuition and a more automated method, the resulting time series was able to perform just as well as closely-watched survey data in terms of nowcasting changes in UK unemployment.

The resulting conclusion that Google Trends series offers reliable signals, which are then corroborated by later-released data sources, is an especially important finding as it adds credibility to Google Trends data for institutions wishing to receive an even earlier steer on economic trends than currently offered by more traditional data sources.

In terms of future research developments, interest on the work contained in the thesis will likely be focused on the emerging field of “big data” and how to use these new sources of information in nowcasting applications.

As demonstrated in chapter 4, tools such as MIDAS regressions offer a flexible way to incorporate various mixed time frequency data within nowcasting applications, providing a platform to observe the contribution that new indicators make in reducing uncertainty and strengthening accuracy through a nowcasting cycle.

Outside of internet search data and the opportunities already mentioned in chapter 4, the creation of indicators from other sources of information that can then be incorporated into nowcasting frameworks seems a natural research area for future research to focus on.

For example, using text analytics to analyse written economic reports is an interesting development and seems to offer a number of opportunities to extract new quantitative measures of current economic conditions see e.g. Balke, Fulmer, and Zhang (2015).

Electronics payments data is another exciting development opportunity due to not only the considerable population coverage that these data sources provide, but also the option to observe in a timely fashion the impacts on consumption of economic shocks or the insight provided on sectors of the economy not well covered by official statistics. Similar goals can of course also be applied to internet search data.

Greater sectoral analysis would be an important development in the nowcasting sphere as the majority of nowcasting tends to be concerned with the macroeconomic picture (e.g. GDP, unemployment). Such a focus therefore offers the chance for greater "story-telling" and the general economic interpretation currently lacking in the field.

Finally, an observation emanating from the thesis research is that nowcasting models continue to struggle to capture turning points in the economic cycle in a timely fashion. This in part seems to reflect the use of AR(1) terms or the assumption of constant parameters within many nowcasting regressions. While probably fine in periods of relative economic calm, the increased macroeconomic volatility seen during recent years nonethless means the predictive power of nowcasting models remains under pressure.

With this in mind, the incorporation of time-varying characteristics within nowcasting frameworks, such as the Markov-Switching MIDAS models proposed by Guerin and Marcellino (2013), is a promising area.

Focusing on the signals provided by the timely indicators such as PMIs, and how these could be better incorporated and quantified in terms of changes in GDP, is an especially interesting development opportunity.

Appendix A

Appendix: Notes Related to Chapter 4

A.1 Conversion of Google data into Monthly Values

The monthly values for the Google search volumes are derived by assuming that the weekly values are the same throughout the Google reporting week (Sunday through Saturday). These daily numbers are then multiplied by the number of occurances for the week (usually seven, but this can vary for those weeks that straddle two months). The sum is then divided by the number of days in the month, as per the numerical example below:

Week	Search Volume
28/09/2014 to 04/10/2014	57
05/10/2014 to 11/10/2014	65
12/10/2014 to 18/10/2014	72
19/10/2014 to 25/10/2014	50
26/10/2014 to 01/11/2014	65

$$October = \frac{(4 * 57 + 7 * 65 + 7 * 72 + 7 * 50 + 65 * 6)}{31} = 62.2$$

A.2 Google Search Terms in GRI

redundancy	redundancy pay	redundancy calculator
redundancy payments	uk redundancy	statutory redundancy
statutory redundancy pay	redundancy notice	how much redundancy
redundancy law	redundancy payment	tax redundancy
redundancy notice period	redundancy letter	redundancy uk
voluntary redundancy	redundancy pay calculator	tax redundancy pay
rpo	redundancy rights	redundancy consultation

A.3 Chapter 4 Explanatory Variables: Details

GoogJSA: Google Search Term "JSA", weekly normalised search volumes provided by the Google Trends Interface.
Data Source: Google Trends

GoogRed: Google Search Term "Redundancy", weekly normalised search volumes provided by the Google Trends Interface
Data Source: Google Trends

GRI: Google Redundancy Index: Index derived from weighted Google terms related to redundancy (see appendix A.2 for these terms).
Data Source: Google Trends

RoJPP: Report on Jobs Permanent Placements: Seasonally adjusted diffusion index from the KPMG/REC Report on Jobs survey of recruitment consultants. Data are derived from the question: "Please compare the number of staff placed in permanent positions with the number one month ago."
Data Source: KPMG/REC

RoJTB: Report on Jobs Temporary Billings: Seasonally adjusted diffusion index taken from the KPMG/REC Report on Jobs survey of recruitment consultants.

Data are derived from the question: "Please compare your billings received from the employment of temporary and contract staff with the situation one month ago."

Data Source: KPMG/REC

RoJVC: Report on Jobs Vacancies Index: Seasonally adjusted diffusion index taken from the KPMG/REC Report on Jobs survey of recruitment consultants. Recruitment consultants are asked to specify whether the demand for staff from employers has changed on the previous month, thereby providing an indicator of the number of job vacancies.

Data Source: KPMG/REC

RoJSA: Report on Jobs Staff Availability Index: Seasonally adjusted diffusion index taken from the KPMG/REC Report on Jobs survey of recruitment consultants. Data are derived from the question: "Is the availability of candidates for permanent vacancies better, the same or worse than one month ago?"

Data Source: KPMG/REC

PMIEM: Purchasing Managers' Index Employment Index: Seasonally adjusted diffusion index taken from the UK Purchasing Managers' Index (PMI) dataset, specifically the question related to changes in employment from one month to the next. The index is based on survey responses from the manufacturing, services and construction sectors.

Data Source: Markit Economics

ConExp: Consumer expectations for unemployment: Taken from the DG EcFin consumer survey of UK households. Respondents are asked to provide their expectations for unemployment over the next 12 months. Data are provided as a seasonally adjusted net balance.

Data Source: DG EcFin

BoESR: Bank of England Agents Survey (Services): Net balance regarding the

employment intentions of service sector companies. Taken from the Bank of England Agents Summary of Business Conditions.
Data Source: Bank of England

BoEMN: Bank of England Agents Survey (Manufacturing): Net balance regarding the employment intentions of manuacturing companies. Taken from the Bank of England Agents Summary of Business Conditions.
Data Source: Bank of England

F1: Derived factor based purely on the survey-based indicators. Factor derived from static principal components analysis.
Data Source: Author's Calculations

F2: Derived factor based on the survey-based indicators plus the Google Redundancy and GRI variables. Factor derived from static principal components analysis.
Data Source: Author's Calculations

Note: All data are available on a monthly basis, except the Google data, which are provided by the Google Trends interface on a weekly time frequency.

A.4 Conversion of Google data into Weekly Values

As per the weekly into monthly conversion, first assume that the weekly values are unchanged through the reporting week of Sunday through Saturday. Daily numbers are then averaged over the respective calendar periods of 1-7, 8-14, 15-21 and 22nd to month end as per the numerical example below for October 2014:

Week	Search Volume
28/09/2014 to 04/10/2014	57
05/10/2014 to 11/10/2014	65
12/10/2014 to 18/10/2014	72
19/10/2014 to 25/10/2014	50
26/10/2014 to 01/11/2014	65

$$W1 = \frac{(4 * 57 + 3 * 65)}{7} = 60.4$$

$$W2 = \frac{(4 * 65 + 3 * 72}{7} = 68.0$$

$$W3 = \frac{(4 * 72 + 3 * 50)}{7} = 62.6$$

$$W4 = \frac{4(*50 + 6 * 65)}{10} = 59.0$$

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.

Alessi, L, E Ghysels, L Onorante, R Peach, and S Potter (2014). "Central Bank Macroeconomic Forecasting during the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York Experiences". *ECB Working Paper Series* 1688.

Allan, G, G Koop, S McIntyre, and P Smith (2014). "Nowcasting Scottish GDP Growth". *University of Strathclyde Discussion Papers in Economics* 14-11.

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.

Angelini, E, G Camba-Mendez, G Runstler D Giannone, and L Reichlin (2008). "Short term forecasts of euro area GDP growth". *ECB Working Paper Series* 949.

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.

Askitas, N and K F Zimmermann (2009). "Google Econometrics and Unemployment Forecasting". *Institute for the Study of Labour (IZA) Discussion Paper* No. 4201 June 2009.

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, E Ghysels, and J Wright (2013). "State Space Models and MIDAS Regressions". *Econometric Reviews* 32.7, pp. 779–813.

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

Balke, S, M Fulmer, and R Zhang (2015). "Incorporating the Beige Book into a Quantitative Index of Economic Activity". *Southern Methodist University* November 2015.

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, D Giannone, and L Reichlin (2011). *Nowcasting.* Ed. by M P Clements and D F Hendry. Chapter 10, Oxford Handbook on Economic Forecasting. Oxford University Press.

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.

Bangwayo-Skeete, P F and R W Skeete (2015). "Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach". *Tourism Management* 46.February 2015, pp. 454–464.

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

Bentley, R A, R Nyman, and P Ormerod (2014). *Nowcasting economic and social data: when and why search engine data fails, an illustration using Google Flu Trends*. URL: http://arvix.org/abs/1408.0699.

Bholat, D (2015). "Big Data and Central Banks". *Bank of England Quarterly Bulletin* Q1 2015.

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.

Britton, E, P Fisher, and J Whitley (1998). "The Inflation Report Projections: Understanding the Fan Chart". *Bank of England Quarterly Bulletin* 38, pp. 30–37.

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.

Bundesbank (2014). "Forecasting models in short-term business cycle analysis – a workshop report". *Deutsche Bundesbank Monthly Report* September 2013.69–83.

Camacho, M and G Perez-Quiros (2010). "Introducing the EURO-STING: Short Term Indicator of Euro Area Growth". *Journal of Applied Econometrics* 25.4, pp. 663–694.

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.

Chamberlain, G (2010). "Googling the present". *Office for National Statistics Economic and Labour Market Review* 4.12.

Choi, H and H R Varian (2009a). "Predicting the Present with Google Trends". *Technical Report, Google Inc* April 2009.

— (2009b). "Using Internet Search Data as Economic Indicators". *Technical Report, Google Inc* July 2009.

Clements, M P and A B Galvao (2008). "Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US output growth". *Journal of Business and Economic Statistics* 26.4, pp. 546–554.

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

Croushore, D and T Stark (2001). "A real-time dataset for macroeconomists". *Journal of Economic Econometrics* 105.1, pp. 111–130.

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.

D'Amuri, F (2009). "Predicting Unemployment in Short Samples with Internet Job Search Query Data". *MPRA Paper* No. 18403.

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.

Diron, M (2008). "Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data". *Journal of Forecasting* 27.5, pp. 371–390.

Doz, C, D Giannone, and L Reichlin (2007). "A two-step estimator for large approximate dynamic factor models based on Kalman Filtering". *CEPR Discussion Papers* 6043.

ECB (2008). “Short-term forecasts of economic activity in the euro area”. *ECB Monthly Bulletin* April 2013, pp. 69–74.

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

Esteves, P S (2009). “Are ATM/POS data relevant when nowcasting private consumption”. *Banco de Portugal Working Papers* No 2009-25.

Ettredge, M, J Gerdes, and G Karuga (2005). “Using Web-based Search Data to Predict Macroeconomic Statistics”. *Communications of the ACM* 48.11, pp. 87–92.

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.

Fondeur, Y and F Karame (2013). “Can Google data help predict French youth unemployment?” *Economic Modelling* 30.c, pp. 117–125.

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

Forni, M, M Hallin, M Lippi, and L Reichlin (2002). “The generalized factor model: identification and estimation”. *Review of Economics and Statistics* 82.4, pp. 540–554.

Galbraith, J W and G Tkacz (2007). “Electronic Transactions as High Frequency Indicators of Economic Activity”. *Bank of Canada Working Paper* No 2007-58.

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.

Gill, T, D Perera, and D Sunner (2012). "Electronic Indicators of Economic Activity". *Australian Reserve Bank Bulletin* June (2012).

Ginsberg, J, M H Mohebbi, R S Patel, L Brammer, M S Smolinski, and L Brilliant (2009). "Detecting influenza epidemics using search engine query data". *Nature* 457, pp. 1012–1014.

Grasmann, P and F Keereman (2001). "An indicator-based short-term forecast for quarterly GDP in the euro area". *European Commission Economic Paper* 154.

Guerin, P and M Marcellino (2013). "Markov-Switching MIDAS Models". *Journal of Business and Economic Statistics* 31.1, pp. 45–56.

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.

Hamilton, J D and J C Wu (2014). "Risk Premia in Crude Oil Futures Prices". *Journal of International Money and Finance* 42.April 2014, pp. 9–37.

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.

Kholodilin, K A, M Podstawski, and B Silverstovs (2010). "Do Google searches help in nowcasting private consumption? A real-time evidence for the US". *KOF Working Papers* April.256.

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.

Koenig, E, S Dolmas, and J M Piger (2001). "The Use and Abuse of Real-Time Data in Economic Forecasting". *Federal Reserve Bank of St. Louis, Working Paper* 2001-015A.

Koop, G and L Onorante (2014). "Macroeconomic nowcasting using Google probabilities". *Mimeo* August 2014.

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.

Lazer, D, R Kennedy, G King, and A Vespignani (2014). "The parable of Google Flu: Traps in big data analysis". *Science* 343.6176, pp. 1203–1205.

Libero, M and G Moretti (2013). "Real Time Forecasts of Inflation: The Role of Financial Variables". *Journal of Forecasting* 32.1, pp. 51–61.

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.

Lima, F (2013). "The use of payments data to improve monetary and financial analysis". *Banco De Portugal Workshop on Integrated Management of Micro-databases.*

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.

Matheson, T D (2010). "An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys". *Economic Modelling* 27.1, pp. 304–314.

Mazzi, GL, J Mitchell, and G Montana (2010). "Density Nowcasts and Model Combination: Nowcasting Euro-Area GDP Growth over the 2008-9 Recession". *NIESR Discussion Papers* 368.

McLaren, N and R Shanbhogue (2011). "Predicting Initial Claims for Unemployment Benefits". *Bank of England Quarterly Bulletin* 51.2, pp. 134–140.

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.

Mitchell, J, RJ Smith, M Weale, S Wright, and EL Salazar (2005). "An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth". *Economic Journal* 115.1, pp. 108–129.

Mourougane, A and M Roma (2002). "Can confidence indicators be useful to predict short-term real GDP growth?" *ECB Working Paper Series* 133.

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.

Parigi, G and G Schlitzer (1995). "Quarterly Forecasts of the Italian Business Cycle by Means of Monthly Economic Indicators". *Journal of Forecasting* 14.2, pp. 117–141.

Pesaran, MH and M Weale (2006). *Survey expectations.* Ed. by G Elliott, CWJ Granger, and A Timmermann. Handbook of Economic Forecasting, Volumne 1. North-Holland: Amsterdam, pp. 715–776.

Pesaran, MH and M Weale (2013). *Advances in Forecasting under Model Instability*. Ed. by G Elliott and A Timmermann. Handbook of Economic Forecasting, Volumne 2B. Elsevier Publications, pp. 1203–1324.

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.

Rünstler, G and F Sédillot (2003). “Short-Term Estimates of Euro Area Real GDP By Means Of Monthly Data”. *European Central Bank* Working Paper Series 276.

Sargent, T J and C A Sims (1977). *Business cycle modeling without pretending to have too much a-priori economic theory*. New Methods in Business Cycle Research. Federal Reserve Bank of Minneapolis.

Schmidt, T and S Vosen (2009). “Forecasting private consumption: survey-based indicators vs. Google trends”. *RUHR Economic Paper* 155.

Schorfheider, F and D Song (2011). “Real-time forecasting with a mixed frequency VAR”. *mimeo*.

Schumacher, C (2014). “MIDAS and Bridge Equations”. *Deutsche Bundesbank Discussion Paper* No 26/2014.

Smith, G (2012). “Google Internet search activity and volatility prediction in the market for foreign currency”. *Finance Research Letter* 9.2, pp. 103–110.

Stock, J H and M W Watson (2002). “Macroeconomic forecasting using diffusion indexes”. *Journal of Business and Economic Statistics* 20.2, pp. 147–162.

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

Suhoy, T (2009). “Query indices and a 2008 downturn: Israeli data”. *Bank of Israel Discussion Paper* 2009.06.

Tay, AS and KF Wallis (2000). “Density Forecasting: A Survey”. *Journal of Forecasting* 19.4, pp. 235–254.

Taylor, L, R Schroeder, and E Meyer (2014). *Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same.* URL: http://dx.doi.org/10.1177/2053951714536877.

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).

Varian, H R (2014). “Big Data. New Tricks for Econometrics”. *Journal of Economic Perspectives* 28.2, pp. 3–28.

Yang, X, P Pan, J A Evans, and B Lv (2015). “Forecasting Chinese Tourist Volume with Search Engine Data”. *Tourism Management* 46, pp. 386–397.

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