

UNCERTAIN KINGDOM: NOWCASTING GDP AND ITS REVISIONS[☆]

Nikoleta Anesti*
Bank of England

Ana Beatriz Galvão[†]
University of Warwick

Silvia Miranda-Agrippino[‡]
Bank of England and CFM

August 22, 2018

Abstract

We design a new econometric framework to nowcast macroeconomic data subject to revisions, and use it to predict UK GDP growth in real-time. To this aim, we assemble a novel dataset of monthly and quarterly indicators featuring over ten years of real-time data vintages. Successive monthly estimates of GDP growth for the same quarter are treated as correlated observables in a Dynamic Factor Model (DFM) that also includes a large number of mixed-frequency predictors, leading to the release-augmented DFM (RA-DFM). The framework allows for a simple characterisation of the stochastic process for the revisions as a function of the observables, and permits a detailed assessment of the contribution of the data flow in informing (i) forecasts of quarterly GDP growth; (ii) the evolution of forecast uncertainty; and (iii) forecasts of revisions to early released GDP data. By evaluating the real-time performance of the RA-DFM, we find that the model's predictions have information about the latest GDP releases above and beyond that contained in the statistical office earlier estimates; predictive intervals are well-calibrated; and UK GDP growth real-time estimates are commensurate with professional nowcasters. We also provide evidence that statistical office data on production and labour markets, subject to large publication delays, account for most of the forecastability of the revisions.

Keywords: Nowcasting; Data Revisions; Dynamic Factor Model.

JEL Classification: C51; C53; C55

*Monetary Analysis, Bank of England, Threadneedle Street, London EC2R 8AH, UK.
E-mail: nikoleta.anesti@bankofengland.co.uk

[†]Warwick Business School, University of Warwick, CV4 7AL, Coventry, UK.
E-mail: ana.galvao@wbs.ac.uk Web: <https://sites.google.com/site/anabgalvao/>

[‡]corresponding author: Monetary Analysis, Bank of England, Threadneedle Street, London EC2R 8AH, UK. W: www.silviamirandaagrippino.com E: silvia.miranda-agrippino@bankofengland.co.uk

☆ We are grateful to Domenico Giannone, Simon van Norden, Gianni Amisano, Juan Antolin-Diaz, Ivan Petrella, and conference participants at the St Louis Fed Central Bank Forecasting Conference, the BCRA Nowcasting and Big Data Workshop, the 1st Vienna Workshop on Economic Forecasting, the 2018 IAAE Annual Meeting, and the Now-Casting.com 2018 Meeting for helpful comments and discussions. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of England or any of its Committees.

1 Introduction

Macroeconomic forecasters face a dilemma when nowcasting current economic conditions and key macroeconomic aggregates such as GDP, consumption, and investment. Because the release of these data is subject to potentially large revisions, it is not normally clear which release (or estimate) a forecaster should aim to nowcast. On the one hand, early releases permit a timely assessment of current economic conditions, but are inherently incomplete. On the other, the publication delay of more mature estimates can act as a deterrent even if these data are potentially more reliable. Choosing a specific forecast target is ultimately a question of the implicit loss function of the forecaster. Policy institutions may prefer to target more mature releases, as they are likely to paint a more realistic portrait of current conditions, conditional on which policies are then designed and implemented. Other forecasters, such as e.g. market participants, may prefer to forecast earlier releases instead, as the payoff of their trading strategies may be linked to that particular realisation.

Observers, market participants, and professional forecasters partially resolve this trade-off by typically considering large numbers of predictors when forming their expectations, including successive monthly releases of GDP growth for any given quarter (see e.g. [Clements and Galvão, 2017](#)). Conversely, statistical nowcasting models tend to discard information on GDP data revisions; even with real-time data, the focus is usually on one specific release at a time, normally, either the latest available vintage of data, or the first release (see e.g. [Bańbura et al., 2013](#), for a comprehensive review). In this paper, we propose a new econometric framework to nowcast data subject to revisions that mimics as closely as possible the behaviour of a professional forecaster by explicitly taking into account the information contained in subsequent estimates of the target GDP variable. We call this model the Release-Augmented Dynamic Factor Model, or RA-DFM. In the RA-DFM, we augment an otherwise standard mixed-frequency DFM with successive monthly estimates of quarterly GDP growth for the same quarter, and treat them as separate but correlated observables, thereby exploiting their intrinsic factor structure. Subsequent releases relative to the same reference period can in fact be thought of as increasingly more accurate estimates of the same object. Statistical nowcasting models

are built on the intuition that other data that correlate with GDP growth but are released in a more timely fashion, and also sampled at higher frequencies, can help form an early assessment of current economic conditions before the official GDP numbers are published. The RA-DFM extends this concept also to revisions to early GDP estimates. Initial rounds of revisions to early released data are typically due to the fact that as time goes by and more information is accumulated, the national statistical offices can refine their assessment of past events. Hence, there is scope for using the real-time data flow to inform forecasts of the successive releases for any given quarter. Our framework allows for a simple characterisation of the stochastic process for the revision as a function of the observables, and permits a detailed assessment of the contribution of the real-time data flow in informing (i) forecasts of quarterly GDP growth; (ii) the evolution of forecast uncertainty around point estimates; and (iii) the forecasts of revisions to early released GDP data.

Augmenting an otherwise standard mixed-frequency DFM with successive monthly GDP releases has a number of advantages compared to models in the previous literature. First, the common component of these releases is driven by the full set of economic and financial indicators in the model. As a consequence, we are able to exploit the information content of a large number of indicators when predicting the initial monthly data revisions. This is an advantage compared to existing models that allow to incorporate only a small number of indicators, if any at all (see [Kishor and Koenig, 2012](#); [Cunningham et al., 2012](#)). Second, the states vector grows linearly in the number of revision rounds considered. This allows us to retain parsimony, and facilitates the estimation of these types of models with real-time data, in contrast to specifications in e.g. [Kishor and Koenig \(2012\)](#); [Galvão \(2017\)](#). Third, we allow for serial and (weak) cross-sectional correlation among the idiosyncratic revision components, which can improve forecasting performance as suggested by [Clements and Galvão \(2013a\)](#). Our approach to modelling is thus compatible with the fact that initial data revisions are predominantly caused by the statistical office having access to a larger number of input data. Finally, with a straightforward extension of the ‘news’ decomposition of [Bańbura and Modugno \(2014\)](#), we are able to discern the contribution of different pieces of data to the forecast of subsequent releases, effectively providing a framework to forecast the official revisions of the

statistical office.

We use the RA-DFM to compute point and density nowcasts of UK real GDP growth in real-time. To this aim, we assemble a comprehensive mixed-frequency real-time dataset that features over 10 years of real-time data vintages with history going back to January 1990. The main source for the construction of our real-time dataset is the archive of the Bank of England, in which data released by the UK Office of National Statistics (ONS) have been carefully stored over the years. We make this dataset available to the broader research community. To the best of our knowledge, ours is the most comprehensive real-time mixed-frequency dataset for the UK economy in terms of breadth and coverage. The Bank of England maintains a real-time database that only covers quarterly variables, and details on its construction are in [Castle and Ellis \(2002\)](#); [Garratt and Vahey \(2006\)](#). An early mixed-frequency real-time dataset for the UK economy was introduced in [Egginton et al. \(2002\)](#); this dataset, however, was only last updated at the end of 1999, and covered a smaller cross-section compared to ours. Our data covers the indexes of production and services, labour market indicators, macroeconomic aggregates such as consumption, investment, and international trade, as well as surveys, and credit measures and financial variables. The complete list encompasses all the ‘market movers’ that feature in the most prominent economic calendars, such as those distributed by Bloomberg and Thomson Reuters.

Nowcasting the UK economy is particularly challenging, and as consequence interesting for academic researchers, for a number of reasons. First, the data flow in the UK is substantially ‘slower’ than in the US, for example. Industrial production data are released 15 days after the end of the reference month in the US, while it takes an extra month to have the equivalent number for the same period released in the UK. Second, the UK is fundamentally a service-based economy. Hence, the conventional wisdom whereby “industrial production and PMIs are sufficient to gauge the stance of current real activity” does not necessarily apply, and can in fact lead to severe forecasting misjudgements. Yet, however, the first official data on the output produced in the service sector are released with almost two months of delay with the respect to the end of the reference period (i.e. at the end of March for January), and are hence of little relevance for real-time nowcasting. Finally, UK data are characterised by a substantial degree of noisiness, which adds

complexity to any signal-extraction exercise.

Revisions to the official preliminary ONS GDP estimates can be substantial. In the UK the Office for National Statistics publishes a first estimate for GDP growth 4 weeks after the end of the reference quarter; these preliminary GDP figures are typically estimated based on data covering only a fraction (about 44%) of the required survey data.¹ As new information accumulates, a second and third estimates for the same quarter are published after 8 and 13 weeks respectively. These numbers are further revised in the following three years. Revisions to early released data have in some cases led to a complete reversal of early assessments of UK growth. A notable case is the so-called ‘Double-Dip’ recession episode of late 2011/early 2012, that was erased from the data following a major revision in 2013. Hence, accounting for data uncertainty when nowcasting the UK economy is important.

Combining our real-time dataset with the calendar of data releases back to 2006 yields over 1,500 real-time data vintages over which we evaluate the out-of-sample forecasting performance of the RA-DFM model. We find that the model’s forecasts are robust to variations in the model specification, that the model produces well-calibrated predictive intervals, and that nowcasts and forecasts for the first release beat by a large margin naïve forecasts that discard information in the real-time data flow. This is in line with earlier studies (see e.g. [Bańbura et al., 2013](#)). Our framework, however, permits going beyond the first release and analysing the contribution of data releases in forecasting revisions to earlier GDP estimates. We find that monthly indicators are informative for predicting the first revision rounds, and that their predictive content is exhausted after the second month after the end of the reference quarter. ‘Hard’ data, such as production and labour market indicators partially owe their predictive content to their large publication delay. Figures relative to the last month of the reference quarter are released with up to two months of delay in the UK. This makes them relevant in forecasting the first GDP revision. Moreover, we find that the RA-DFM forecasts contain information that is relevant for forecasting ‘true’ growth – measured using a later vintage – above and beyond that contained in official ONS estimates. While retaining parsimony, and without

¹<https://www.ons.gov.uk/economy/grossdomesticproductgdp/qmis/grossdomesticproductgdpqmi>

the introduction of any element of judgement, the model performs well when compared to model averages and judgement-based forecasts embedded in the predictions of institutional forecasters. For what concerns in particular the ‘Double-Dip’ recession episode, we find that conditional on information available back in late 2011, the RA-DFM model was able to extract a reliable signal from the real-time data flow which never pointed towards a recession.

Our paper builds on the large body of literature on nowcasting GDP growth using Dynamic Factor Models (DFM), firstly introduced by [Giannone, Reichlin and Small \(2008\)](#). Numerous applications to point forecasts using (typically pseudo) real-time data have been proposed over the years, and for a number of different countries (see [Bańbura et al., 2013](#), for a review). [Aastveit et al. \(2014\)](#)’s evaluation includes a DFM for nowcasting US growth using fully real time data. Forecasting models with factors have been previously applied for predicting UK GDP growth ([Artis et al., 2005](#); [Mitchell, 2009](#); [Miranda-Agrippino, 2012](#); [Anesti et al., 2017](#)), but forecasting evaluations typically use only the latest available vintage for monthly and quarterly indicators. This might be linked to the lack of availability of real-time databases for UK monthly economic indicators. By building a mixed-frequency real-time dataset for the UK economy, we are able to deliver a truly real-time forecasting evaluation for UK GDP growth.

Our paper is also related to the literature that have proposed methods for macroeconomic forecasting when dealing with data subject to revisions. The literature has typically approached the issue of data uncertainty in forecasting in three ways. The first one evaluates a model’s forecasts using only the data that actually were available to a professional forecaster at each forecast origin (see e.g. survey in [Croushore, 2006](#)). The second improves over the first one by estimating forecasting models using data with the same number of data revision rounds as the data actually available to condition on when computing forecasts in real time ([Koenig et al., 2003](#); [Clements and Galvão, 2013b](#)). The third one requires the joint estimation of the forecasting model and the data revision process with the aim of predicting revised future values of the variable of interest ([Kishor and Koenig, 2012](#); [Cunningham et al., 2012](#); [Clements and Galvão, 2013a](#); [Carriero et al., 2015](#); [Galvão, 2017](#)). Our approach not only addresses the issue of how to compute short-term forecasts of data subject to revision, but it also provides a method to evaluate the

effects of data news on the prediction of data revisions. In this respect, our approach is closest to the third branch of the literature reviewed above.

The paper is organised as follows. Section 2 describes our model, and details our approach to handle data revisions and update point and density forecasts in line with the real-time arrival of new pieces of data. We discuss the characteristics of UK data revisions and the construction of the real-time dataset in Section 3. All the empirical results are collected in Section 4. Finally, Section 5 concludes.

2 A Nowcasting Model for Macroeconomic Data Subject to Revisions

In this section we describe the framework that we design to model and forecast subsequent GDP releases relative to the same quarter, the model-implied stochastic process for the revisions, and how we assess the relevance of ‘data news’ in forecasting GDP growth and revisions to early released data. While our application focuses on GDP growth, the framework can be applied to any macroeconomic aggregate for which the statistical office releases successive updates to early estimates for the same quarter.

2.1 The DFM model

We begin with the mixed-frequency Dynamic Factor Model (DFM) in [Bańbura et al. \(2013\)](#). Let x_t^M denote a generic $n_M \times 1$ vector of demeaned stationary monthly variables observed at $t = 1, 2, \dots, T$. Similarly, x_t^Q denotes a $n_Q \times 1$ vector of quarterly zero-mean stationary variables observed at $t = 3, 6, \dots, T$.

We assume that x_t^M has a factor structure, and write:

$$x_t^M = \mathbf{A}_M f_t + \zeta_t^M, \tag{1}$$

$$f_t = \mathbf{A}_1 f_{t-1} + \dots + \mathbf{A}_p f_{t-p} + \eta_t \quad \eta_t \sim \mathcal{N}(0, \mathbf{\Sigma}), \tag{2}$$

$$\zeta_t^M = \mathbf{D} \zeta_{t-1}^M + \epsilon_t \quad \epsilon_{i,t} \sim \mathcal{N}(0, \varsigma_i^2), \tag{3}$$

where $i = 1, \dots, n_M$ and $f_t \equiv (f_{1,t}, \dots, f_{r,t})'$ is an $r \times 1$ vector of zero-mean unobserved factors. The variables in x_t^M load on the factors via the coefficients in $\mathbf{\Lambda}_M$. The factors follow a VAR(p), and \mathbf{A}_j , $j = 1, \dots, p$ are r -dimensional square matrices of autoregressive coefficients. ζ_t^M is a vector of AR(1) idiosyncratic terms, with \mathbf{D} a diagonal matrix of univariate autoregressive coefficients. Zero restrictions can be imposed on $\mathbf{\Lambda}_M$ to facilitate the interpretation of the factors as being specific to e.g. labour market conditions, surveys and financial markets.

The approximation for flow variables of [Mariano and Murasawa \(2003\)](#) allows to incorporate quarterly variables into the framework by constructing partially observed monthly series that are assumed to follow the same factor structure in Eq. (1). Specifically, let X_t^Q denote the level of a quarterly variable, e.g. log level of consumption. Let \tilde{X}_t^M be its unobservable monthly counterpart, and define $\tilde{x}_t^M \equiv \tilde{X}_t^M - \tilde{X}_{t-1}^M$. \tilde{x}_t^M is a monthly variable observed only once every three months, and hence characterised by a systematic pattern of missing data. The approximation allows to write²

$$x_t^Q \simeq (1 + 2L + 3L^2 + 2L^3 + L^4)\tilde{x}_t^M. \quad (4)$$

Combining Eq. (4) with Eq. (1) yields:

$$x_t^Q \simeq (1 + 2L + 3L^2 + 2L^3 + L^4)(\mathbf{\Lambda}_Q f_t + \tilde{\zeta}_t^M) \quad (5)$$

$$= (\mathbf{\Lambda}_Q + 2\mathbf{\Lambda}_Q L + 3\mathbf{\Lambda}_Q L^2 + 2\mathbf{\Lambda}_Q L^3 + \mathbf{\Lambda}_Q L^4) f_t + (1 + 2L + 3L^2 + 2L^3 + L^4)\tilde{\zeta}_t^M, \quad (6)$$

hence,

$$\underset{(n \times 1)}{x_t} \equiv \begin{pmatrix} x_t^M \\ x_t^Q \end{pmatrix} \simeq \begin{pmatrix} \mathbf{\Lambda}_M & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Lambda}_Q & 2\mathbf{\Lambda}_Q & 3\mathbf{\Lambda}_Q & 2\mathbf{\Lambda}_Q & \mathbf{\Lambda}_Q \end{pmatrix} F_t + \begin{pmatrix} \zeta_t^M \\ \zeta_t^Q \end{pmatrix}, \quad (7)$$

where $F_t = (f_t', f_{t-1}', \dots, f_{t-4}')'$, $\mathbf{\Lambda}_M$ and $\mathbf{\Lambda}_Q$ are of dimensions $n_M \times r$ and $n_Q \times r$ respectively, and $\mathbf{0}$ denotes an $n_M \times r$ matrix of zeros. Finally, $\zeta_t^Q \equiv R(L)\tilde{\zeta}_t^M$, where $R(L) \equiv 1 + 2L + 3L^2 + 2L^3 + L^4$ and $\tilde{\zeta}_t^M$ denotes the idiosyncratic component of \tilde{x}_t^M . As before, restrictions can be imposed on $\mathbf{\Lambda}_Q$ such that global and group-specific factors are defined (see [Bańbura](#)

² $x_t^Q \equiv X_t^Q - X_{t-3}^Q = (1 - L^3)X_t^Q \simeq (1 - L^3)(\tilde{X}_t^M + \tilde{X}_{t-1}^M + \tilde{X}_{t-2}^M) = (1 - L^3)(1 + L + L^2)\tilde{X}_t^M = (1 - L)(1 + L + L^2)^2\tilde{X}_t^M = (1 + L + L^2)^2\tilde{x}_t^M$, see [Mariano and Murasawa \(2003\)](#).

et al., 2013).

Stacking the vectors $x_t \forall t = 1, \dots, T$ yields an unbalanced monthly panel. Variables may have a different start date, may contain systematically missing values as in the case of the quarterly indicators, and they may display the ‘ragged-edge’ that is typical of real-time data vintages whose entries are not released in a synchronous manner. All these features of the data can be broadly characterised by allowing for an arbitrary pattern of missing data in each vintage of x_t , and efficiently dealt with by estimating the model in Eqs. (2 - 3) and (7) using the algorithm in Bańbura and Modugno (2014). Prior to estimation, variables are transformed to achieve stationarity (see Table 2 for details on transformations) and standardised. We report details on the estimation in Appendix A.

2.2 Augmenting the DFM with Subsequent GDP Releases: the RA-DFM

In what follows, we detail the characteristics of our modelling approach. We focus specifically on GDP, however the framework is seamlessly extended to any quarterly variable for which the statistical office publishes updated estimates following a monthly schedule, such as GDP components as aggregate consumption and investment.

In most advanced economies, statistical offices release a first estimate of GDP based on a partial coverage of the economy a few weeks after the end of the relevant quarter. This first and incomplete number is then updated as more data are collected, and improved estimates are released in the following months. We think of these successive estimates of GDP for the same quarter as separate but correlated observables. This allows us to exploit the intrinsic factor structure of the official estimates (i.e. they are all estimates of the same object), and to essentially switch forecast target in line with the publication calendar of the statistical office. Intuitively, this is achieved by augmenting Eq. (7) with a set of monthly releases of GDP growth for the same reference quarter, and hence assuming that they have a factor structure – the same as in Eqs. (1 - 3).

Formally, let y_t denote quarterly GDP growth, and let $y_t^{(k)}$ denote the k^{th} monthly update for the estimate of y_t released by the statistical office, such that for every quarter $y_t^{(1)}$ denotes the first release published one month after the end of the quarter, $y_t^{(2)}$ denotes

the second release published two months after the reference quarter, and so on.³

We augment Eq. (7) with the time series for the first $k = 1, \dots, n_K$ GDP releases as follows:

$$x_t \equiv \begin{pmatrix} x_t^M \\ x_t^Q \\ y_t^{(1)} \\ \vdots \\ y_t^{(n_K)} \end{pmatrix}_{(n \times 1)} \simeq \begin{pmatrix} \Lambda_M & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \Lambda_Q & 2\Lambda_Q & 3\Lambda_Q & 2\Lambda_Q & \Lambda_Q \\ \Lambda^{(1)} & 2\Lambda^{(1)} & 3\Lambda^{(1)} & 2\Lambda^{(1)} & \Lambda^{(1)} \\ \vdots & & \vdots & & \vdots \\ \Lambda^{(n_K)} & 2\Lambda^{(n_K)} & 3\Lambda^{(n_K)} & 2\Lambda^{(n_K)} & \Lambda^{(n_K)} \end{pmatrix} F_t + \begin{pmatrix} \zeta_t^M \\ \zeta_t^Q \\ \varepsilon_t^{(1)} \\ \vdots \\ \varepsilon_t^{(n_K)} \end{pmatrix}, \quad (8)$$

where we use $\Lambda^{(k)}$ and $\varepsilon_t^{(k)}$ to denote, respectively, the loadings and the idiosyncratic component of each $y_t^{(k)}$, $k = 1, \dots, n_K$. The framework is general enough to include an arbitrary number of subsequent GDP releases, however, in our empirical application we choose to focus only on the first four release rounds (i.e. $n_K = 4$) for a number of reasons. First, the first revision rounds are those more likely to be originated by the availability of new information relative to the reference quarter. Second, and related, these are those typically regarded as being ‘market movers’ (see e.g. Bloomberg/Econday economic calendars). Third, the publication of the fourth update ($y_t^{(4)}$) typically coincides with the publication of the first release relative to the following quarter ($y_{t+1}^{(1)}$) – hence, for what concerns real-time nowcasting, at that point in time the relevant target switches to y_{t+1} , and subsequent revisions to past quarters become less relevant.

As is the case for the monthly and other quarterly variables, we allow the idiosyncratic components of GDP releases to display some degree of autocorrelation (i.e. serial correlation within the same vintage – $\mathbb{E}[\varepsilon_t^{(k)} \varepsilon_{t-\tau}^{(k)'}] \neq 0$, $\tau \neq 0$). However, because successive GDP releases for the same quarter are improved estimates of the same object, the elements of the idiosyncratic vector $\varepsilon_t = (\varepsilon_t^{(1)}, \dots, \varepsilon_t^{(n_K)})'$ may also be correlated across vintages – i.e. $\mathbb{E}[\varepsilon_t \varepsilon_t'] \neq 0$. If GDP-specific information is correlated across releases, systematic correlation may persist beyond that accounted for by the common factors (i.e. correlation across all variables in the same data vintage, $\mathbb{E}[x_t x_t']$). To account for these features, we

³The timing of the publication can vary across countries, but it is typically the case that the statistical offices publish monthly releases for quarterly variables.

model the vector of GDP idiosyncratic terms as a VAR(1),

$$\begin{matrix} \varepsilon_t \\ (n_K \times 1) \end{matrix} = \mathbf{\Phi} \varepsilon_{t-1} + \nu_t \quad \nu_t \sim \mathcal{N}(0, \mathbf{\Gamma}), \quad (9)$$

where $\mathbf{\Gamma}$ is a full matrix that enables the presence of a common component across the GDP releases. This common component can be given a news interpretation in the sense of [Jacobs and van Norden \(2011\)](#): if revisions are due to the incorporation of new information over time, the element of news in the earlier releases, $y_t^{(k)}$, will persist in the later releases, $y_t^{(k+1)}$. This generates contemporaneous correlation among subsequent GDP releases – $\mathbb{E}[\varepsilon_t \varepsilon_t']$ – that is captured in our framework via $\mathbf{\Gamma}$. The innovations ν_t can be thought of as information incorporated at each individual GDP release.

We call the model in Eqs. (2, 3, 8, 9) the Release-Augmented DFM, or RA-DFM.

Lastly, we define the k^{th} revision as

$$rev_t^{(k)} = y_t^{(k+1)} - y_t^{(k)}, \quad (10)$$

that is, in terms of the difference between consecutive publications for the same reference period.⁴ Combining Eq. (10) with Eq. (8) implies the following stochastic process for the GDP revisions:

$$rev_t^{(k)} = [\bar{\mathbf{\Lambda}}^{(k+1)} - \bar{\mathbf{\Lambda}}^{(k)}] F_t + [\varepsilon_t^{(k+1)} - \varepsilon_t^{(k)}], \quad (11)$$

where $\bar{\mathbf{\Lambda}}^{(k)} \equiv [\mathbf{\Lambda}^{(k)} \ 2\mathbf{\Lambda}^{(k)} \ 3\mathbf{\Lambda}^{(k)} \ 2\mathbf{\Lambda}^{(k)} \ \mathbf{\Lambda}^{(k)}]$, $\forall k$. The term $[\bar{\mathbf{\Lambda}}^{(k+1)} - \bar{\mathbf{\Lambda}}^{(k)}] F_t$ captures GDP revisions caused by information conveyed by the common factors, that is, information in all the vector of variables x_t . If between $y_t^{(k)}$ and $y_t^{(k+1)}$ information about either x_t^M or x_t^Q is published, this will affect the estimate of F_t (i.e. the conditional expectation of F_t given the information set that includes the release of x_t^M or x_t^Q), as we detail in the next subsection. Conversely, any information relative to either y_t or its lags that is released prior to $y_t^{(k)}$ will trigger an update in the estimate of the conditional expectation of both terms of Eq. (11). Predictability that instead arises from the autocorrelation of the revision process – $\mathbb{E}[rev_t^{(k)} rev_{t-\tau}^{(k)'}]$ – will be captured by the autoregressive structure

⁴Alternatively, we could adopt the convention whereby $rev_t^{(k)} = y_t^{(k+1)} - y_t^{(1)}$, that measures the revision always relative to the first release. In practice, for our purpose, the two definitions are equivalent.

of both F_t and $\varepsilon_t^{(k)}$, $\forall k$. Random noise in $rev_t^{(k)}$ will instead be a function of ν_t (see Eq. 9). In this sense, the model-implied revision process in Eq. (11) can in principle accommodate both ‘noise’ and ‘news’ revisions (see e.g. Mankiw and Shapiro, 1986; Faust et al., 2005). We return to this point in the next subsection. Cunningham et al. (2012) assume that the variability of each successive revision declines with the data maturity k . We do not impose such restrictions, but empirically, depending on the values of $\Lambda^{(k)}$ $\forall k$ and of the idiosyncratic errors variances, we could have that $\text{Var}(rev_t^{(k+1)}) \leq \text{Var}(rev_t^{(k)})$, $\forall k$.

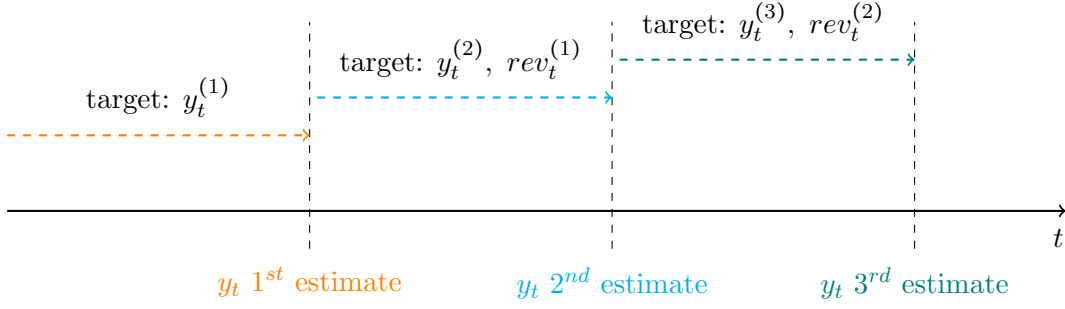
2.3 Real-Time Forecasts within the RA-DFM: Forecast Targets, Revisions, and the Role of Data News

Statistical nowcasting models are built on the intuition that other data that correlate with GDP growth but are released in a more timely fashion, and also sampled at higher frequencies, can help form an early assessment of current economic conditions before the official GDP numbers are published. The RA-DFM extends this concept also to revisions to early GDP estimates.

Time-Varying Forecast Targets Traditional models such as e.g. Bańbura et al. (2013) have a unique forecast target, typically either the first GDP release, or the latest available vintage. Conversely, the RA-DFM allows to switch forecast targets as time goes by, and in accordance with the publication calendar of the statistical agency. Prior to the publication of any official GDP data for the reference quarter, the target is the first estimate of GDP – $y_t^{(1)}$. Once this number is published, the target shifts to the second estimate for the same reference quarter – $y_t^{(2)}$. Note that conditional on having observed $y_t^{(1)}$, targeting $y_t^{(2)}$ is the same as targeting the first revision round $rev_t^{(1)}$. With the publication of $y_t^{(2)}$, the target moves to be $y_t^{(3)}$ (or, equivalently, $rev_t^{(2)}$) and so on. We sketch the intuition in Figure 1.

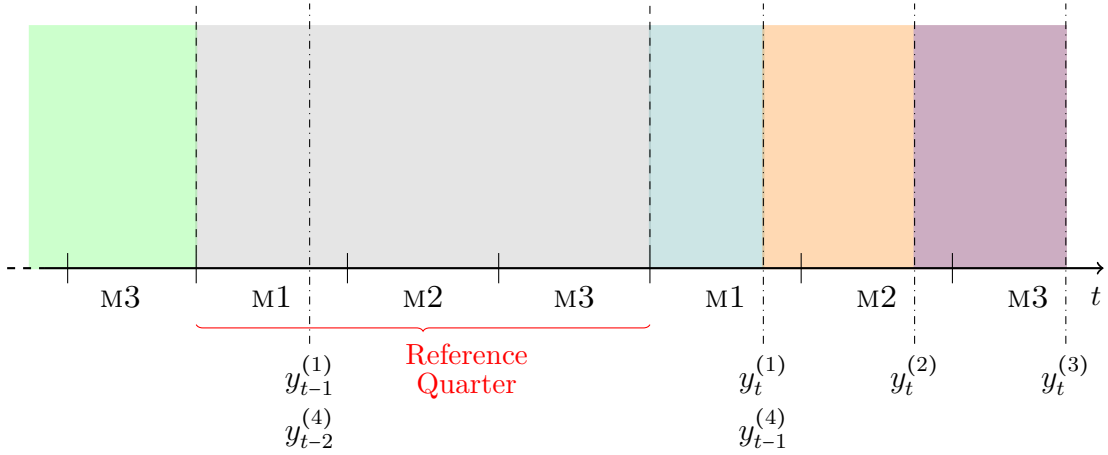
Let Ω_v denote the information set in a data vintage v – i.e. a snapshot of x_t at a particular date. We use the following convention for the forecasts horizons, summarised in Figure 2. If the timing of v falls within the current quarter (grey area), conditional on Ω_v we nowcast $y_t^{(1)}$ and forecast $y_{t+1}^{(1)}$. Similarly, if it falls within the first month following

FIGURE 1: TIME-VARYING FORECAST TARGETS



Note: Example of time-varying forecast targets for the RA-DFM.

FIGURE 2: TRACKING WINDOW



Note: The figure sketches the convention for the forecast horizons in the RA-DFM. The green area is the forecast period, where the forecast horizon, expressed in quarters, is $h = 1$. The grey area is the nowcast period, with $h = 0$. The backcast period is the sum of the teal, orange, and purple areas ($h = -1$).

the reference quarter, but still before the publication of the first release (teal area), we backcast $y_t^{(1)}$, and nowcast $y_{t+1}^{(1)}$. Once $y_t^{(1)}$ is released (orange area), we drop it from the set of active targets and substitute with $y_t^{(2)}$, which is then further substituted with $y_t^{(3)}$ once the release for $y_t^{(2)}$ is out (purple area), as in Figure 1. We track each quarter for a total of roughly 270 days. The first 90 are a pure forecast; i.e. the forecast horizon, expressed in quarters, is $h = 1$ (green area). The second set of 90 days corresponds to the nowcast period ($h = 0$, grey area), and we generally refer to the backcast period as the sum of the teal, orange, and purple areas ($h = -1$).

The Role of Data News We use the RA-DFM to forecast GDP growth in a data-

rich environment in real-time. In doing so, we mimic as closely as possible the behaviour of a professional forecaster and use at each forecast origin only information that was effectively available at each point in time, taking into account the release schedule and publication delay of each piece of data. In order to address the contribution of all the different data to GDP forecasts updates and to updates to the uncertainty around the forecasts, we first rewrite the RA-DFM model in its state-space form

$$x_t = \mathbf{C}s_t + e_t \quad e_t \sim \mathcal{N}(0, \mathbf{R}), \quad (12)$$

$$s_t = \mathbf{A}s_{t-1} + u_t \quad u_t \sim \mathcal{N}(0, \mathbf{Q}), \quad (13)$$

where x_t is defined as in Eq. (8), and the vector of unobserved states is $s_t \equiv (f'_t, \dots, f'_{t-4}, \zeta_t^{M'}, \zeta_t^{Q'}, \dots, \zeta_{t-4}^{Q'}, \varepsilon_t, \dots, \varepsilon_{t-4})'$.⁵ Let Ω_{v-1} and Ω_v denote the information set in two consecutive x_t vintages. Note that consecutive data vintages are not generally equally spaced, i.e. they may be a few hours apart in the case of data being released at two different times within the same day, or they may be days or even weeks apart depending on the characteristics of the release calendar. With real-time data, $\Omega_v \setminus \Omega_{v-1}$ will contain first releases of some of the variables in x_t , and revisions to older data, such that $\Omega_{v-1} \not\subseteq \Omega_v$.

For simplicity, consider the case in which only one variable x_τ^* is released between Ω_{v-1} and Ω_v . Between consecutive vintages, the forecast for $y_t^{(k)}$ is updated as follows

$$\mathbb{E}\left[y_t^{(k)} \mid \Omega_v\right] - \mathbb{E}\left[y_t^{(k)} \mid \Omega_{v-1}\right] = \mathbb{E}\left[y_t^{(k)} \mid I_v\right] + \mathbb{E}\left[y_t^{(k)} \mid O_v\right]. \quad (14)$$

In Eq. (14), $I_v \equiv x_\tau^* - \mathbb{E}(x_\tau^* \mid \Omega_{v-1})$. Using the terminology in Bańbura and Modugno (2014), we refer to I_v as *data news*, or the news component in the release of x_τ^* . That is, I_v constitutes an element of surprise with respect to the model's forecast $\mathbb{E}(x_\tau^* \mid \Omega_{v-1})$, i.e. it is an innovation with respect to Ω_{v-1} . O_v contains instead revisions to past data that are released together with x_τ^* and, potentially, a term which relates to the correlation among these. In the analysis of the contribution of data news to the forecast updates we focus on the terms that condition on I_v and disregard O_v .

⁵All the details on the state-space representation in Eqs. (12 - 13) are reported in Appendix A.

Using the properties of the conditional expectation, we further obtain

$$\mathbb{E}\left[y_t^{(k)} \mid I_v\right] = \mathbb{E}\left[y_t^{(k)} I_v'\right] \mathbb{E}\left[I_v I_v'\right]^{-1} I_v. \quad (15)$$

The elements in Eq. (15) are obtained from the Kalman smoother,

$$\mathbb{E}\left[y_t^{(k)} I_v'\right] = \mathbf{C}_{(k)} \mathbb{E}\left[(s_t - \mathbb{E}[s_t \mid \Omega_v]) (s_\tau - \mathbb{E}[s_\tau \mid \Omega_v])'\right] \mathbf{C}'_{x^*}, \quad (16)$$

$$\mathbb{E}\left[I_v I_v'\right] = \mathbf{C}_{x^*} \mathbb{E}\left[(s_\tau - \mathbb{E}[s_\tau \mid \Omega_v]) (s_\tau - \mathbb{E}[s_\tau \mid \Omega_v])'\right] \mathbf{C}'_{x^*} + \mathbf{R}_{y^{(k)}, x^*}, \quad (17)$$

where \mathbf{C}_j denotes the j -th row of \mathbf{C} , and we refer to $\mathbb{E}\left[y_t^{(k)} I_v'\right] \mathbb{E}\left[I_v I_v'\right]^{-1}$ as the news weights. Similarly, we can look at the contribution of data releases in the update of the uncertainty around the forecast for $y_t^{(k)}$. Specifically,

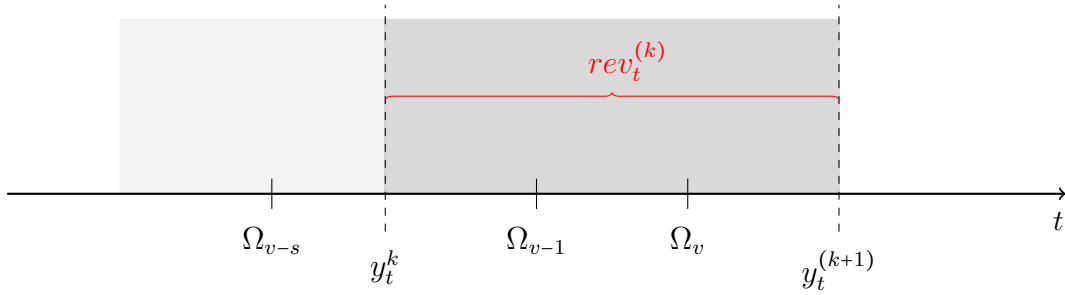
$$\mathbb{E}\left[y_t^{(k)} y_t^{(k)'} \mid \Omega_v\right] - \mathbb{E}\left[y_t^{(k)} y_t^{(k)'} \mid \Omega_{v-1}\right] = \mathbb{E}\left[y_t^{(k)} y_t^{(k)'} \mid I_v\right] + \mathbb{E}\left[y_t^{(k)} y_t^{(k)'} \mid O_v\right]. \quad (18)$$

News and Noise in GDP Revisions Conditional on having observed $y_t^{(k)}$, and extending the argument in [Bańbura and Modugno \(2014\)](#), we can decompose the forecast update for $rev_t^{(k)}$ with a straightforward generalisation of Eq. (14)

$$\begin{aligned} \mathbb{E}\left[rev_t^{(k)} \mid \Omega_v, y_t^{(k)}\right] - \mathbb{E}\left[rev_t^{(k)} \mid \Omega_{v-1}, y_t^{(k)}\right] &= \mathbb{E}\left[y_t^{(k+1)} \mid \Omega_v, y_t^{(k)}\right] - \mathbb{E}\left[y_t^{(k+1)} \mid \Omega_{v-1}, y_t^{(k)}\right] \\ &= \mathbb{E}\left[y_t^{(k+1)} \mid I_v, y_t^{(k)}\right] + \mathbb{E}\left[y_t^{(k+1)} \mid O_v, y_t^{(k)}\right]. \end{aligned} \quad (19)$$

Consider now the case of news and noise in GDP revisions, as defined by [Mankiw and Shapiro \(1986\)](#), and the timing of events summarised in [Figure 3](#). A ‘pure news’ $rev_t^{(k)}$ is an innovation with respect to the information set at the time of the release of $y_t^{(k)}$ (pale grey area in the figure). Hence, it is triggered purely by information released between $y_t^{(k)}$ and $y_t^{(k+1)}$ (dark grey area in the figure). A ‘pure noise’ $rev_t^{(k)}$, on the other hand, is a random measurement error that is uncorrelated with the ‘true’ value of GDP growth, and hence with any information released between $y_t^{(k)}$ and $y_t^{(k+1)}$. As discussed in the previous section, only GDP releases move the vector ε_t and, by construction, there are no GDP releases between $y_t^{(k)}$ and $y_t^{(k+1)}$. Hence, in our real-time nowcasting environment,

FIGURE 3: NEWS AND NOISE IN GDP REVISIONS



Note: News and Noise GDP Revisions in the RA-DFM.

we account for the contribution of data news in forecasting $rev_t^{(k)}$, insofar as $rev_t^{(k)}$ has at least some element of news.

3 A Real-Time Mixed-Frequency Dataset for the UK Economy

In this section, we describe the real-time dataset we compiled for nowcasting UK GDP growth. We start by describing GDP data releases and revisions, then we move to summarise our novel real-time data set of monthly and quarterly economic and financial indicators, and we end by overviewing the real-time data flow in the UK.

3.1 UK GDP: Releases and Revisions

The UK’s Office of National Statistics (ONS) measures the Gross Domestic Product of the total national economic activity in the UK in three different ways: the output approach, the expenditure approach and the income approach. GDP estimates are produced quarterly and annually, and there are three publication stages for the quarterly estimates: the Preliminary Estimate, the Second Estimate, and the UK Quarterly National Accounts.⁶ In its current schedule (May 2018), the ONS’s preliminary GDP estimate is published 4 weeks after the end of the reference quarter.⁷ This first estimate is based solely on

⁶<https://www.ons.gov.uk/economy/grossdomesticproductgdp/methodologies/grossdomesticproductgdpqmi>

⁷Since July 2018, the ONS has introduced the publication of a monthly GDP estimate, while the first GDP release is now delayed by another 10 days. Our model can easily be generalised to formally link

TABLE 1: SUMMARY STATISTICS FOR GDP RELEASES AND IMPLIED REVISIONS

RELEASES				
	First	Second	Third	Fourth
Mean	0.359	0.366	0.367	0.369
Stdev	0.560	0.559	0.582	0.599
AC(1)	0.608	0.616	0.619	0.631
REVISIONS				
	–	First	Second	Third
Mean	–	0.008	0.008	0.010
Stdev	–	0.095	0.131	0.153
AC(1)	–	-0.204	-0.107	0.012
Q(4)	–	6.952	3.988	1.896
p-values	–	[0.138]	[0.408]	[0.755]
SNR	–	0.971	0.950	0.935

Note: Revisions are defined with respect to the first release. AC(1) is the first order autocorrelation coefficients. Q(4) denotes the Lyung-Box Q(4) test for a serial correlation of order 4 with p-values reported in square brackets. $SNR = 1 - \left[\text{Var} \left(y_t^{(k+1)} - y_t^{(1)} \right) / \text{Var} \left(y_t^{(k+1)} \right) \right]$.

44% of actual output data, thus leaving a large scope for projection and imputations. In contrast to the first estimate, the second estimate, published 8 weeks after the reference quarter, is based on information from all three approaches. This second estimate covers 80% of the required data for the output approach, and 50% to 60% of the data for the income and expenditure approaches. The UK quarterly national accounts (QNA) are then published 13 weeks after the end of the reference quarter, and are based on 90% of the data for the output approach, 70% for the expenditure approach, and 70% for the income approach. Hence, in each given quarter, and for each given quarter, there are three monthly ONS releases of GDP numbers. These first rounds of revisions are primarily due to the inclusion of information that was not available at the time of the preliminary estimate.

Following these initial revision rounds, annual revisions of GDP estimates are published either in June or September every year as part of the Blue Book publication.⁸ During the first three years after the reference quarter, a further source of annual revision

monthly to quarterly GDP releases, following the intuition in [Bragoli and Modugno \(2017\)](#)

⁸<https://www.ons.gov.uk/economy/grossdomesticproductgdp/compendium/unitedkingdomnationalaccountsthebluebook/2017>

sions is due to the balancing of real GDP computed with the income and expenditure approaches. After three years, GDP numbers computed under the two approaches are matched to yield the same value in (chained) British Pounds.⁹

The first four monthly releases of real UK GDP growth for all quarters since 1990 are charted in Figure 4, together with the latest available vintage at the time of writing dating May 2018. Table 1 reports summary statistics for both the first four GDP releases and the implied revisions, computed here always with respect to the first estimate. The table reports the sample mean, standard deviation and the first order serial correlation for all quarters between 1990-Q1 and 2016-Q4. For the revisions, we also report the Lyung-BOx Q(4) test for 4th order serial correlation, and a measure of the signal-to-noise ratio (SNR) computed as $SNR = 1 - \left[\text{Var} \left(y_t^{(k+1)} - y_t^{(1)} \right) / \text{Var} \left(y_t^{(k+1)} \right) \right]$. A SNR near 1 implies that the noisiness of the revision is small.

The numbers reported in the table suggest that initial monthly revisions raise the sample mean and standard deviation of GDP growth, while there is little evidence of serial correlation. The increase of about 5% in the standard deviation is compatible with news revisions, and in accord with results in Galvão and Mitchell (2018).¹⁰ These initial monthly revisions are sizeable. The third revision’s standard deviation is 1/3 of the standard deviation of the preliminary estimate.¹¹

3.2 The Real-Time Mixed-Frequency Dataset

We assembled a mixed-frequency dataset counting 8 quarterly and 25 monthly indicators.¹² These variables are listed in Table 2, and are grouped into six types: economic

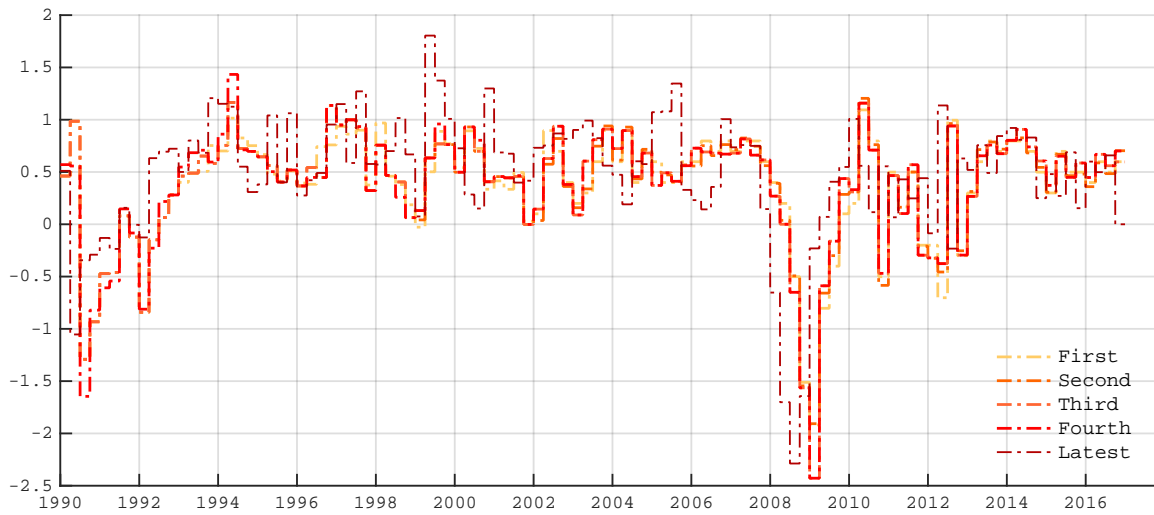
⁹This is in contrast with US GDP growth data revisions where differences between GDP (expenditure) and GDI estimates persist (see e.g. Aruoba et al., 2016).

¹⁰Note that Galvão and Mitchell (2018) provide statistically significant evidence that data revisions raise the year-on-year GDP growth mean for data from 1993 to 2013, we, however, find that this is not a consequence of the initial monthly revisions since Table 1 suggests that the average revision is nearly zero.

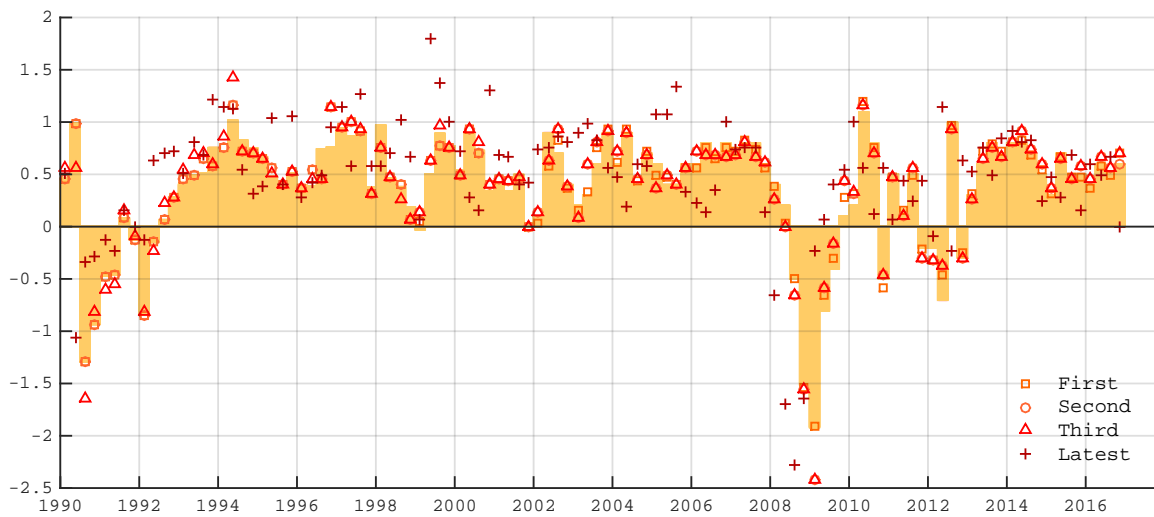
¹¹In contrast, Clements and Galvão (2017) indicate that US initial monthly revisions account for a larger 3/5 ratio of the initial release standard deviation. This suggests that UK GDP initial revisions are in general smaller than the equivalent US values, and motivates future research with the aim to apply the RA-DFM model to US data.

¹²The number of monthly indicators that we consider is comparable to that in Bańbura et al. (2013), but smaller than that in other applications such as Artis et al. (2005).

FIGURE 4: UK GDP INITIAL MONTHLY RELEASES



(A) Note: UK GDP Monthly Releases.



(B) Note: UK GDP First Release (yellow bar) and subsequent releases (markers).

TABLE 2: REAL-TIME DATASET FOR THE UK ECONOMY

	CODE	VARIABLE NAME	SOURCE	TYPE	FREQ	REVISED	TRANSF
1	GDP1	Gross Domestic Product, 1st rel	ONS	Activity	Q	✓	LD
2	GDP2	Gross Domestic Product, 2nd rel	ONS	Activity	Q	✓	LD
3	GDP3	Gross Domestic Product, 3rd rel	ONS	Activity	Q	✓	LD
4	GDP4	Gross Domestic Product, 4th rel	ONS	Activity	Q	✓	LD
5	QCONSTR	Quarterly Construction Output	ONS	Activity	Q	✓	LD
6	CONS	Private Consumption	ONS	Activity	Q	✓	LD
6	INV	Total Business Investment	ONS	Activity	Q	✓	LD
8	HINV	Housing Investment	ONS	Activity	Q	✓	LD
9	IOP	Industrial Production	ONS	Activity	M	✓	LD
10	MPROD	Manufacturing Production	ONS	Activity	M	✓	LD
12	IOS	Index of Services	ONS	Activity	M	✓	LD
13	BOPEXP	BOP Total Exports (Goods)	ONS	Trade	M	✓	LD
14	BOPIMP	BOP Total Imports (Goods)	ONS	Trade	M	✓	LD
15	RSI	Retail Sales Index	ONS	Trade	M	✓	LD
16	CCOUNTR	Claimant Count Rate	ONS	Labour	M		D
17	LFSE	LFS Number of Employees	ONS	Labour	M		LD
18	LFSU	LFS Unemployment Rate	ONS	Labour	M		D
18	MTGAPP	Mortgages Approved	BOE	Credit	M		LD
19	CREDIT	Net Consumer Credit	BOE	Credit	M		LD
20	UKBASKET	UK Focused Equity Index	BOE	Fin'l	M		LD
21	SERI	Sterling Effective Exchange Rate	LSE	Fin'l	M		LD
22	TERMSP	Term Spread	BOE	Fin'l	M		D
23	CORPSP	Corporate Bond Spread	ML	Fin'l	M		D
24	PMIM	PMI Manufacturing	IHS Markit	Survey	M		L
25	CIPSEM	CIPS-E-Manufacturing	IHS Markit	Survey	M		L
26	PMIC	PMI Construction	IHS Markit	Survey	M		L
27	CIPSEC	CIPS-E-Construction	IHS Markit	Survey	M		L
28	PMIS	PMI Services	IHS Markit	Survey	M		L
29	CIPSES	CIPS-E-Services	IHS Markit	Survey	M		L
30	CBIORDER	CBI Industrial Trends	CBI	Survey	M		L
31	CBISALE	CBI Distributive Trade	CBI	Survey	M		L
32	LLOYBB	Lloyds Business Barometer	Lloyds	Survey	M		L
33	ASCORE	Agents' Scores	BOE	Survey	M		L

Note: Sources are the Office for National Statistics (ONS), the Bank of England (BOE), Bank of America Merrill Lynch (ML), IHS Markit/CIPS, the Confederation of British Industries (CBI), Lloyds Bank. Revisions in survey data occur primarily due to rebasing and are hence treated as unrevised. Transformation codes: LD = log difference, L = levels, D = first difference.

activity, labour market statistics, business and output surveys, credit and financial data.¹³

The data span the years from 1990 to 2016 with full real-time vintages since 2006.

Quarterly variables include the first four GDP releases and some of the output and expenditure components of GDP: Construction, Business Investment, Housing Investment and Private Consumption. These are all revised following a monthly schedule. The

¹³While we collected real time data also for price variables, such as the Consumer and Retail Price Indices, and price indices for both imports and exports, we do not consider inflation measures in our benchmark model. The UK measures of inflation are published by the ONS with no seasonal adjustment; this implies that in a real-time nowcasting exercise, we need to either seasonally adjusted the data each time a new vintage becomes available, or use year-on-year growth rates. We attempt the latter approach and preliminary results (available on request) suggest the inclusion of inflation measures deteriorate the forecasting performance of the model due to the lagging nature of year-on-year growth rates. This is consistent with findings in e.g. [Giannone et al. \(2008\)](#).

monthly indicators are intended to provide additional information related to economic activity in more timely fashion, and the choice is based on their relevance for policy makers, statistical agencies, and market participants. The variables populating the top portion of Table 2 are all subject to revision. In order to construct real-time vintages for these variables we relied on the archives of the Bank of England, where vintages of data released by the ONS data have been carefully stored over the years. These data are available in their original release unit, and we were able to reconstruct real-time vintages for these variables from 2006-Q4.¹⁴ Other monthly indicators such as surveys, prices and labour market statistics can get lightly revised. These revisions are almost exclusively due to re-basing and/ or changes in measurements or seasonal adjustment rather than to the addition of extra information. For these variables, we construct real-time vintages by starting from the latest available vintage available at the time of the assembly of the dataset (July 2017), and work backward using the release calendar of each of these data. The same procedure is used for credit and financial market variables that are also not revised. Asset prices enter the dataset in monthly averages. We also include in our dataset the Agents' Score, a survey compiled by the regional agents of the Bank of England. This survey is based on questions relative to both current and expected economic conditions that the regional agents ask firms and businesses during their regular visits, and its timing is tied to the Bank of England's monetary policy cycle.¹⁵

To the best of our knowledge, ours is the most comprehensive real-time mixed-frequency dataset for the UK economy in terms of breadth and coverage. The Bank of England maintains a real-time database that only covers quarterly variables. This dataset is updated regularly with the publication of the ONS Blue Book, and details on its construction are in [Castle and Ellis \(2002\)](#). For some of the variables in Table 2, such as the Index of Industrial Production and Retail Sales, our dataset extends previous work of [Egginton et al. \(2002\)](#); this dataset, however, was only last updated at the end of 1999.

TABLE 3: THE UK DATA FLOW

RELEASE DAY	CODE	VARIABLE NAME	FREQUENCY	REFERENCE PERIOD	PUBLICATION DELAY
1	PMIM	PMI Manufacturing	M	m-1	1
1	CIPSEM	CIPS-E-Manufacturing	M	m+3	-89
3	PMIC	PMI Construction	M	m-1	3
3	CIPSEC	CIPS-E-Construction	M	m+3	-87
5	PMIS	PMI Services	M	m-1	5
5	CIPSES	CIPS-E-Services	M	m+3	-85
9	IOP	Industrial Production	M	m-2	39
9	MPROD	Manufacturing Production	M	m-2	39
10	BOPEXP	BOP Total Exports (Goods)	M	m-2	40
10	BOPIMP	BOP Total Imports (Goods)	M	m-2	40
17	CCOUNTR	Claimant Count Rate	M	m-1	17
17	LFSE	LFS Number of Employees	M	m-2	47
17	LFSU	LFS Unemployment Rate	M	m-2	47
20	RSI	Retail Sales Index	M	m-1	20
20	CBIORDER	CBI Industrial Trends	M	m	-10
22	IOS	Index of Services	M	m-2	52
22	GDP	Either GDP1, GDP2 or GDP3	Q	q-1	22, 52, 82 [†]
22	QCONSTR	Construction Output	Q	q-1	22, 52, 82 [†]
22	CONS	Private Consumption	Q	q-1	52 and 82 [†]
22	INV	Total Business Investment	Q	q-1	52 and 82 [†]
22	HINV	Housing Investment	Q	q-1	52 and 82 [†]
23	CBISALE	CBI Distributive Trade	M	m	-7
26	LLOYBB	Lloyds Business Barometer	M	m	-4
30	ASCORE	Agents' Score	M	m	0
30	UKBASKET	UK Focused Equity Index	M	m	0
30	SERI	Sterling Effective Exchange Rate	M	m	0
30	TERMSP	Term Spread	M	m	0
30	CORPSP	Corporate Bond Spread	M	m	0
30	MTGAPP	Mortgages Approved	M	m-1	30
30	CREDIT	Net Consumer Credit	M	m-1	30

Note: The table sketches the data flow within a typical month. The first column is the average release day for each variable. Column five reports the reference period: m and q denote the current month and current quarter; hence, $m + 3$ refers to three months ahead, and $q - 1$ to the previous quarter. The last column reports the average publication delay (in days) from the end of the reference period. [†] The publication delay of quarterly variables varies depending on which month in the quarter is considered (i.e. 25 days for preliminary estimates, 85 days for the third estimate).

3.3 The UK Data Flow

In the context of real-time forecasting, addressing the timeliness and publication calendar of the different indicators is as important as assembling the relevant data. We recovered date and time of official data releases for all the variables in Table 2 combining information provided by the original data suppliers with the economic calendar of data releases distributed by Bloomberg. The latter is populated by all the ‘market movers’, most of which appear in our set.

The data flow within a typical month in the UK is summarised in Table 3. The first column reports the average publication day for each indicator, while in the fifth column we report the period the release refers to. These two pieces of information are combined in the last column of the table where we report the typical publication delay for each variable, expressed in days from the end of the reference period. For example, Manufacturing PMI is released on the first day of each month for the previous month. The publication delay in this case is only one day. Within the same release, Markit/IHS also publishes a forward looking index that summarises expectations relative to the next quarter (CIPSEM). Hence, in this case the publication delay is negative, as the number refers to the following 90 days.

Production and international trade data are published in the second week of each month, and refer to two months prior to the one in which they are published. E.g. the index of industrial production (IOP) for March is released on the 9th of May, with a publication delay of 39 days counting from the end of the reference month (Mar). Hence, in each quarter, the first production data relative to that quarter are only released in the third month of that quarter. This constitutes a considerable delay, particularly when compared to US production data which are published only two weeks after the reference period. A similarly long publication delay characterises labour market statistics. These get published by the ONS on the third week of every month. The timeliest labour market data are those relative to the Claimants Count: these count the number of unemployment benefits claims every month, and have the shortest publication delay. Conversely, the unemployment rate and employment data, part of the same release, have an extra 30

¹⁴Further details are reported in Table B.1 in the Appendix.

¹⁵<https://www.bankofengland.co.uk/about/people/agents>

days of delay. The scarcity of timely ‘hard’ data in the UK makes nowcasting the UK economy a much tougher exercise when compared to other countries, since most of the information at early stages of each quarter only come from either surveys or credit and financial markets data. The majority of UK surveys are releases towards the end of each month for the current month, with zero (negative) delay. Since we use monthly averages for asset prices we assume that they become available at the end of the month for the current month, similar to surveys. Contrary to the latter, we assume that their release time is the end of the trading day on the last day of the month.

Combining the real-time dataset with the calendar of data releases from 2006-Q4 yields over 1500 real-time mixed-frequency data vintages over which we evaluate the performance of the RA-DFM model in nowcasting UK economic activity.

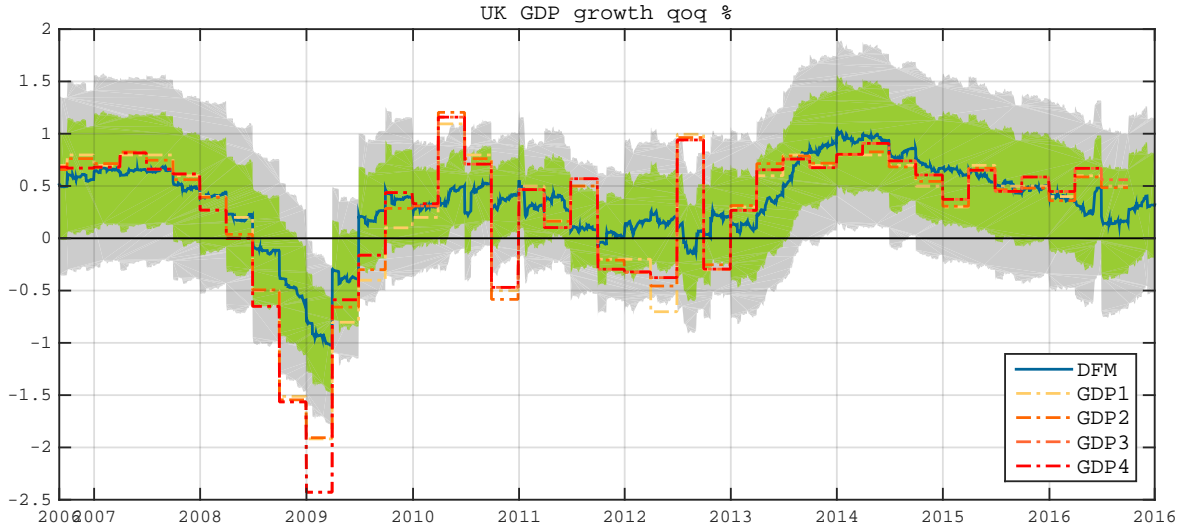
4 Nowcasting UK GDP Growth in Real-Time

In this section we apply the RA-DFM model developed in Section 2 to the novel dataset described in Section 3 to predict UK real quarterly GDP growth in real-time. We start by analysing the model’s average forecasting performance under a benchmark specification, and evaluate variations induced by different parametrisations in terms of both point and density forecasts (Sections 4.1 and 4.2). Next, we move to evaluate the contribution of data news in informing forecast updates for the revision process (Section 4.3). In Section 4.4 we evaluate the informativeness of the predictions of the benchmark specification for the latest available vintage of UK GDP growth. Section 4.5 compares the RA-DFM model against a selection of institutional forecasters. Finally, in Section 4.6 we zoom on two interesting episodes where uncertainty about the current state of the UK economy has been particularly relevant for policy analysis.

4.1 The benchmark RA-DFM model in Practice

The benchmark RA-DFM specification includes all the 25 monthly and 8 quarterly series listed in Table 2. Monthly variables enter the model either in levels (e.g. surveys), or in month-on-month growth rates. Quarterly variables all enter in quarter-on-quarter growth rates. We estimate the RA-DFM model with three factors (i.e. $r = 3$). The first factor

FIGURE 5: NOWCAST OF $y_t^{(1)}$ OVER THE EVALUATION SAMPLE



Note: Evolution of nowcast for $y_t^{(1)}$ over the evaluation sample (blue line) with one and two standard deviation forecast error bands (green and grey areas respectively). Dash-dotted lines are for subsequent GDP releases with darker shades of red used for more recent publications.

loads on all variables – this can be interpreted as a synthetic indicator for economic activity in the UK. The second factor only loads on ONS economic activity measures, that is, ‘hard data’ (top panel of Table 2). The third factor summarises information in surveys (bottom panel of Table 2). We set $p = 1$ in Eq. (2), model each ζ_t^M and ζ_t^Q as independent AR(1) processes, and the vector of GDP idiosyncratic terms ε_t as a VAR of order 1.

The parameters of the model – i.e. \mathbf{C} , \mathbf{R} , \mathbf{A} , \mathbf{Q} in Eqs. (12 - 13) –, are estimated using data since January 1992, and an expanding time window at the beginning of every new calendar year in the evaluation sample (2006-2016). Forecast updates at each forecast origin within the year are then computed using the parameters estimated at the beginning of the year. Real-time out-of-sample forecast, nowcast and backcasts (see Figure 2) for real quarter-on-quarter UK GDP growth are produced for the 10-year period between Q4 2006 and Q4 2016. As discussed in Section 3, combining the real-time dataset with the actual publication calendar for all the data between 2006 and 2016 delivers over 1,500 real-time data vintages over which the performance of the model is evaluated.

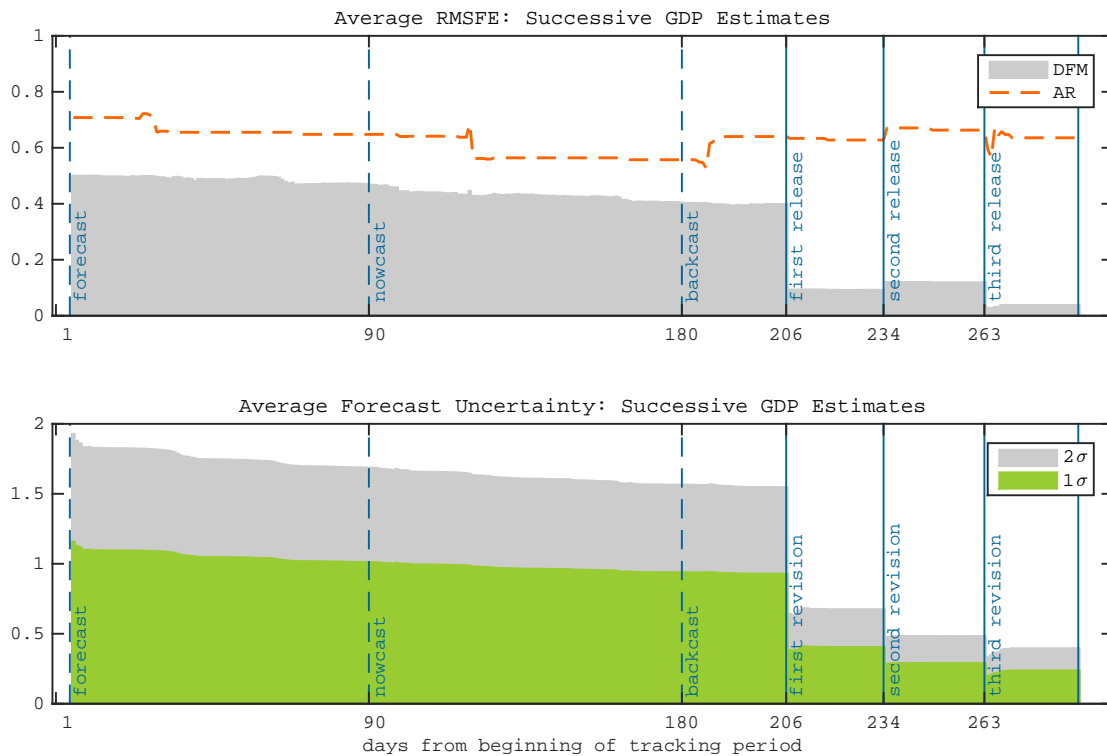
Figure 5 plots the evolution of the nowcast for the first GDP release – $\hat{y}_t^{(1)}$ – together with one and two standard deviation error bands against the first four GDP releases. In

the picture, darker shades of red refer to more mature estimates of UK GDP growth. The two standard deviation bands are equivalent to a 95% predictive interval for a Gaussian predictive density (which is an assumption compatible with our linear Gaussian state-space model). Because the out-of-sample period counts 41 quarters, we would expect the realisations to be outside the two standard deviation intervals two times. Figure 5 reveals that in fact GDP estimates lie outside the bands around 5 times (6 in case of the first estimate), and we discuss these cases below. From 2013 onwards, we observe that the uncertainty around the model's forecasts increases, but outturns tend to be very close to the RA-DFM predictions. The nowcasting uncertainty of the later period seems to be due to the large data uncertainty that characterised 2012, rather than the great recession.

The first episode in which outturns are outside the RA-DFM bands is the 2008/2009 recession. Here, while the nowcast is adjusted downward in a relatively timely fashion, the magnitude of the downturn is severely underestimated, even when compared to the first GDP numbers (yellow dash-dotted line). The remaining episodes are all related to idiosyncratic events that the model would not be able to capture, virtually by construction. The exceptionally low reading of Q4 2010, revised away in latest vintages, was thought to have been mostly induced by particularly adverse weather conditions. Similarly, the strong positive growth rate registered in Q3 2012 and both preceded and followed by negative-growth quarters is typically attributed to the joint effect of the Diamond Jubilee and the Olympic Games. In both these episodes, the construction sector is thought to have been the one responsible for the changes in GDP growth. Unfortunately, due to monthly construction data only starting in 2010, our data set only includes construction output at quarterly frequency. The slow movements in the quarterly variable, coupled with its publication delay, effectively limits the ability of the model to detect changes in the construction sector in a timely fashion. Another interesting episode is the first half of 2012. Here, and despite three consecutive negative-growth quarters according to official first estimates (yellow line), the model never predicts a recession. However, as we shall see, this recession too was eventually revised away. We discuss this particular episode in Section 4.6.

Figure 6 summarises the average (point) forecasting performance of the benchmark RA-DFM specification over the forecast, nowcast and backcast periods. The top panel

FIGURE 6: AVERAGE PERFORMANCE OVER THE EVALUATION SAMPLE



Note: Average RMSFE (top panel) and forecast uncertainty (bottom panel) over the evaluation sample. Numbers on the x axis denote days since the beginning of the tracking window for the average reference quarter.

reports the root mean squared forecast error (RMSFE), an *ex post* measure of uncertainty, while the bottom panel reports the average predicted uncertainty, an *ex ante* measure. These are both averages over the 41 quarters in the out-of-sample period. Numbers on the y axis are quarter-on-quarter percentage points. On the x axis we report the number of days since the beginning of the tracking window for each quarter (see Figure 2), such that the first 90 days are a pure forecast period, and the backcast starts after 180 days, at the end of the reference quarter, and before any of the GDP releases is published. The figure also includes the RMSFE for a naïve AR(1) estimated on the different vintages of GDP growth in line with their real-time availability. In both panels, the two uncertainty measures decline as the forecasting horizon shortens indicating the model’s efficiency in incorporating new data as they become available. Contrary to a simple AR which only uses GDP data, the model’s accuracy improves steadily over the nowcast and forecast periods as more data related to the reference quarter are released. This confirms the

importance of introducing correlated but more timely data in forecasting real growth. Furthermore, once the first GDP release is out, and due to the high correlation among subsequent releases, the forecast error for the later releases becomes virtually zero, on average, and conditional on the publication of the first figures. Similarly, conditional on the first release, the forecast uncertainty around the successive revisions is minimal.

4.2 RA-DFM Accuracy Across Model Specifications

We assess the accuracy of the RA-DFM across specification both in terms of point and density forecasts using average RMSFEs, and Logarithmic Score (LS)/Continuous Ranked Probability Score (CRPS) respectively. The numbers in the tables reported in this section are averages over the out-of-sample period (Q4 2006-Q4 2016). The LS is defined as the logarithm of the probability density function evaluated at the realisation. By minimising the negative of the LS, we would choose the model that on average gives the higher probability to the actual realisation, and minimises the KLIC distance between the model’s predictive density and the true unknown density. The CRPS on the other hand measures the average absolute distance between the model’s empirical cumulative distribution function (CDF) and the empirical CDF that is associated with the realisation. As in the case of the Mean Absolute Error, the CRPS is more robust to the presence of outliers, and a model with the smallest CRPS is the most accurate.

Table 4 compares the forecast accuracy of different model specifications at selected points in time in the tracking window (see Figure 2). RMSFE values for the benchmark specification are in column (a). All other columns report the RMSFEs of different specifications, and RMSFEs relative to benchmark (i.e. numbers smaller than 1 denote improvements relative to benchmark) are in the middle panel of the table. Table 5 and Table 6 replicate the comparison in terms of CRPS and LS respectively.

We consider different specifications. First, we change the number of lags in the factors VAR (Eq. 2) from 1 to 4 in column (b). Then in column (c) we replace the VAR specification in Eq. (9) with a factor that loads only on the vector y_t . This is an alternative specification that similarly allows for the presence of a common component specific to

TABLE 4: RMSFE ACROSS MODEL SPECIFICATIONS

POSITION IN TRACKING WINDOW	(a)	(b)	(c)	(d)	(e)	(f)	(g)
FORECAST STARTS	0.500	0.492	0.502	0.555	0.507	0.510	0.510
NOWCAST STARTS	0.465	0.455	0.468	0.502	0.467	0.472	0.469
BACKCAST STARTS	0.403	0.406	0.427	0.429	0.391	0.396	0.404
GDP1	0.399	0.405	0.424	0.418	0.421	0.422	0.429
GDP2	0.093	0.086	0.388	0.179	0.092	–	0.158
GDP3	0.119	0.116	0.406	0.117	0.124	–	0.158
GDP4	0.033	0.034	0.411	0.068	0.029	–	0.158
RELATIVE TO BENCHMARK (a)		(b)	(c)	(d)	(e)	(f)	(g)
	1	0.983	1.004	1.109	1.013	1.018	1.019
	1	0.978	1.006	1.079	1.003	1.014	1.008
	1	1.006	1.060	1.064	0.970	0.982	1.002
	1	1.014	1.062	1.048	1.054	1.058	1.074
	1	0.929	4.198	1.934	0.998	–	1.710
	1	0.970	3.405	0.980	1.036	–	1.321
	1	1.015	12.370	2.033	0.877	–	4.753
GDP TARGETS	4	4	4	4	4	1	4 [†]
FACTORS VAR LAGS	1	4	1	1	1	1	1
VAR IDIOSYNCRATIC	✓	✓		✓	✓	✓	✓
GDP FACTOR			✓				
REAL-TIME	✓	✓	✓	✓			
DIAGONALS					✓	✓	✓
ROLLING				✓			

Note: Top panel: RMSFE across different specifications. Mid panel: RSMFEs relative to benchmark (other columns). [†] includes more mature GDP releases. All models include 25 monthly and 4 other quarterly variables, and 3 factors (i.e. 4 if also GDP factor is added).

the four GDP releases. Next we assess the effects of breaks and sample instabilities by estimating the RA-DFM model parameters using a rolling 10-year fixed-length window instead of the expanding one in the benchmark. These results are in column (d).

We next move to assess the information content in the real-time vintages of the monthly and quarterly data used in the model. In column (e), we replace the time series actually available at each point in time (i.e. the full real-time revision triangles) for the variables in x_t^Q and x_t^M with their first release only (diagonals). Column (f) is a standard DFM model for nowcasting where only the first releases for all variables in x_t are used; as consequence, the model's predictions stop when the first GDP estimate

TABLE 5: CRPS ACROSS MODEL SPECIFICATIONS

POSITION IN TRACKING WINDOW	(a)	(b)	(c)	(d)	(e)	(f)	(g)
FORECAST STARTS	0.342	0.345	0.492	0.365	0.350	0.351	0.349
NOWCAST STARTS	0.317	0.314	0.476	0.334	0.324	0.326	0.322
BACKCAST STARTS	0.289	0.288	0.454	0.306	0.291	0.294	0.295
GDP1	0.286	0.287	0.451	0.303	0.298	0.300	0.300
GDP2	0.103	0.104	0.464	0.140	0.103	–	0.150
GDP3	0.084	0.084	0.466	0.111	0.085	–	0.150
GDP4	0.055	0.056	0.467	0.050	0.055	–	0.150
RELATIVE TO BENCHMARK (a)		(b)	(c)	(d)	(e)	(f)	(g)
	1	1.009	1.439	1.068	1.024	1.027	1.021
	1	0.988	1.498	1.052	1.020	1.027	1.015
	1	0.998	1.572	1.059	1.009	1.020	1.021
	1	1.002	1.577	1.058	1.043	1.050	1.050
	1	1.011	4.487	1.351	0.999	–	1.451
	1	0.995	5.540	1.325	1.014	–	1.782
	1	1.024	8.485	0.907	0.999	–	2.725
GDP TARGETS	4	4	4	4	4	1	4 [†]
FACTORS VAR LAGS	1	4	1	1	1	1	1
VAR IDIOSYNCRATIC	✓	✓		✓	✓	✓	✓
GDP FACTOR			✓				
REAL-TIME	✓	✓	✓	✓			
DIAGONALS					✓	✓	✓
ROLLING				✓			

Note: Top panel: CRPS across different specifications. Mid panel: CRPSs relative to benchmark (other columns). [†] includes more mature GDP releases. All models include 25 monthly and 4 other quarterly variables, and 3 factors (i.e. 4 if also GDP factor is added).

is released. Lastly, in column (g) we replace the consecutive GDP releases with more mature data, namely, the 6th and the 12th monthly releases.

We note that the RA-DFM forecasting performance is similar across specifications for both point and density forecasts. We observe gains in the range of 3 to 7% in predicting the second and third estimates when allowing for longer lags in the VAR for the factors, and in predicting the last estimate when using only diagonals for x_t^Q and x_t^M . It is clear though that the VAR for the GDP idiosyncratic (Eq. 9) in our benchmark specification performs better than the alternative of having an extra ($r = 4$) factor only for the GDP releases as in column (c). Compared to the VAR specification in Eq. (9), allowing for an

extra GDP-specific factor introduces a further set of 4 AR(1) idiosyncratic terms which may complicate identification. The results in Table 4 also suggest that predictions for the first GDP estimate benefit from the inclusion of the first four monthly estimates of GDP growth in y_t . The comparison between columns (a) and (e) gives us a sense of the information contained in the revisions to the other data in the model. We note that this tends to be generally small. Column (f) shows that the inclusion of the different GDP releases does not essentially alter the performance of the model with respect to the first release, but impairs forecasts of future releases by construction. Finally, comparison between columns (a) and (g) reveals that while the information in more mature GDP vintages may be relevant in absolute terms, it is discounted in the model due to its large publication delay.

The evaluation of the density forecasts in terms of CRPS and Logscores reveals a similar pattern; although the performance of the various specifications is not too different, it seems that the inclusion of the real time vintages of monthly and quarterly series yields more accurate density forecasts than accounting for the first release only (diagonals). Finally, the model appears to benefit from considering multiple targets rather than using information in the first release of GDP only, as it traditionally happens in the nowcasting literature.

4.3 Forecasting the Revisions to Early GDP Releases: the Role of Data News

We now turn to analyse the role of different data releases in informing successive forecast updates using the benchmark RA-DFM model. To this aim, we define average impacts for each variable in x_t as the product of average weights times the average standard deviation of data news. Specifically, recall Eq. (15) from Section 2, reported below for convenience

$$\mathbb{E}\left[y_t^{(k)} \mid I_v\right] = \underbrace{\mathbb{E}\left[y_t^{(k)} I'_v\right] \mathbb{E}\left[I_v I'_v\right]^{-1}}_{\text{weight}} \underbrace{\left[x_\tau^* - \mathbb{E}(x_\tau^* \mid \Omega_{v-1})\right]}_{\text{news}}. \quad (15)$$

TABLE 6: LOGSCORES ACROSS MODEL SPECIFICATIONS

POSITION IN TRACKING WINDOW	(a)	(b)	(c)	(d)	(e)	(f)	(g)
FORECAST STARTS	1.114	1.134	1.585	1.147	1.137	1.140	1.133
NOWCAST STARTS	1.032	1.023	1.556	1.051	1.054	1.060	1.047
BACKCAST STARTS	0.954	0.950	1.513	0.985	0.971	0.982	0.976
GDP1	0.947	0.945	1.508	0.987	0.984	0.993	0.987
GDP2	0.024	0.041	1.553	0.204	0.023	–	0.379
GDP3	-0.251	-0.247	1.552	0.077	-0.245	–	0.377
GDP4	-0.591	-0.570	1.554	-0.836	-0.583	–	0.378
RELATIVE TO BENCHMARK (a)		(b)	(c)	(d)	(e)	(f)	(g)
		0.020	0.471	0.034	0.024	0.027	0.019
		-0.009	0.524	0.020	0.022	0.029	0.016
		-0.005	0.559	0.031	0.017	0.028	0.022
		-0.002	0.560	0.040	0.036	0.046	0.040
		0.017	1.528	0.180	-0.001	–	0.355
		0.004	1.803	0.327	0.006	–	0.628
		0.021	2.145	-0.245	0.008	–	0.969
GDP TARGETS	4	4	4	4	4	1	4 [†]
FACTORS VAR LAGS	1	4	1	1	1	1	1
VAR IDIOSYNCRATIC	✓	✓		✓	✓	✓	✓
GDP FACTOR			✓				
REAL-TIME	✓	✓	✓	✓			
DIAGONALS					✓	✓	✓
ROLLING				✓			

Note: Top panel: the negative of Logscores across different specifications across different specifications. Mid panel: Difference in Logscores from the benchmark (other columns). [†] includes more mature GDP releases. All models include 25 monthly and 4 other quarterly variables, and 3 factors (i.e. 4 if also GDP factor is added).

For each month, the average impacts for each variable x_τ^* are constructed as

$$i_{x^*} = \frac{1}{V} \sum_{v=1}^V b_{x^*} \bar{\sigma}_{x^*}, \quad (20)$$

where V is the number of data vintages in which x_τ^* is released, b_{x^*} are the weights in Eq. (15), and $\bar{\sigma}_{x^*}$ denotes the average standard deviation of the model's forecast errors (i.e. the data news, I_v) for x_τ^* .

Results for the forecast and nowcast period largely confirm previous findings in the literature (surveyed in Bańbura et al., 2013), and are reported in Appendix C. Over the

forecast period, the lion’s share is played by survey information. The largest impact variables are the forward looking components of the Markit CIPS/PMIs, particularly the components related to manufacturing production (CIPSEM) and services (CIPSES), followed by the CBI Survey on Industry Orders, and the Bank of England’s Agents Scores (see Section 3 for details on the input data). The impact of all the remaining data, including from financial and credit markets, is virtually null.

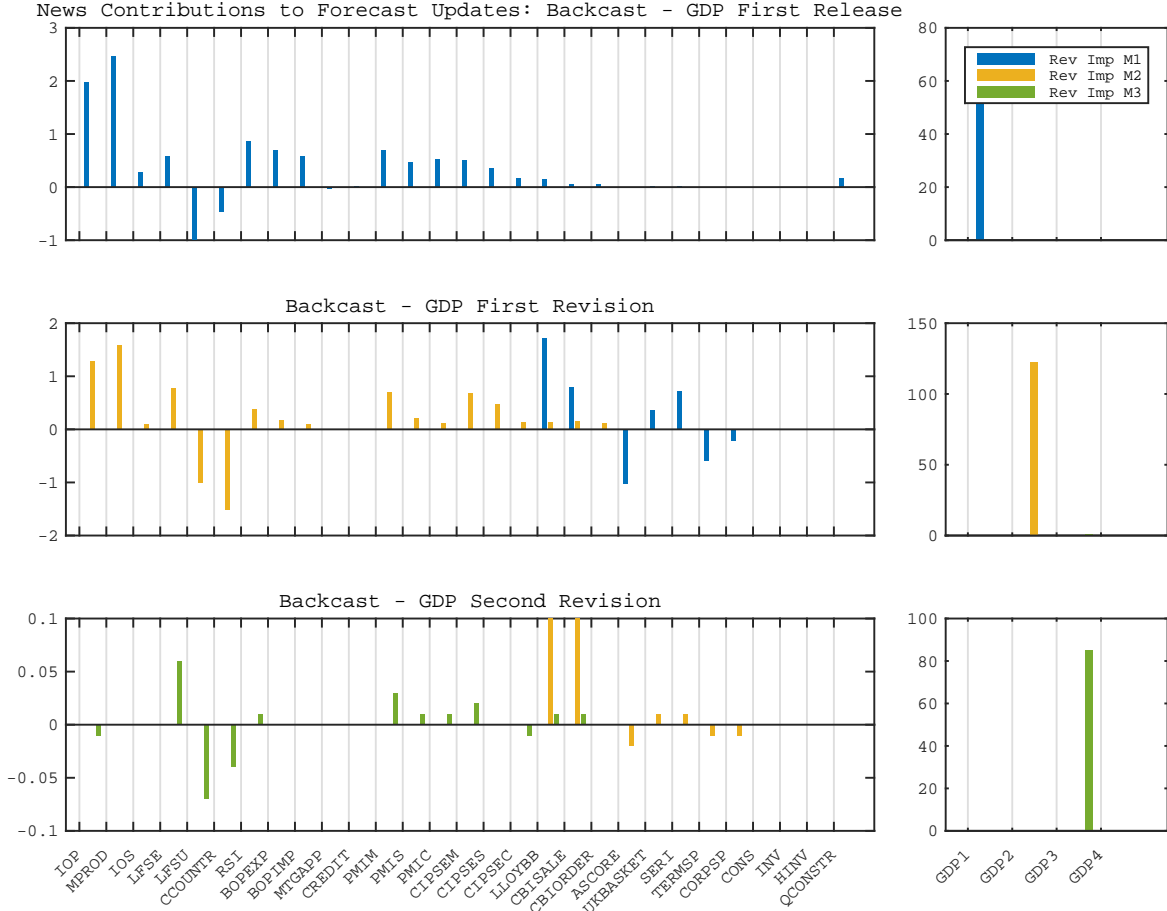
During the nowcast period, the largest impact variable is the first GDP release for the previous quarter, published towards the end of the first month (see Figure 2). Here both soft (survey) and hard data (ONS measures primarily) play an important role, with impacts that are comparable in magnitude. Importantly, however, while surveys tend to lose importance over time, the reverse is true for hard data such as the index of Production (IOP). Production data are in fact most informative in the third month of the nowcast period as this is when the first data relative to the reference quarter are published.

Figure 7 reports the average impacts of data releases on the backcast updates for the first (top panels), second (middle panels), and third (bottom panels) GDP release. The middle and bottom panels of Figure 7 equivalently report the contribution of data news in forecast updates for the first and second revision rounds (i.e. $rev_t^{(1)}$ and $rev_t^{(2)}$ using the notation of Section 2). For ease of comparison, average impacts are scaled such that all the targets have a standard deviation of 1, and M1, M2, M3 refer to the three months in the backcast period.¹⁶

In terms of timing, the panels of Figure 7 correspond (from top to bottom rows) to the teal, orange and purple areas of Figure 2 respectively. The impacts in the top row of Figure 7 are also in line with the previous literature, and confirm the importance of hard data, and production in particular, in backcasting the initial estimates of GDP growth. Compared to the nowcast period, here we note that the relative importance of surveys and hard data is reversed. Hard data are also important for predicting the first revision (middle panel). Production and labour data published in the second month after the end of the reference quarter (yellow bars) refer to the third month of the reference quarter,

¹⁶M1, M2 and M3 correspond to Apr, May, Jun for each Q1, to Jul, Aug, Sep for each Q2 and so on. The bars in Figure 7 are averages for each month in the backcast over the evaluation period, such that M1 is the average of releases published in Jan, Apr, Jul, Oct, M2 averages over Feb, May, Aug, Nov, and M3 averages over Mar, Jun, Sep, Dec.

FIGURE 7: PREDICTING THE REVISIONS: IMPACT OF DATA NEWS



Note: Impact of data releases for the backcast of the first (top panel), second (middle panel) and third (bottom panel) GDP release. Average impacts are constructed as average weights times the average standard deviation of the data news.

and are hence directly informative for the first revision of the GDP estimate. This is in line with results for the US reported in [Clements and Galvão \(2017\)](#). Similarly, surveys published in the first month after the end of the reference quarter, but after the first GDP release (blue bars) contain useful information for the first revision. And this is particularly true for the Lloyds Business Barometer (LLOYBB). Comparing the magnitude of the data impacts in the middle and bottom panels, we note that the predicting content of data releases is largely exhausted with the publication of the second GDP estimate. This is a direct consequence of the timeliness of the data included in our set – numbers published beyond the end of the second month after the end of the reference quarter start referring to the following one, and are hence little informative. In relative terms, an exception is

labour market data. Due to their longer publication delay, they still have some impact for the second revision, but magnitudes in absolute terms are negligible.

4.4 Forecasts for Later GDP Vintages

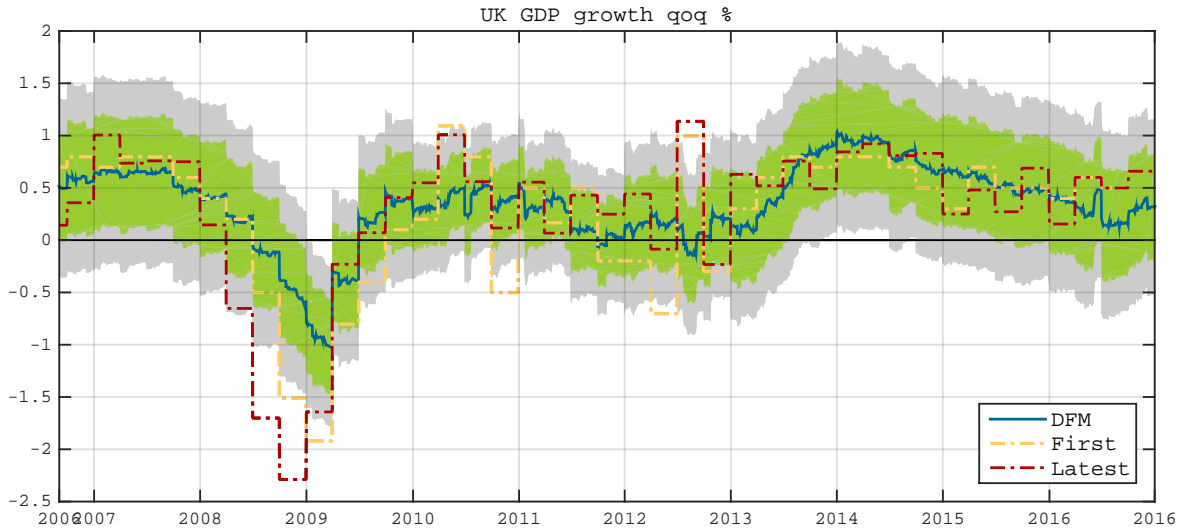
A standard way to think about subsequent GDP releases for the same quarter is as increasingly more accurate estimates of a ‘true’ growth figure. In this sense, latest vintages of GDP data can be thought of as being the best available estimates of historical growth, given that the majority of observations have gone through many rounds of revisions. In this section we evaluate the RA-DFM predictions against the latest available vintage of GDP data available at the time of writing (dating May 2018).

Figure 8 compares the benchmark RA-DFM nowcasts (i.e. the same as in Figure 5) against the first GDP release (yellow) and the May 2018 GDP vintage (maroon). Interestingly, the latest GDP growth estimate is outside the 95% predictive interval only 3 times compared with the 6 of the first release (see Section 4.1). This is a 50% reduction in comparison with the first release, and suggests that the RA-DFM model is able to provide a well-calibrated assessment of the revised estimate of GDP growth. Of the three events that lie outside the predictive interval, two are entirely idiosyncratic, as discussed in the comment to Figure 5. The remaining one is the 2008-2009 recession. It is worthwhile to note that, as documented also in Galvão and Mitchell (2018), the revisions in this period have shifted the turning points in an out of the recession phase such that the trough of the recession was anticipated to Q4 2008.

Building on this illustrative evidence, we look more formally at the informativeness of the RA-DFM model in predicting ‘true’ GDP data in Table 7. Here again ‘true’ are the growth values in the May 2018 vintage. Formally, we use a forecast encompassing test to compare the predictive content of the official ONS releases $y_t^{(1)}$, $y_t^{(2)}$, $y_t^{(3)}$ and $y_t^{(4)}$ against model’s forecasts (benchmark RA-DFM specification) conditional on information sets dating the day before each of the official releases. In other words, model’s forecasts are aligned such that the timing of the information set of the RA-DFM and the ONS coincide. The regression used to compute the test statistic is as follows:

$$y_t^{(2018M5)} = c + \lambda y_t^{(k)} + (1 - \lambda) \widehat{y}_t^{(k)} + \xi_t, \quad (21)$$

FIGURE 8: NOWCAST VS FIRST AND LATEST GDP RELEASES



Note: Evolution of nowcast for $y_t^{(1)}$ over the evaluation sample (blue line) with one and two standard deviation forecast error bands (green and grey areas respectively). Dash-dotted lines are for subsequent GDP releases with darker shades of red used for more recent publications. Latest vintage dates May 2018.

where $y_t^{(k)}$ denotes the k^{th} monthly release of GDP growth and $\hat{y}_t^{(k)}$ is the RA-DFM back-cast for $y_t^{(k)}$ dating a day before $y_t^{(k)}$ is due for publication. The coefficient λ determines the optimal forecast combination weights. If $\lambda = 0$, then the model's forecast $\hat{y}_t^{(k)}$ encompasses the corresponding official release of the statistical office, i.e., given $\hat{y}_t^{(k)}$, the information in $y_t^{(k)}$ can be dispensed with in order to forecast $y_t^{(2018M5)}$.

We estimate the regression in Eq. (21) with the observations of the out-of-sample period for the first four GDP releases and report the results in Table 7. The columns of Table 7 compare the information content of the ONS first, second, and third release with the corresponding RA-DFM predictions. The estimated weights, λ and $(1-\lambda)$, show that the RA-DFM predictions provide information for the 'true' state of the economy beyond what is contained in the official first estimates. This information is also sizeable, with estimated $(1-\lambda)$ roughly equal to one half for the first release and up to 0.7 for the fourth release. This is an important result, as it shows that the RA-DFM contains information that is useful to assess the true state of the economy, and that is not contained in the official estimates.

Results relative to the second and the third ONS releases are suggestive of the presence of multicollinearity between the official estimates and the RA-DFM forecasts. Indeed,

TABLE 7: OFFICIAL ESTIMATES VS RA-DFM PREDICTIONS

	FIRST	SECOND	THIRD	FOURTH
ONS GDP release	0.629*** (4.20)	0.598 (0.80)	1.187* (1.78)	0.555*** (5.10)
RA-DFM forecast	0.522** (2.08)	0.328 (0.44)	-0.349 (-0.52)	0.677*** (3.79)
c	-0.054 (-0.48)	0.029 (0.34)	0.063 (0.73)	-0.103 (-1.53)

Note: Forecast encompassing test. The dependent variable is UK GDP growth as per latest available vintage (May 2018). t statistics are reported in brackets, robust standard errors, $n=41$ quarters. The columns report the coefficients of Eq. (21) estimated using the first, second, third, and the fourth ONS releases respectively. Model's forecasts are aligned such that the timing of the information set of the RA-DFM and the ONS coincide.

the correlation between the series is 0.99 for both the second and third releases, compared to 0.82 and 0.75 for the first and fourth releases respectively. The estimates in Table 7, however, suggest that the RA-DFM does provide additional information to the statistical office release.

4.5 The RA-DFM and Institutional Forecasters

In this section we compare the RA-DFM forecasting performance with institutional forecasters. We restrict our attention to those institutions which produce nowcasts for the quarterly GDP growth rate as that produced by the model. These are the Bank of England (BoE), the National Institute for Economic and Social Research (NIESR), and the Bloomberg Survey of Economists (BSE).

The Bank of England disseminates its official forecasts once a quarter in its Inflation Report, published in February, May, August and November every year. The official forecasts represent the Monetary Policy Committee's best collective judgement and are informed by model-based (DSGE) forecasts, and by the staff's view. Additionally, for the nowcast quarter, the staff's view is informed by a suite of nowcasting models. As such, being a combination of different models and judgement, these forecasts effectively combine a much wider array of information than that in the benchmark RA-DFM, and are likely to account for idiosyncratic events such as e.g. weather conditions. The BoE

publishes its nowcasts for the current quarter once the first estimates for the previous quarter have been released, and roughly 70 days before the official first numbers for the current quarter are out.

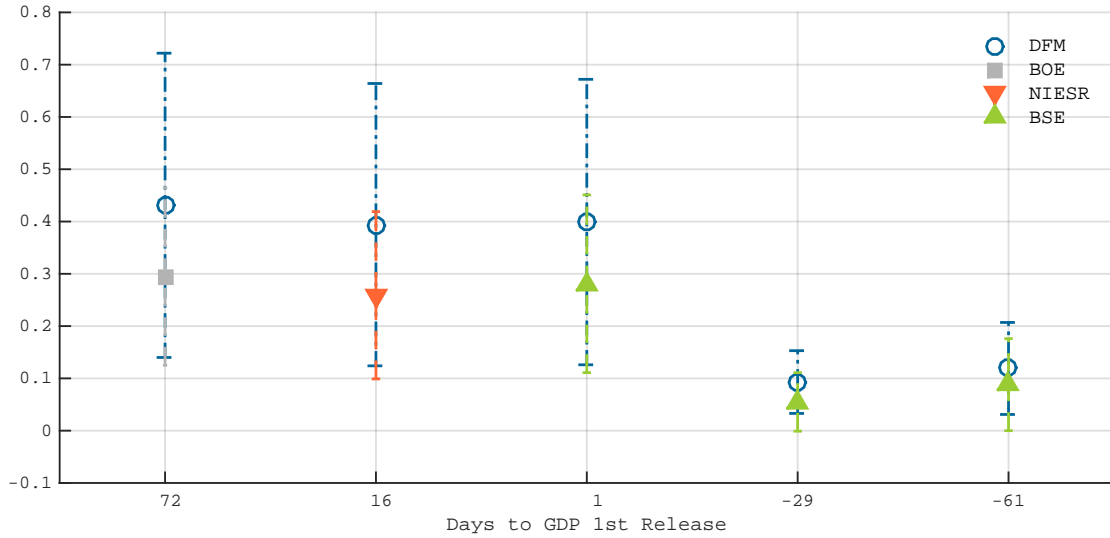
The NIESR currently publishes a monthly report on the status of the UK economy on the day of the release of the index of industrial production. The publication includes an estimate of monthly and 3-month growth rates, e.g. in May, the forecasts are for growth between March and April, and between February and April. Hence, to match the timing of the forecasts, we compare model's backcasts with NIESR forecasts published in January, April, July and October. NIESR nowcasts are based on monthly GDP estimates, constructed by output data (the methodology is described in the [Mitchell et al. \(2005\)](#) paper) and also include staff judgement.

Finally, we compare the model's forecasts with the median response of the survey of professional forecasters distributed by the financial services provider Bloomberg. The survey typically starts around two weeks before each GDP release is due, and respondents can contribute forecasts up to 24 hours before the official release date. The survey is finalised on the day before each GDP release, and collects forecasts from economists in major businesses and the financial services industry. The panels are unbalanced over time, but count an average of around 100 contributors. Individual forecasts can be based on models, pure judgement, or a combination of the two.

Hence, the characteristics of the professional forecasters we compare against make them tough to beat: on the one hand, the inclusion of judgement accounts for idiosyncratic events, on the other, being essentially forecast combinations, they are also more resilient to outliers and other possible structural breaks.

We compare the RMSFEs of the model and the three external forecasters in [Figure 9](#). For each external forecaster, we align the model's forecasts such that the dates in which the forecasts are produced coincide; these are expressed in distance from the publication of the first official GDP release, and reported on the x axis. [Figure 9](#) also reports confidence intervals for these forecasts constructed from the average standard deviation of the forecast errors over the evaluation sample (Q4 2006:Q4 2016). Inspection of [Figure 9](#) reveals that, unsurprisingly, judgement-based forecast combinations tend to be more accurate. However, and notwithstanding the larger model's implied uncertainty com-

FIGURE 9: COMPARISON WITH PROFESSIONAL FORECASTERS



Note: Average RMSFE of benchmark specification (blue circles) compared with Bank of England (BOE, grey square), NIESR (orange triangle), and the Bloomberg Survey (BSE, green triangle). Vertical lines are two standard deviation error bars computed over the quarters in the evaluation sample (Q4 2006:Q4 2016). Days to the release of GDP first estimates are reported on the x axis and correspond to the publication of the institutional forecasts. Forecast errors are computed against the ONS release targeted by the institutional forecaster, i.e. first release for the BOE and NIESR, and first, second and third for the consecutive BSEs.

pared to that of the institutional forecasters, the average RA-DFM RMSFEs still remain within the confidence bands of the competing forecaster. This is important, since both Bloomberg and NIESR forecasts often produce zero forecast errors. These differences naturally decrease after the first GDP estimate becomes available and the model targets the next GDP release. To assess the model’s ability to predict the ‘true’ state of the economy, as illustrated in Section 4.4, we repeated the same exercise but this time using the latest available GDP vintage (May 2018 in our case). Although the latest vintage is a harder benchmark to beat and the RMSFEs of both the RA-DFM and the external forecasters are higher in *absolute* terms, the RA-DFM performs better against the latest vintage in *relative* terms.

It is important to finally note here that the model attains these levels of forecast accuracy long before the institutional forecasters publish their predictions (see Figure 6) and while institutional forecasters only publish one prediction for every quarter, the model is able to provide with continuous forecast updates immediately after every data release and the RA-DFM forecasts come with a well calibrated predictive density.

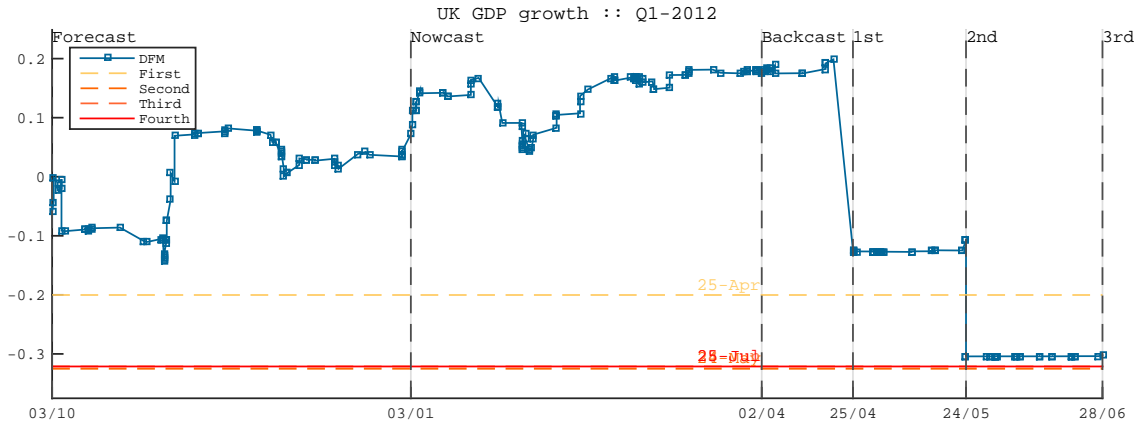
4.6 The ‘Double-Dip’ Recession and the EU Referendum

We conclude this section with an illustration of two interesting case studies that highlight the usefulness of our modelling approach to measure UK GDP growth in circumstances of high uncertainty about the current state of the economy.

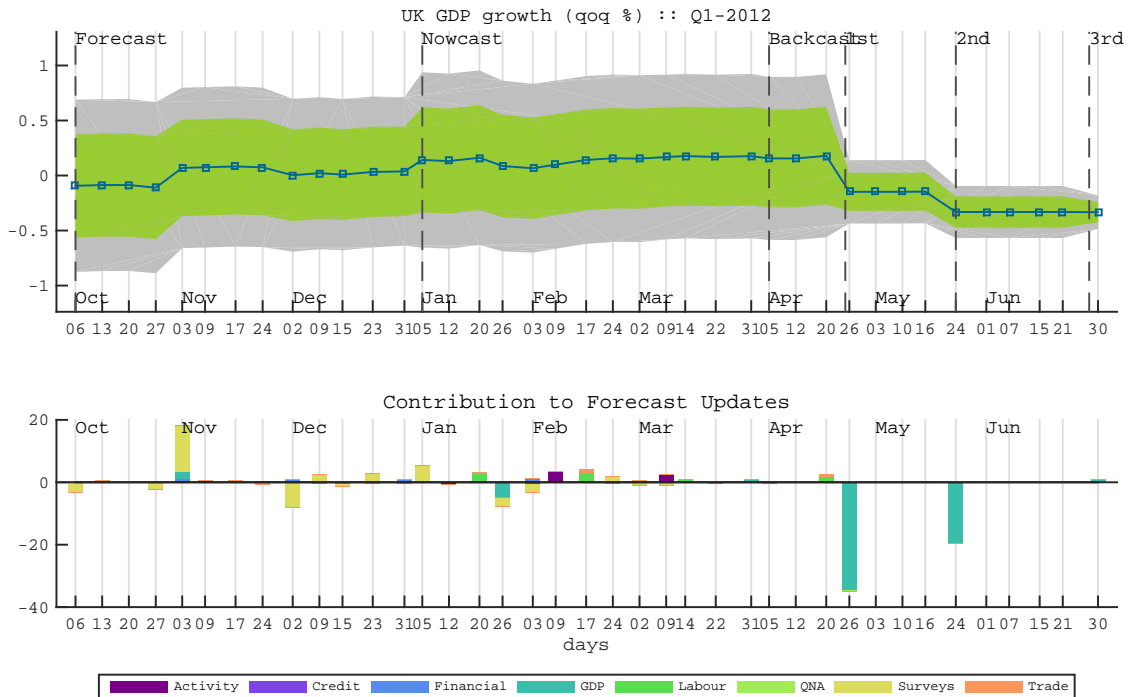
The first episode is the so-called ‘Double-Dip’ Recession of 2012. When numbers for Q1 2012 were first published, in April 2012, analysts concluded that the UK had entered a new recessionary phase. Indeed, according to April 2012 GDP vintage, the growth of -0.2% in Q1 2012 was the second consecutive quarter of negative GDP growth, since the value for Q4 2011 was -0.3% (see also Figure 5). Following a major annual revision of historic data published by the ONS in June 2013, the evidence of two-consecutive quarters of negative growth was removed. According to the latest figures (May 2018), the growth rates in both Q4 2011 and Q1 2012 were positive. Figure 10 shows the evolution of forecasts, nowcasts and backcasts for the Q1 2012 reference quarter, computed with all real-time data vintages from October 2011 up to June 2012. By inspecting the RA-DFM measurement of UK growth before the publication of the preliminary release in April in Figure 10, we can conclude that data news have been on balance positive, in agreement with the latest vintage values for Q1 2012: our measurements only turned negative after the publication of the ONS preliminary estimate. In contrast, professional forecasters predicted negative growth rates for the same quarter. Even though the model’s nowcasts are in line with the latest vintage values, the value of the ONS preliminary estimate is still within the model’s 68% predictive interval, providing evidence that the RA-DFM model is able to provide a good assessment of the Q1 2012 GDP uncertainty.

The second episode is the quarter following which the UK voted out of the European Union. The “Brexit Referendum” took place on June 23, 2016. Following the referendum results, all professional forecasters, including the Bank of England, revised downward their short-term forecast for UK growth by up to 0.25%. The downward revisions were primarily driven by the publication of rather pessimistic survey data, such as the PMIs. Figure 11 reports the RA-DFM forecasts, nowcasts and backcasts for Q3 2016 using all the real-time data vintages between April 2016 and December 2016. Inspection of Figure 11 reveals that while surveys did indeed have an initial negative impact on

FIGURE 10: THE DOUBLE-DIP RECESSION

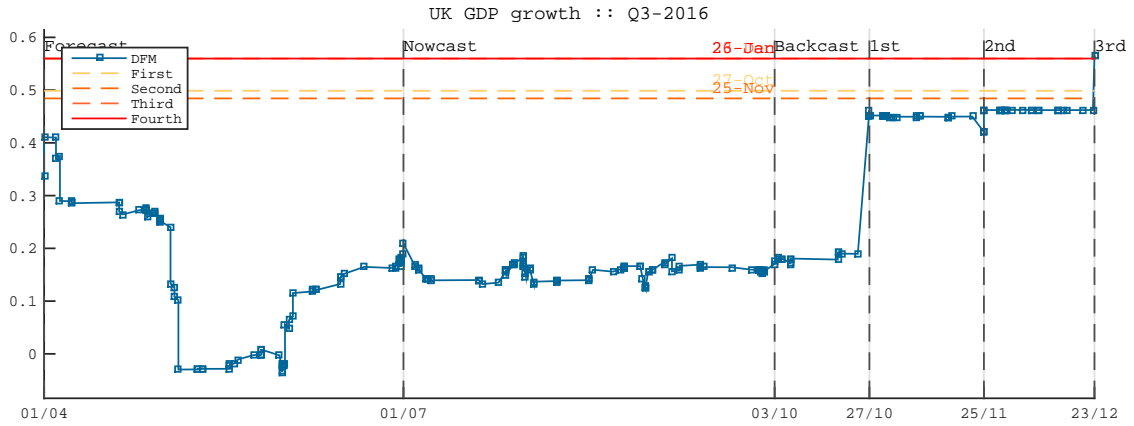


(A) Note: Evolution of forecast for Q1 2012 (blue line) and first four releases for the same quarter.

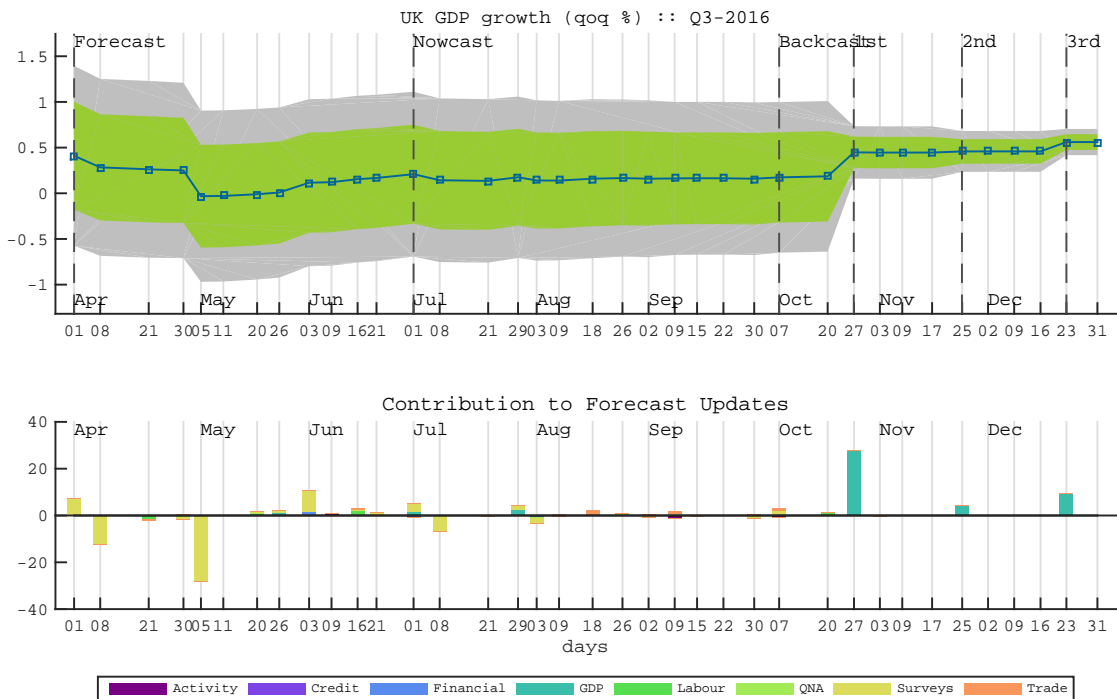


(B) Note: Contribution of data to updates in the forecasts for Q1 2012.

FIGURE 11: THE EU REFERENDUM



(A) Note: Evolution of forecast for Q3 2016(blue line) and first four releases for the same quarter.



(B) Note: Contribution of data to updates in the forecasts for Q3 2016.

the model’s forecast, this was subsequently balanced out by other positive data news. As a consequence, while the RA-DFM forecast was updated downwards by about 0.25 percentage points following the publication of survey data in May, positive contributions from surveys in June reverted the nowcasts upwards, and new information published after the referendum had little impact on the nowcasts. The model’s best prediction at the time of the publication of the first release was 0.2%, just 0.1 percentage points below out pre-referendum forecasts of about 0.3% growth. As in the previous episode, the model’s 68% predictive interval includes the first outturn (0.5%).

In summary, both examples illustrate how the RA-DFM can extract a reliable signal to measure GDP growth by filtering out the noise that may contaminate the early statistical releases, while providing a good assessment of the uncertainty around the point estimate.

5 Conclusions

In this paper, we have proposed a new framework for nowcasting data subject to revision, and applied it to nowcasting UK real GDP growth in real-time. To this aim, we have also compiled a rich and comprehensive real-time mixed-frequency dataset for the UK economy, assembled using official data stored over the years in the archives of the Bank of England. The dataset covers the years 1990-2016, and full real-time revision triangles since 2006.

The econometric framework that we propose is a Release-Augmented Dynamic Factor Model, or RA-DFM. The novelty with respect to previous nowcasting models resides in augmenting the measurement equation of the state-space representation of the DFM with consecutive official estimates for the same target variable, real UK GDP growth in our case, and relative to the same quarter. The model permits forecasting the target variable beyond the release of first statistical office estimate, and allows for a simple characterisation of the stochastic process for the revisions to initial releases of macroeconomic data. Within this framework, we are able to assess the contribution that different pieces of data have in informing updates to *i*) nowcast of current GDP growth, *ii*) uncertainty around point forecasts, and *iii*) forecasts of the revisions to previously released GDP data.

We find that the model produces accurate estimates of ‘true’ UK GDP growth, mea-

sured using the latest available vintage of data, and that it contains information useful to predict ‘true’ growth beyond that contained in official earlier estimates. While retaining parsimony, and without the introduction of any element of judgement, the RA-DFM real-time GDP growth estimates are commensurate with model combinations and judgement-based forecasts embedded in the predictions of institutional forecasters. Finally, we found that the RA-DFM yields well calibrated predictive intervals, and that ‘hard data’ are most informative in forecasting future revisions to GDP estimates.

References

- Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2014) “Nowcasting GDP in real time: a density combination approach,” *Journal of Business and Economic Statistics*, Vol. 32, pp. 48–68.
- Anesti, N., S. Hayes, A. Moreira, and J. Tasker (2017) “Peering into the present: the Bank’s approach to GDP nowcasting,” *Bank of England Quarterly Bulletin*, Vol. Q2, pp. 122–133.
- Artis, M. J., A. Banerjee, and M. Marcellino (2005) “Factor Forecasts for the UK,” *Journal of Forecasting*, Vol. 24, pp. 279–298.
- Aruoba, S. B., F. X. Diebold, F. Nalewaik, and F. Schorfheide (2016) “Improving GDP Measurement: A Measurement-Error Perspective,” *Journal of Econometrics*, Vol. 191, pp. 384–397.
- Bañbura, M. and M. Modugno (2014) “Maximum Likelihood estimation of factor models on datasets with arbitrary pattern of missing data,” *Journal of Applied Econometrics*, Vol. 29, No. 133-160.
- Bañbura, Marta, D. Giannone, M. Modugno, and L. Reichlin (2013) “Now-casting and the Real-Time Data Flow,” in *Handbook of Economic Forecasting, volume 2A*: Elsevier, Chap. 4, pp. 195–237.
- Bragoli, Daniela and Michele Modugno (2017) “A now-casting model for Canada: Do U.S. variables matter?,” *International Journal of Forecasting*, Vol. 33, No. 4, pp. 786–800.
- Carriero, A., M. P. Clements, and A. B. Galvão (2015) “Forecasting with Bayesian Multivariate Vintage-Based VARs,” *International Journal of Forecasting*, Vol. 35, pp. 757–768.
- Castle, J. and C. Ellis (2002) “Building a real-time database for GDP(E),” *Bank of England Quarterly Bulletin*, No. Spring, pp. 42–48.
- Clements, M. P. and A. B. Galvão (2013a) “Forecasting with Vector Autoregressive Models of Data Vintages: US output growth and inflation,” *International Journal of Forecasting*, Vol. 29, pp. 698–714.
- (2013b) “Real-time Forecasting of Inflation and Output Growth with Autoregressive Models in the Presence of Data Revisions,” *Journal of Applied Econometrics*, Vol. 28, pp. 458–477. DOI: 10.1002/jae.2274.

- Clements, M. P and A. B. Galvão (2017) “Predicting Early Data Revisions to US Output Growth and the Effects of Releases on Equity Markets,” *Journal of Business and Economic Statistics*, Vol. 35, No. 591-604.
- Croushore, D. (2006) “Forecasting with Real-Time Macroeconomic Data,” in *Handbook of Economic Forecasting I*: Elsevier, pp. 961–982.
- Cunningham, A., J. Eklund, C. Jeffery, G. Kapetanios, and V. Labhard (2012) “A state space approach to extracting the signal from uncertain data,” *Journal of Business & Economic Statistics*, Vol. 30, pp. 173–80.
- Egginton, Don M., Andreas Pick, and Shaun Vahey (2002) “‘Keep it real!’: a real-time UK macro data set,” *Economics Letters*, Vol. 77, No. 1, pp. 15–20.
- Faust, J., J. H. Rogers, and J. H. Wright (2005) “News and Noise in G-7 GDP Announcements,” *Journal of Money, Credit and Banking*, Vol. 37 (3), pp. 403–417.
- Galvão, A. B. (2017) “DSGE Models and Data revisions,” *Journal of Econometrics*, Vol. 196, pp. 215–232.
- Galvão, A. B. and J. Mitchell (2018) “Measuring GDP Growth Data Uncertainty,” *University of Warwick, mimeo*.
- Garratt, A. and S. P. Vahey (2006) “UK real-time macro data characteristics,” *Economic Journal*, Vol. 116, pp. F119–F135.
- Giannone, D., L. Reichlin, and D. Small (2008) “Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases,” *Journal of Monetary Economics*, Vol. 55, pp. 665–676.
- Jacobs, J. P. A. M and S. van Norden (2011) “Modeling Data Revisions: Measurement Error and Dynamics of ‘True’ Values,” *Journal of Econometrics*, Vol. 161, pp. 101–109.
- Kishor, N. K. and E. F. Koenig (2012) “VAR Estimation and Forecasting When Data are Subject to Revision,” *Journal of Business and Economic Statistics*, Vol. 30, pp. 181–190.
- Koenig, E. F., S. Dolmas, and J. Piger (2003) “The use and abuse of real-time data in economic forecasting,” *The Review of Economics and Statistics*, Vol. 85(3), pp. 618–628.
- Mankiw, N. G. and M. D. Shapiro (1986) “News or Noise: An Analysis of GNP Revisions,” *Survey of Current Business (May 1986)*, US Department of Commerce, Bureau of Economic Analysis, pp. 20–25.

- Mariano, Roberto S. and Yasutomo Murasawa (2003) “A new coincident index of business cycles based on monthly and quarterly series,” *Journal of Applied Econometrics*, Vol. 18, No. 4, pp. 427–443.
- Miranda-Agrippino, Silvia (2012) “Nowcasting UK GDP,” Technical report, London Business School.
- Mitchell, J. (2009) “Where are we now? The UK recession and nowcasting GDP growth using statistical models,” *National Institute Economic Review*, Vol. 209, pp. 60–69.
- Mitchell, J., R. J. Smith, M. R. Weale, S. Wright, and E. L. Salazar (2005) “An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth*,” *The Economic Journal*, Vol. 115, No. 501, pp. F108–F129.

A Estimation and State-Space Representation

A.1 State Space Representation of the RA-DFM

The full RA-DFM of Section 2 with $n_K = 4$ is

$$x_t \equiv \begin{pmatrix} x_t^M \\ x_t^Q \\ y_t^{(1)} \\ \vdots \\ y_t^{(4)} \end{pmatrix}_{(n \times 1)} \simeq \begin{pmatrix} \mathbf{\Lambda}_M & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Lambda}_Q & 2\mathbf{\Lambda}_Q & 3\mathbf{\Lambda}_Q & 2\mathbf{\Lambda}_Q & \mathbf{\Lambda}_Q \\ \mathbf{\Lambda}^{(1)} & 2\mathbf{\Lambda}^{(1)} & 3\mathbf{\Lambda}^{(1)} & 2\mathbf{\Lambda}^{(1)} & \mathbf{\Lambda}^{(1)} \\ \vdots & & \vdots & & \vdots \\ \mathbf{\Lambda}^{(4)} & 2\mathbf{\Lambda}^{(4)} & 3\mathbf{\Lambda}^{(4)} & 2\mathbf{\Lambda}^{(4)} & \mathbf{\Lambda}^{(4)} \end{pmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-4} \end{pmatrix} + \begin{pmatrix} \zeta_t^M \\ \zeta_t^Q \\ \varepsilon_t^{(1)} \\ \vdots \\ \varepsilon_t^{(4)} \end{pmatrix}, \quad (8)$$

$$f_t = \mathbf{A}_1 f_{t-1} + \dots + \mathbf{A}_p f_{t-p} + \eta_t \quad \eta_t \sim \mathcal{N}(0, \mathbf{\Sigma}), \quad (2)$$

$$\zeta_t^M = \mathbf{D} \zeta_{t-1}^M + \epsilon_t^M \quad \epsilon_{i,t}^M \sim \mathcal{N}(0, \varsigma_{M,i}^2), \quad (3)$$

$$\zeta_t^Q = \mathbf{\Psi} \zeta_{t-1}^Q + \epsilon_t^Q \quad \epsilon_{j,t}^Q \sim \mathcal{N}(0, \varsigma_{Q,i}^2), \quad (\text{A.1})$$

$$\varepsilon_t = \mathbf{\Phi} \varepsilon_{t-1} + \nu_t \quad \nu_t \sim \mathcal{N}(0, \mathbf{\Gamma}). \quad (9)$$

\mathbf{D} and $\mathbf{\Psi}$ are both diagonal, and $\mathbf{\Phi}$ is a full 4×4 matrix. We can rewrite the equations above in state-space representation

$$x_t = \mathbf{C} s_t + e_t \quad e_t \sim \mathcal{N}(0, \mathbf{R}), \quad (\text{A.2})$$

$$s_t = \mathbf{A} s_{t-1} + u_t \quad u_t \sim \mathcal{N}(0, \mathbf{Q}), \quad (\text{A.3})$$

where x_t and e_t are $n \times 1$, \mathbf{C} is $n \times n_s$, s_t and u_t are $n_s \times 1$, \mathbf{R} is $n \times n$ and \mathbf{A} and \mathbf{Q} are $n_s \times n_s$. Recall that x_t^M and x_t^Q are vectors of dimensions n_M and n_Q respectively. Let q denote the number of lagged factors needed for the approximation in Eq. (4), i.e. $q = 4$, and n_K denote the number of GDP releases the DFM is augmented with, also equal to 4 in our case. We define:

- $n = n_M + n_Q + n_K$: number of observables,
- $n_s = n_s^f + n_s^M + n_s^Q + n_s^K$: number of unobserved states,
- $n_s^f = r(\max\{p, q\} + 1)$: number of states for factors,
- $n_s^M = n_M$: number of states for monthly idiosyncratic,
- $n_s^Q = n_Q(\max\{p, q\} + 1)$: number of states for quarterly idiosyncratic.

- $n_s^K = 4(\max\{p, q\} + 1)$: number of states for GDP releases idiosyncratic.

With $p < q$, $q = 4$, $k = 4$, and $n_Q = 1$ (i.e. there is one quarterly variable besides GDP):

$$s_t = \begin{pmatrix} s_t^f \\ (n_s^f \times 1) \\ \hline s_t^M \\ (n_s^M \times 1) \\ \hline s_t^Q \\ (5 \times 1) \\ \hline s_t^K \\ (n_s^K \times 1) \end{pmatrix} = \begin{pmatrix} (f'_t \ \cdots \ f'_{t-4})' \\ \hline (\zeta_{1,t}^M \ \cdots \ \zeta_{n_M,t}^M)' \\ \hline (\zeta_t^Q \ \cdots \ \zeta_{t-4}^Q)' \\ \hline (\varepsilon_t \ \cdots \ \varepsilon_{t-4})' \end{pmatrix}, \quad (\text{A.4})$$

where $\varepsilon_t \equiv (\varepsilon_t^{(1)}, \dots, \varepsilon_t^{(4)})'$ and the partitions identify (from top to bottom) the states referring to the factors (s_t^f), and to the idiosyncratic for the monthly variables (s_t^M), the quarterly variables (s_t^Q), and the GDP releases (s_t^K).

$$\mathbf{C}_{(n \times n_s)} = \begin{pmatrix} \Lambda_M & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I}_{n_M} & \mathbf{0} & \mathbf{0} \\ \hline \Lambda_Q & 2\Lambda_Q & 3\Lambda_Q & 2\Lambda_Q & \Lambda_Q & \mathbf{0} & \mathcal{R} & \mathbf{0} \\ \hline \Lambda^{(1)} & 2\Lambda^{(1)} & 3\Lambda^{(1)} & 2\Lambda^{(1)} & \Lambda^{(1)} & \mathbf{0} & \mathbf{0} & \mathcal{R} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \Lambda^{(4)} & 2\Lambda^{(4)} & 3\Lambda^{(4)} & 2\Lambda^{(4)} & \Lambda^{(4)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{R} \end{pmatrix} \quad (\text{A.5})$$

where $\mathbf{0}$ denotes matrices of zeros of conformable dimensions, \mathbb{I}_m is the identity matrix of dimension m , and $\mathcal{R} = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \end{bmatrix}$.

$$\mathbf{R}_{(n \times n)} = \varrho \mathbb{I}_n, \quad (\text{A.6})$$

where ϱ is a very small number.

$$\mathbf{A}_{(n_s \times n_s)} = \begin{pmatrix} A_1 & \dots & A_p & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbb{I}_{r(q-1)} & \mathbf{0} & & & & \\ \hline & & \mathbf{0} & \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \hline & & \mathbf{0} & \mathbf{0} & \mathbf{\Psi} & \mathbf{0} \\ \hline & & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Phi} \end{pmatrix}. \quad (\text{A.7})$$

Finally,

$$\mathbf{Q}_{(n_s \times n_s)} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} & & & \\ \mathbf{0} & \mathbf{0} & & & \\ \hline & & \mathbf{Q}^M & & \\ \hline & & \mathbf{0} & \mathbf{Q}^Q & \\ \hline & & \mathbf{0} & & \mathbf{Q}^K \end{pmatrix}, \quad (\text{A.8})$$

where \mathbf{Q}^M is a diagonal matrix with the variances of the idiosyncratic monthly variables $\zeta_{M,i}^2$, \mathbf{Q}^Q is a block diagonal matrix with a block for each quarterly variable, and each block is all zeros except for the element (1,1) which equals the variance of the idiosyncratic quarterly variables $\zeta_{Q,i}^2$. Finally, \mathbf{Q}^K is a sparse matrix with the elements of $\mathbf{\Gamma}$ appropriately placed in correspondence of the contemporaneous covariances of the idiosyncratic terms for the GDP releases.

The structure in Eqs. (A.5 - A.8) is easily extended to accommodate the presence of block structures in the specification of f_t , by appropriately modifying the relevant matrix partitions.

A.2 Estimation

Maximum Likelihood estimation of the RA-DFM can be carried using the EM Algorithm, where the Kalman Filter is used to calculate the expected conditional likelihood, and the Kalman Smoother updates the estimates of the states vector and relevant auto-covariance matrices at each iteration. The presence of missing values in x_t is handled by appropriately modifying the two algorithms such that the weight assigned to the missing

observations vanishes at each $t \in [1, T]$ (see [Bańbura and Modugno, 2014](#)).

Let $\mathbf{C}_{[\iota]}$, $\mathbf{R}_{[\iota]}$, $\mathbf{A}_{[\iota]}$, $\mathbf{Q}_{[\iota]}$ denote the system matrices estimated at iteration ι of the EM. Moreover:

- $\Theta_{[\iota]}$: collects all parameters at iteration ι ,
- Ω_v : information set at data vintage v ,
- $\mathbb{E}_{\Omega, \iota} \equiv \mathbb{E}[\cdot | \Omega_t, \Theta_{[\iota]}]$: expectation conditional on all data and parameters at ι ,
- $s_{t|T, \iota} \equiv \mathbb{E}_{\Omega, \iota}[s_t]$: smoothed states,
- $P_{t|T, \iota}$: smoothed states variance,
- $P_{t, t-1|T, \iota}$: smoothed states first order autocovariance.

Further, partition $\mathbf{C}_{[\iota]}$, $\mathbf{A}_{[\iota]}$, $\mathbf{Q}_{[\iota]}$ such that

$$\mathbf{C}_{[\iota]} = \begin{bmatrix} \mathbf{C}_{[\iota]}^M & \mathbb{I} & \mathbf{0} \\ \mathbf{C}_{[\iota]}^Q & \mathbf{0} & \mathbb{I}_{k+1} \otimes \mathcal{R} \end{bmatrix}, \quad (\text{A.9})$$

$$\mathbf{A}_{[\iota]} = \begin{bmatrix} \mathbf{A}_{[\iota]}^f & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{[\iota]}^M & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{[\iota]}^Q \end{bmatrix}, \quad (\text{A.10})$$

$$\mathbf{Q}_{[\iota]} = \begin{bmatrix} \mathbf{Q}_{[\iota]}^f & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{[\iota]}^M & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{[\iota]}^Q \end{bmatrix}. \quad (\text{A.11})$$

In Eqs. (A.10 - A.11), $\mathbf{A}_{[\iota]}^M = \mathbf{D}$, $\mathbf{Q}_{[\iota]}^M$ is defined in ??, and with GDP the only quarterly variable $\mathbf{A}_{[\iota]}^Q = \mathbf{\Phi}$ and $\mathbf{Q}_{[\iota]}^Q = \mathbf{\Gamma}$. With the states being non observable, for each partition of $s_{t|T, \iota}$ the set of relevant sufficient statistics is given by:

$$\mathbb{E}_{\Omega, \iota}[s_t^j s_t^{j'}] = s_{t|T, \iota}^j s_{t|T, \iota}^{j'} + \sum_{t=1}^T P_{t|T, \iota}^j, \quad (\text{A.12})$$

$$\mathbb{E}_{\Omega, \iota}[s_{t-1}^j s_{t-1}^{j'}] = s_{t-1|T, \iota}^j s_{t-1|T, \iota}^{j'} + \sum_{t=1}^T P_{t-1|T, \iota}^j, \quad (\text{A.13})$$

$$\mathbb{E}_{\Omega, \iota}[s_t^j s_{t-1}^{j'}] = s_{t|T, \iota}^j s_{t-1|T, \iota}^{j'} + \sum_{t=1}^T P_{t, t-1|T, \iota}^j \quad j \in \{f, M, Q\}. \quad (\text{A.14})$$

Lastly, if at any $t \in [1, T]$ x_t contains missing observations, define W_t to be an $n \times n$ diagonal matrix of logical identifiers which singles out the available information discarding the unknowns.

The components in Eqs. (A.9 - A.11) at iteration $\iota + 1$ are the maximizers of the expected log likelihood conditional on Ω_t and $\Theta_{[\iota]}$. For the measurement equation:

$$\text{vec}(\mathbf{C}_{[\iota+1]}^M) = \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_t^f s_t^{f'}] \otimes W_t \right]^{-1} \left[\text{vec} \left(\sum_{t=1}^T W_t (x_t s_{t|T, \iota}^{f'} - \mathbb{E}_{\Omega, \iota} [s_t^f s_t^{f'}]) \right) \right], \quad (\text{A.15})$$

$$\text{vec}(\mathbf{C}_{[\iota+1]}^Q) = \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_t^f s_t^{f'}] \otimes W_t \right]^{-1} \left[\text{vec} \left(\sum_{t=1}^T W_t (x_t s_{t|T, \iota}^{f'} - \mathbb{E}_{\Omega, \iota} [s_t^f s_t^{f'}]) \right) \right]. \quad (\text{A.16})$$

When restrictions on the quarterly loadings are active, then those are enforced using the standard constrained least squares formula on the relevant partition of the parameters ($\mathbf{\Lambda} \in \mathbf{C}_{[\iota+1]}^Q$). In our case, restrictions are in place to bridge the monthly and quarterly observations. Write the restrictions as $\mathbf{B}\mathbf{\Lambda} = \mathbf{b}$, where $\mathbf{\Lambda}$ is the partition of $\mathbf{C}_{[\iota+1]}^Q$ which is subject to restriction, and \mathbf{b} is a vector of zeros. The restricted loadings are given by:

$$\mathbf{\Lambda}_c = \mathbf{\Lambda} - \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_t^f s_t^{f'}] \right]^{-1} \mathbf{B}' \left[\mathbf{B} \left(\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_t^f s_t^{f'}] \right)^{-1} \mathbf{B}' \right]^{-1} (\mathbf{B}\mathbf{\Lambda} - \mathbf{b}), \quad (\text{A.17})$$

and Eq. (A.16) is adapted conformably.

For the parameters of the state equation:

$$\mathbf{A}_{[\iota+1]}^f = \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_t^f s_{t-1}^{f'}] \right] \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_{t-1}^f s_{t-1}^{f'}] \right]^{-1}, \quad (\text{A.18})$$

$$\mathbf{A}_{[\iota+1]}^{\forall j \in \{M\}} = \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_t^j s_{t-1}^{j'}] \right] \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_{t-1}^j s_{t-1}^{j'}] \right]^{-1}, \quad (\text{A.19})$$

$$\mathbf{A}_{[\iota+1]}^Q = \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_t^K s_{t-1}^{K'}] \right] \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \iota} [s_{t-1}^K s_{t-1}^{K'}] \right]^{-1}, \quad (\text{A.20})$$

and

$$\mathbf{Q}_{[\ell+1]}^f = \frac{1}{T} \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \ell} [s_t^f s_t^{f'}] - \mathbf{A}_{[\ell+1]}^f \sum_{t=1}^T \mathbb{E}_{\Omega, \ell} [s_t^f s_{t-1}^{f'}] \right], \quad (\text{A.21})$$

$$\mathbf{Q}_{[\ell+1]}^{\forall j \in \{M\}} = \frac{1}{T} \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \ell} [s_t^j s_t^{j'}] - \mathbf{A}_{[\ell+1]}^j \sum_{t=1}^T \mathbb{E}_{\Omega, \ell} [s_t^j s_{t-1}^{j'}] \right], \quad (\text{A.22})$$

$$\mathbf{Q}_{[\ell+1]}^Q = \frac{1}{T} \left[\sum_{t=1}^T \mathbb{E}_{\Omega, \ell} [s_t^t s_t^{K'}] - \mathbf{A}_{[\ell+1]}^Q \sum_{t=1}^T \mathbb{E}_{\Omega, \ell} [s_t^K s_{t-1}^{K'}] \right]. \quad (\text{A.23})$$

B The UK Real-Time Data

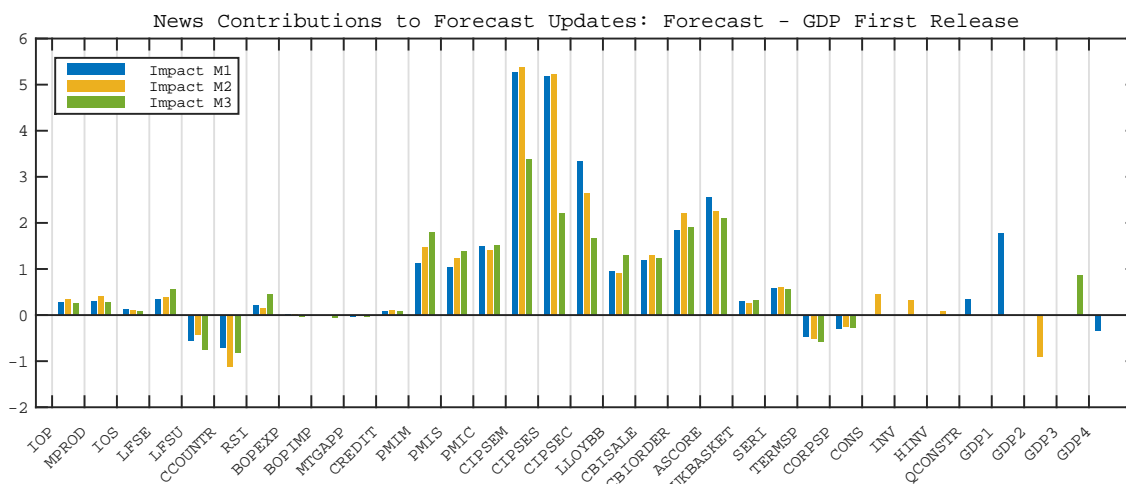
TABLE B.1: REAL-TIME VINTAGES

VARIABLE NAME	ONS CODE		EARLIEST VINTAGE	EARLIEST DATA POINT	UNITS
	PRE 2011	POST 2011			
GDP	ABMI.Q	ABMI.Q	04-1990	1990Q1	£ million, CVM, SA
CONS	ABJR.Q+ HAYO.Q	ABJR.Q+ HAYO.Q	11-2006	1990Q1	£ million, CVM, SA
INV	NPEL.Q	NPEL.Q	11-2006	1990Q1	£ million, CVM, SA
HINV	DFEG.Q+ L635.Q+ L637.Q	DFEG.Q+ L635.Q+ L637.Q	11-2006	1990Q1	£ million, CVM, SA
QCONSTR	L2N8.Q	L2N8.Q	10-2006	1993Q1	
IOP	CKYW.M	K222.M	09-2006	01-1990	Index, SA
MPROD	CKYY.M	K22A.M	09-2006	01-1990	Index, SA
IOS	FVQQ.M	S2KU.M	09-2006	01-1995	Index, SA
BOPEXP	BOKG.M	BOKG.M	09-2006	01-1990	£ million, SA
BOPIMP	BOKH.M	BOKH.M	09-2006	01-1990	£ million, SA
RSI	EAPS.M	J5EK.M	09-2006	01-1990	Index, SA

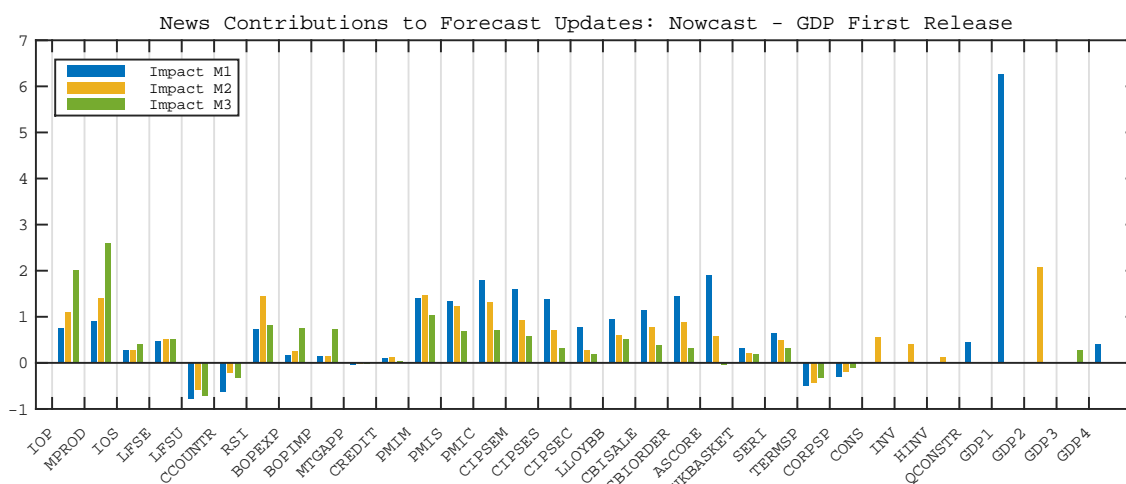
Note: The table summarised the availability of real-time data. The Variable Name is the same used in Table 2. The second and third column report the official ONS identifiers. The Earliest Vintage refers to the timing of first available real-time vintage. The Earliest Data Point indicates the starting point of the time series and the Units correspond to the exact format in which the series is stored in the Bank of England internal database. CVM stands for Chained Volume Measures.

C Additional Charts

FIGURE C.1: IMPACT OF DATA NEWS – FORECAST AND NOWCAST



(A) Forecast



(B) Nowcast

Note : Impact of data releases for the forecast (top panel) and the nowcast (middle panel) of the first GDP release. Average impacts are constructed as average weights times the average standard deviation of the data news.