Uncertain Kingdom: A Framework for Nowcasting GDP and its Revisions☆

Nikoleta Anesti* Ana Beatriz Galvão† Silvia Miranda-Agrippino‡

Bank of England University of Warwick Bank of England and CFM

First version: November 22, 2016 – This version: October 31, 2017

Preliminary and Incomplete

Do Not Circulate

Abstract

We design a new mixed-frequency framework to nowcast data subject to revisions. Subsequent releases of GDP relative to the same quarter are treated as separate but correlated observables in an augmented state-space that also includes a large number of heterogeneous predictors. The framework allows for a simple characterisation of the revision process as a function of the observables, and permits a detailed assessment of the contribution of different data releases in informing (i) subsequent forecasts updates for the target variable(s); (ii) the evolution of forecast uncertainty over time; and (iii) the forecasts for future revisions of the target variable. We construct a real-time mixed-frequency database for the UK economy and use our model to nowcast UK GDP and its subsequent 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/

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

☆ We are grateful to... The content of this draft may be subject to change. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of England, the Monetary Policy Committee, the Financial Policy Committee or the Prudential Regulation Authority.

1 Introduction

Macroeconomic forecasters face a dilemma when choosing a method for nowcasting economic activity and GDP growth. The academic literature suggests the use of statistical forecasting models because of their simplicity and accuracy. However, it is not always easy to provide an economic narrative to explain the forecasts obtained with these reduced-form models. Also, when the target series is subject to potentially important revision rounds, it is not normally clear which release of GDP the forecaster should aim to nowcast. Early releases permit a timely assessment of current economic conditions, but are inherently incomplete; preliminary GDP figures are typically estimated based on data covering a fraction (about 44%, according to the Office for National Statistics) of the required survey data. Later releases provide more accurate estimates, but the delay in their publication reduces their relevance for real-time forecasting. Observers, market participants, and professional forecasters typically consider large numbers of predictors when forming expectations, including successive monthly releases of GDP growth for any given quarter (see Clements and Galvão, 2014). Conversely, however, statistical nowcasting models typically do not incorporate information on GDP data revisions. Even when using real-time data, the focus is usually on one specific release at a time, normally, the latest available vintage of data (see e.g. Bańbura et al., 2013).

In this paper, we propose a method to compute point and density forecasts for the preliminary (first) GDP release and its subsequent revisions in real-time. We treat different GDP releases for the available observations (quarters) as separated quarterly time series in a Dynamic Factor Model, where we also include a large set of predictors sampled at different frequencies. For each quarter, we compute forecasts by explicitly targeting different GDP releases (e.g. preliminary, advanced, and final) at different points in time. We update our information set following the calendar of data release, hence allowing early releases to inform nowcasts of later ones. Treating different GDP releases as separate observables allows us to incorporate the suggestion in Kishor and Koenig (2012); Clements and Galvão (2013); Clements (2015); Galvao (2017) to avoid mixing data of different maturities, that is, we should separate heavily revised data from lightly revised data when estimating forecasting models.

We use a large set of quarterly and monthly predictors and update the information set every time new data is published. Following Bańbura and Modugno (2010), we extract model-based data news (or surprise components of data releases) every time new data is published, and link them to (*i*) point forecast updates for the relevant releases of GDP growth; (*ii*) the evolution of forecast uncertainty around point estimates; and (*iii*) forecast updates for the revision rounds of GDP growth. This allows us to complement our density forecasts with a fully data-driven narrative as in Bańbura and Modugno (2010).

We apply our new method to compute density nowcasts of UK output growth in real-time. To this end, we also compile a comprehensive mixed-frequency real-time database for the UK economy, which includes national account data, surveys, production, trade and price data, and a number of financial indicators. 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. This includes releases of revised UK GDP data occurring up to one quarter after the preliminary/first release is published.

Revisions to UK GDP growth figures can be quite substantial. One illustrative episode is the recession of 2011/2012, referred to as the double-dip recession because of the previous slowdown of 2008/2009. When GDP data for the first quarter of 2012-Q1 were first released, in April 2012, economists concluded the economy was in recession since the data indicated negative growth for two consecutive quarters. This recession event was revised away from the data in June 2013, when the Office for National Statistics (ONS) released data suggesting a small but positive growth rate for 2012-Q1. Because of the implications that an accurate measurement of the state of the UK economy has for the conduct of policy, institutions may evaluate revisions to past data. The Bank of England, for example, has been publishing backcasts of UK GDP growth in their 'Inflation Report' since November 2007. This paper moves mixed-frequency nowcasting factor models one-step further by incorporating information on releases of data revisions, and using the information contained in early releases to forecast subsequent revisions.

At the time of writing, we have only just assembled the real-time database, which we describe in Section 2. To the best of our knowledge, ours is the first real-time mixed-frequency database for the UK economy. Hence, we currently present the model evaluation using real time vintages for GDP growth only. The model can accurately predict subsequent releases of UK GDP growth. Also, the efficient incorporation of the data flow translates into forecast uncertainty that systematically decreases over time, as more information is used to construct the nowcasts. We also discuss the role of different releases and use the news to produce a narrative around point and density forecasts for all GDP releases, and the forecasts of the upcoming revisions. These tools are particularly useful for policy-makers and market participants.

This paper contributes to the large body of literature on nowcasting GDP using Dynamic Factor Models, which were firstly introduced by Giannone, Reichlin and Small (2008). Their approach to nowcasting is popular with professional forecasters, in particular central banks, as a way to deliver accurate real-time forecasts using a statistical model that is able to provide a systematic analysis of changes in perceptions of current economic conditions. Numerous applications to point forecasts using (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). Evaluation of density forecasts resulting from mixed-frequency dynamic factor models using real-time data are in e.g. Aastveit et al. (2014). The paper also contributes to the literature on the use of real-time data in forecasting (Clements and Galvão, 2012, 2013; Diebold et al., 2016). Miranda-Agrippino (2012) uses a Dynamic Factor Model to nowcast UK GDP growth in a pseudo real-time setting.

The paper is organised as follows. We discuss characteristics of UK data revisions and the construction of the real-time dataset in Section 2. Section 3 describes our approach to handle data revisions and compute density forecasts. Empirical results are collected in Section 4. Section 5 concludes.

2 A UK Real-Time Dataset for Nowcasting

The Office for National Statistics (ONS) produces Gross Domestic Product measures 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 of GDP, and the UK Quarterly

National Accounts. The ONS publishes the first estimate of GDP based on information on output 3.5 weeks after the end of the quarter. The second estimate of GDP, based on information from all approaches, is published 8 weeks after the end of the quarter, and includes information on the level of GDP as well as the growth in GDP. The UK quarterly national accounts are published 12 weeks after the end of the reference quarter. After the first three rounds, the ONS can still revise their GDP estimate. In fact, the window for revisions can be open for up to three years back, as new information might still need to be incorporated. Finally, there are annual revisions published every June (in most of the cases) as part of the Blue Book publication. In this paper, we only explicitly model the first four monthly GDP releases.

The variables included in our real-time dataset (see Table 1 below) comprise indicators followed closely by market participants and the Bank of England, including ONS public releases for output, expenditure, prices and labour market data. We also monitor business and output surveys and financial market data. As we are interested in explicitly incorporating the data revision process of GDP in the model and evaluate it in real time, we need vintages of the GDP and of all the indicators subject to revision. Some indicators, as financial variables, output and business surveys, and prices are only revised if there is a methodological change. As such, we only include a static vintage of these variables. For the remaining indicators subject to revision, we assembled real time vintages that are internally compiled by the Bank of England and for which we have availability as indicated in the Table 1.

Table 1: A Real-Time Dataset for the UK

Vintage	Timing	Release	Publication Lag	Transformation	Real-Time vintages	Frequency	RT vintages availability	Block
1	last week of current m.	Lloyds Business Barometer	m	Level	0	M	No revision	S
2	last week of current m.	CBI Distributive Trade	m	Level	0	M	No revision	S
3	last week of current m.	CBI Industrial Trends	m	Level	0	M	No revision	S
4	last bus.day of current m.	Agents' Scores	m	Level	0	M	No revision	S
5	last bus.day of current m.	UK Focused Equity Index	m	Level	0	M	No revision	F
6	last bus.day of current m.	Sterling Effective Exchange Rate	m	Level	0	M	No revision	F
7	last bus.day of current m.	Term Spread	m	Level	0	M	No revision	F
8	last bus.day of current m.	Corporate Bond Spread	m	Level	0	M	No revision	F
9	1st bus.day of the 1st m.	Markit/CIPS Manufacturing	$m-1$	Level	0	M	No revision	S
10	2nd bus.day of the 1st m.	Markit/CIPS Construction	$m-1$	Level	0	M	No revision	S
11	3rd bus.day of the 1st m.	Markit/CIPS Services	$m-1$	Level	0	M	No revision	S
12	2nd week of the 1st m.	Claimant Count Rate	$m-1$	1st Diff	0	M	No revision	N
13	2nd week of the 1st m.	Retail Sales ex Fuel	$m-1$	Log-Diff	1	M	1997M2	R
14	2nd week of the 1st m.	CPI All Items	$m-1$	Log-Diff	0	M	No revision	N
15	2nd week of the 1st m.	RPI All Items	$m-1$	Log-Diff	0	M	No revision	N
16	2nd week of the 1st m.	PPI Input	$m-1$	Log-Diff	0	M	No revision	N
17	2nd week of the 1st m.	PPI Output	$m-1$	Log-Diff	0	M	No revision	N
18	1st bus.day of 2nd m.	Mortgages Approved	$m-2$	Log-Diff	0	M	No revision	N
19	1st bus.day of 2nd m.	Net Consumer Credit	$m-2$	Level	0	M	No revision	N
20	2nd week of 2nd m.	Index of Production	$m-2$	Log-Diff	1	M	2009M9	R
21	2nd week of 2nd m.	Manufacturing Production	$m-2$	Log-Diff	1	M	2009M9	R
22	2nd week of 2nd m.	Exports Volume (Goods)	$m-2$	Log-Diff	1	M	Quarterly vintages	R
23	2nd week of 2nd m.	Imports Volume (Goods)	$m-2$	Log-Diff	1	M	Quarterly vintages	R
24	last week of 3rd m.	Index of Services	$m-3$	Log-Diff	1	M	2006M2	R
25	last week of 3rd m.	GDP Preliminary Release	$m-3$	Level	1	Q	1990Q1	R
26	2nd week of 4th m.	LFS Number of Employees	$m-3$	Log-Diff	0	M	No revision	R
27	2nd week of 4th m.	LFS Unemployment Rate	$m-3$	1st Diff	0	M	No revision	R
28	last week of 4th m.	GDP Second Estimate	$m-4$	Level	1	Q	1990Q1	R
29	last week of 4th m.	Total Consumption	$m-4$	Log-Diff	1	Q	1990Q1	R
30	last week of 4th m.	Business Investment	$m-4$	Log-Diff	1	Q	1990Q1	R
31	last week of 4th m.	Housing Investment	$m-4$	Log-Diff	1	Q	1997Q1	R
32	last week of 4th m.	Construction Output	$m-4$	Log-Diff	1	Q	1990Q1	R
33	last week of 5th m.	GDP Quarterly National Accounts	$m-5$	Level	1	Q	1990Q1	R
34	last week of 6th m.	GDP Fourth Release	$m-6$	Level	1	Q	1990Q1	R

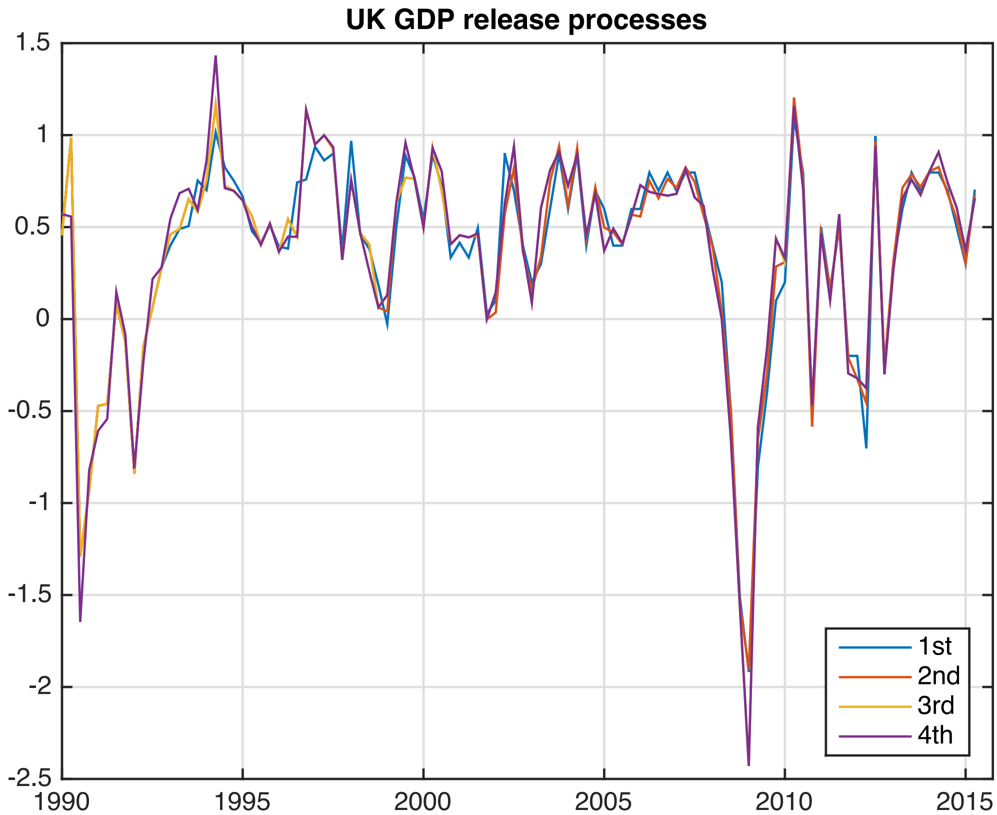
Data releases are indicated in rows. Column 1 indicates the progressive number associated to each 'vintage'. Column 2 indicates the official dates of the publication. Column 3 indicates the releases. Column 4 indicates the publishing lag: e.g. Index of Production is released with 2-months delay (m-2). Column 5 indicates the transformation. Column 6 indicates if we use real-time vintages. Column 7 shows the frequency. Column 8 explains the details of the real-time vintages. Column 9 indicates to which block of data the variable belong to, i.e. S is for 'Soft', N is for Nominal, F is for Financial and R is for Real variables.

2.1 GDP Data Revisions

Initial data revisions are normally caused by information not available at the time of the previous release, taking into account that the statistical office faces a trade-off between timeless and completeness of the surveyed data. The top part of Figure 1 plots the first four monthly releases of UK GDP quarterly growth, as distributed by the ONS – i.e. 'preliminary', 'second estimate, 'Quarterly National Accounts' (third), and fourth and the bottom part of Figure 1 shows the equivalent revision processes, always as the deviations from the preliminary release.

Table 2 presents the characteristics of the first four releases of GDP growth rate that are used in the model and Table 3 presents the same summary statistics of the revision processes that are always defined as deviations from the preliminary release. The standard deviation of the second revision process, i.e. between the Quarterly National Accounts and the preliminary estimate, is largest one, which confirms that in this round the largest amount of information is incorporated and more specifically the Index of Services that is the most relevant for the UK economy. The results of the Lyung-Box Q(4) test for a serial correlation of order 4 of the revision processes suggest that for the first two revision processes there's statistical evidence of serial correlation, although in more mature estimates the results change.

Figure 1: GDP Releases & Revision processes



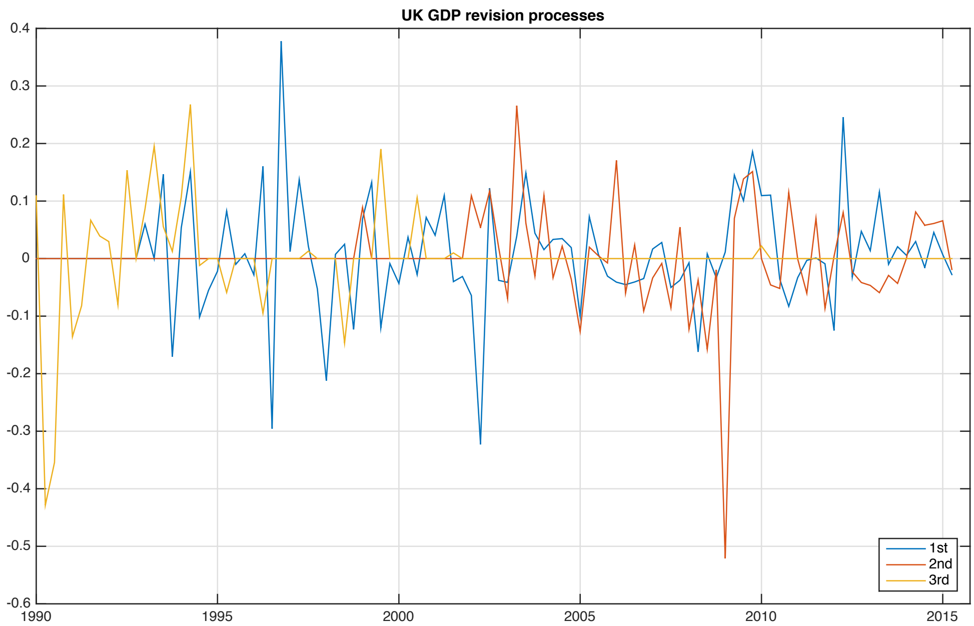


TABLE 2: SUMMARY STATISTICS FOR RELEASES OF DIFFERENT MATURITY

	1st release	2nd estimate	QNA	4th release
Mean	0.3618	0.3690	0.3697	0.3715
Stdev	0.5581	0.5575	0.5801	0.5966
AC(1)	0.6075	0.6150	0.6190	0.6311

TABLE 3: : SUMMARY STATISTICS FOR DATA REVISIONS

	1st	2nd	3rd
Mean	0.0144	-0.0051	0.0004
Stdev	0.0751	0.1107	0.0031
AC(1)	-0.2042	0.0034	0.1648
Q(4)	-0.0288	-0.0288	0.0267
p-values	[0.1385]	[0.9256]	[0.0257]

Note: The revisions are computed as deviations from the first release.

3 A Mixed-Frequency DFM for Data Subject to Revisions

We start by applying the model in Bańbura et al. (2013) to monthly and quarterly data. We use x_t^M to denote a generic $(n_M \times 1)$ vector of demeaned stationary monthly variables, observed at $t = 1, 2, \ldots, T$. Similarly, x_t^Q is a $(n_Q \times 1)$ vector of quarterly zero-mean stationary variables observed at $t = 3, 6, \ldots$. We assume that x_t^M has a factor structure, and write:

$$x_t^M = \Lambda_M f_t + \xi_t^M, \tag{1}$$

$$f_t = A_1 f_{t-1} + \ldots + A_p f_{t-p} + u_t \qquad u_t \sim \mathcal{N}(0, \Sigma_u), \tag{2}$$

$$\xi_{i,t}^M = \rho_i \xi_{i,t-1}^M + \epsilon_{i,t} \qquad \epsilon_{i,t} \sim \mathcal{N}(0, \sigma_{\epsilon_i}^2), \tag{3}$$

where $i = 1, \ldots, n_M$ and $f_t \equiv (f_{1,t}, \ldots, f_{r,t})'$ is an $(r \times 1)$ vector of zero-mean unobserved factors. Λ_M is a $(n_M \times r)$ matrix and collects the loadings for the monthly variables on the factors. The factors follow a VAR(p). A_j, $j = 1, \ldots, p$ are $(r \times r)$ matrices of autoregressive coefficients. ξ_t^M is a vector of AR(1) idiosyncratic terms.

As in Bańbura et al. (2010), we follow Mariano and Murasawa (2003) and incorporate quarterly variables into the framework by constructing partially observed monthly equivalents which we assume follow the same factor structure in Eq. (1-3). Let X_t^Q denote the quarterly level variable, e.g. log level of consumption. Let $\tilde{X}_t^M$ be its unobservable, monthly counterpart such that $X_t^Q = \tilde{X}_t^M + \tilde{X}_{t-1}^M + \tilde{X}_{t-2}^M$, and define $\tilde{x}_t^M \equiv \tilde{X}_t^M - \tilde{X}_{t-1}^M$. The approximation in Mariano and Murasawa (2003) allows to write[1]

$$\begin{aligned} x_t^Q &\simeq R(L)\tilde{x}_t^M, \\ &= (1 + 2L + 3L^2 + 2L^3 + L^4)\tilde{x}_t^M. \end{aligned} \tag{4}$$

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

$$\underset{(n\times1)}{x_t} \equiv \begin{pmatrix} x_t^M \\ \\ x_t^Q \end{pmatrix} \simeq \begin{pmatrix} \Lambda_M & \mathbf{0}_{n_M \times 4r} \\ \\ \underbrace{\Lambda_M R(L)}_{\Lambda_Q} & \end{pmatrix} f_t + \begin{pmatrix} \xi_t^M \\ \\ \underbrace{R(L)\tilde{\xi}_t^M}_{\xi_t^Q} \end{pmatrix}. \tag{5}$$

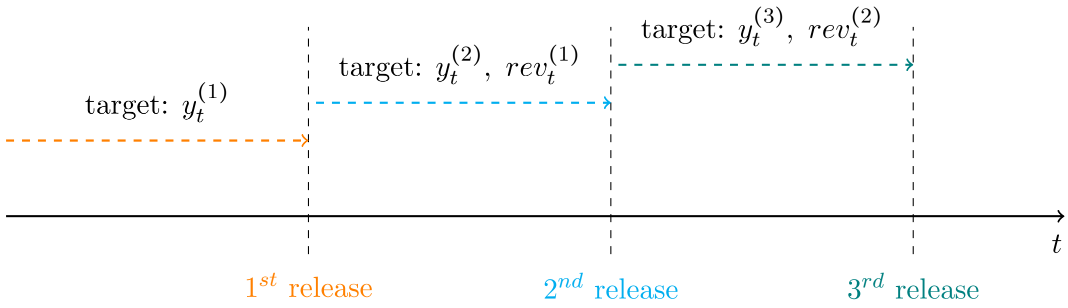
The mixed-frequency model in Eq. (2-3) and (5) can be cast in state-space form and estimated using Maximum Likelihood as in Ba?bura and Modugno (2014) . Prior to estimation, variables are transformed to achieve stationarity (see Table 1 for details on transformations) and standardised. To account for missing values and the 'ragged-edge' that is characteristic of real-time data vintages we implement the algorithm in Ba?bura and Modugno (2014). Details on the estimation and the state-space representation of the model are in Appendix A.

3.1 GDP Releases in the DFM

We assume that (the unobserved monthly components of) successive GDP releases *for the same quarter* follow a factor structure. The same as in Eq. (1-3). As discussed in Section 2, official figures for UK GDP growth, y_t, are sampled at quarterly frequency

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

Figure 2: Nowcast Targets



Add description

and revised in successive monthly rounds. Initial data revisions are normally caused by information not available at the time of the previous release, therefore, there is scope for using the real-time data flow to inform forecasts of the successive releases for any given quarter. Let $y_t^{(k)}$ denote the k^{th} monthly update for y_t released by the statistical office, such that for every quarter $y_t^{(1)}$ is the first release (preliminary estimate), published at the end of the first month following the observational quarter. We define the k^{th} revision as[2]

$$rev_t^{(k)} \equiv y_t^{(k+1)} - y_t^{(k)}. \tag{6}$$

We explicitly target up to three revision rounds or, equivalently, up to four subsequent releases for any given quarter. Given the ONS publication schedule, this means that we stop focusing on the current quarter when the preliminary release relative to the next quarter is published. While we explicitly target multiple releases, we focus on them one at the time, depending on where we are in each quarter. Figure 2 sketches the intuition.

To incorporate the different releases of GDP data, we augment Eq. (5) with the time

[2]Alternatively, we could adopt the alternative convention

$$rev_t^{(k)} \equiv y_t^{(k+1)} - y_t^{(1)},$$

which measures the revision always relative to the first release. Practically, for our purpose, the two definitions are equivalent. Using one rather than the other does not modify our argument.

series for the first ($y_t^{(1)}$), second ($y_t^{(2)}$), third ($y_t^{(3)}$) and fourth release ($y_t^{(4)}$) as follows:

$$\underset{(n\times 1)}{x_t} \equiv \begin{pmatrix} x_t^M \\ x_t^Q \\ y_t^{(1)} \\ \vdots \\ y_t^{(4)} \end{pmatrix} \simeq \begin{pmatrix} \Lambda_M \quad \underset{(n_M\times 4r)}{0} \\ \underset{(n_Q\times 5r)}{\Lambda_Q} \\ \underset{(1\times 5r)}{\Lambda^{(1)}} \\ \vdots \\ \underset{(1\times 5r)}{\Lambda^{(4)}} \end{pmatrix} f_t + \begin{pmatrix} \xi_t^M \\ \xi_t^Q \\ \varepsilon_t^{(1)} \\ \vdots \\ \varepsilon_t^{(4)} \end{pmatrix}, \tag{7}$$

where we use $\varepsilon_t^{(k)}$ to denote the idiosyncratic components of the four GDP releases.

Because of the publication lags, and depending on where we are in the quarter, we may have availability of data up to a given reference period t for a different set of these time series. For example, at the beginning of April 2016, we have values of $y_t^{(1)}$, $y_t^{(2)}$ and $y_t^{(3)}$ up to 2015-Q4 – i.e. relative to $(t-4)$ –, but $y_t^{(4)}$ is available up to 2015-Q3 $(t-7)$ only. This implies that by computing $y_{t-1}^{(1)}$, we are predicting 2016-Q1 first release GDP, but we could also predict the third revision of GDP for 2015-Q4, by computing values for $y_{t-4}^{(4)}$. If we move one month forward, we would observe the first release for 2016-Q1, that is $y_{t-2}^{(1)}$, but we may be interested in using the model to compute predictions for the second estimate of 2016-Q1, that is, $y_{t-2}^{(2)}$.

While we use a fully real-time dataset, and therefore account for the information contained in the revisions for all other variables in x_t as well, we only explicitly model GDP revision rounds. In particular, Eq. (7) implies that we implicitly assume the following process for GDP revisions:

$$rev_t^{(k)} = \left(\Lambda^{(k+1)} - \Lambda^{(k)}\right) f_t + R(L)\left(\varepsilon_t^{(k+1)} - \varepsilon_t^{(k)}\right). \tag{8}$$

Revisions will have a common factor if $\Lambda_{(k+1)} \neq \Lambda_{(k)}$, and they may be serially correlated. The revisions could also change the mean of the underlying process, since we treat successive releases as separated time series. Contrary to Jacobs and van Norden (2011), we do not characterise the revisions as either news or noise. However, we are able to accommodate those revisions that reflect the acquisition of new information by the statistical office if $\Lambda_{(k+1)} \neq \Lambda_{(k)}$, since f_t is our current measurement of economic activity obtained using all information available – i.e. $f_t = \mathbb{E}[f_t \mid \Omega_v]$, where Ω_v denotes the information set at date v. Predictability that instead arises from to the autocorrelation of the revision process, will be captured by the autoregressive structure of $\varepsilon_t^{(k)}$, $\forall k$. Cunningham et al. (2012) assume that the variability of each successive revisions declines with the data maturity k. We do not impose this restriction, but empirically, depending on the values of $\sigma^2_{\epsilon_1}, ..., \sigma^2_{\epsilon_{k+1}}$ and $(\Lambda^{(1)})^2, ..., (\Lambda^{k+1})^2$, we could have that $\mathbb{V}ar\left(rev_t^{(k)}\right) \leq \mathbb{V}ar\left(rev_t^{(1)}\right)$.

Because we assume that the common and idiosyncratic disturbances are Gaussian, and use filtering equations for a Gaussian linear state-space model, it is easy to assume a Gaussian density for our estimates of $y_t^{(k)}$ for $k = 1, ..., K$. The density mean (or point) forecast is:

$$\mu_t^{(k)} \equiv \mathbb{E}\left[y_t^{(k)} \mid \Omega_v\right] = \Lambda^{(k)}\mathbb{E}\left[f_t \mid \Omega_v\right] + \mathbb{E}\left[\varepsilon_t^{(k)} \mid \Omega_v\right]. \tag{9}$$

The forecast variance is:

$$\left(\sigma_t^{(k)}\right)^2 \equiv \mathbb{E}\left[y_t^{(k)} y_t^{(k)\prime} \mid \Omega_v\right] = \Lambda^{(k)}\mathbb{E}\left[f_t f_t' \mid \Omega_v\right]\Lambda^{(k)\prime} + \mathbb{E}\left[\varepsilon_t^{(k)}\varepsilon_t^{(k)\prime} \mid \Omega_v\right], \tag{10}$$

where $\mathbb{E}\left[f_t f_t' \mid \Omega_v\right]$ and $\mathbb{E}\left[\varepsilon_t^{(k)}\varepsilon_t^{(k)\prime} \mid \Omega_v\right]$ are obtained from Kalman filter equations for the state variances. The predictive density is then $\mathcal{N}\left(\mu_t^{(k)}, \left(\sigma_t^{(k)}\right)^2\right)$.

We use the following convention for the forecasts horizons. If v falls within the current quarter, we nowcast $y_t^{(1)}$, and forecast $y_t^{(2)}, \ldots, y_t^{(4)}$. Similarly, if it falls within the first month following the reference quarter, but still before the publication of the first release, we backcast $y_t^{(1)}$, nowcast $y_t^{(2)}$, and forecast $y_t^{(3)}$ and $y_t^{(4)}$.

3.2 News Contributions to Forecasts Updates

TBA

4 Nowcasting UK GDP Growth

4.1 Model Evaluation

We start our empirical exploration by evaluating the models performance in a pseudo real-time environment. To this end, we reconstruct pseudo real-time vintages for the variables in x_t^M and x_t^Q (see Section 3) by applying the effective calendar of data releases to a vintage of data downloaded on the 16/02/2017. We use instead real-time numbers for the four target releases of GDP growth $(y_t^{(1)}, \ldots, y_t^{(4)})$. Given the publication schedule of the data considered (see Table 1), the majority of the monthly 'hard data (e.g. production, services, construction and trade) in our set are only available up to the end of December 2016. Similarly, the most recent figures for quarterly hard data (e.g. consumption and investment) are relative to 2016-Q4. The latest employment figure, with the exception of the Claimant Count Rate, are instead available up to November. Prices and 'soft data (i.e. surveys) are available until January 2017. We convert financial variables into monthly series by taking the average of daily figures. We then adopt the convention that they become available on the last day of the month for the current month. Hence, by mid February, we have financial variables until January. By mid February, we have a preliminary (first) and advanced (second) figure for 2016-Q4 GDP, and all the four relevant releases for 2016-Q3.

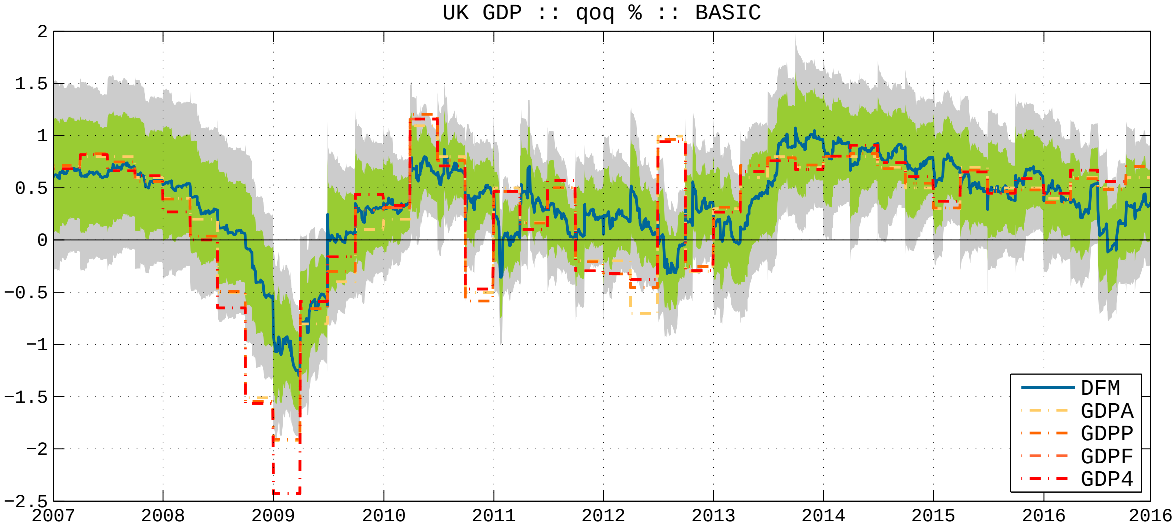
Our sample starts in January 1992, and we evaluate the models performance over the 10-year subperiod 2007-Q1:2016-Q4 in a pseudo out-of-sample forecasting exercise. We use an expanding time window to estimate the parameters of the state space at the beginning of every new calendar year in the evaluation sample. Forecast updates within the year are then computed using the parameters estimated at the beginning of the year, i.e. conditional on information accumulated up to the end of the previous year.

Our benchmark specification includes a subset ($n = 23$) of the variables in Table 1 and three factors ($r = 3$).[3] We set $p = 2$. The predictors that we consider can be broadly classified in four categories. Namely, real variables, nominal and prices, surveys, and financial. As part of our robustness analysis we consider a model specification in which we have one global factor and one factor specific to each of these categories.[4] Also, we

[3]Variables included are: Industrial Production, Manufacturing Production, Index of Services, LFS Number of Employees, Claimant Count Rate, Retail Sales ex Fuel, Exports Volume (Goods), Imports Volume (Goods), CPI All Items, RPI All Items, Net Consumer Credit, PMI Manufacturing, PMI Services, PMI Construction, Lloyds Business Barometer, CBI Industrial Trends, UK Focused Equity Index, Sterling Effective Exchange Rate, Term Spread, GDP Preliminary Release, GDP Second Release, GDP Tjhird Release, GDP Fourth Release.

[4]This is accomplished by imposing zero restrictions in the matrix of loadings (see Eq. (A.5) in Appendix A), and a block diagonal structure for the upper-left blocks of the state-equation matrices in Eq. (A.7) and (A.8).

Figure 3: Nowcast of $y_t^{(1)}$ over the Evaluation Sample

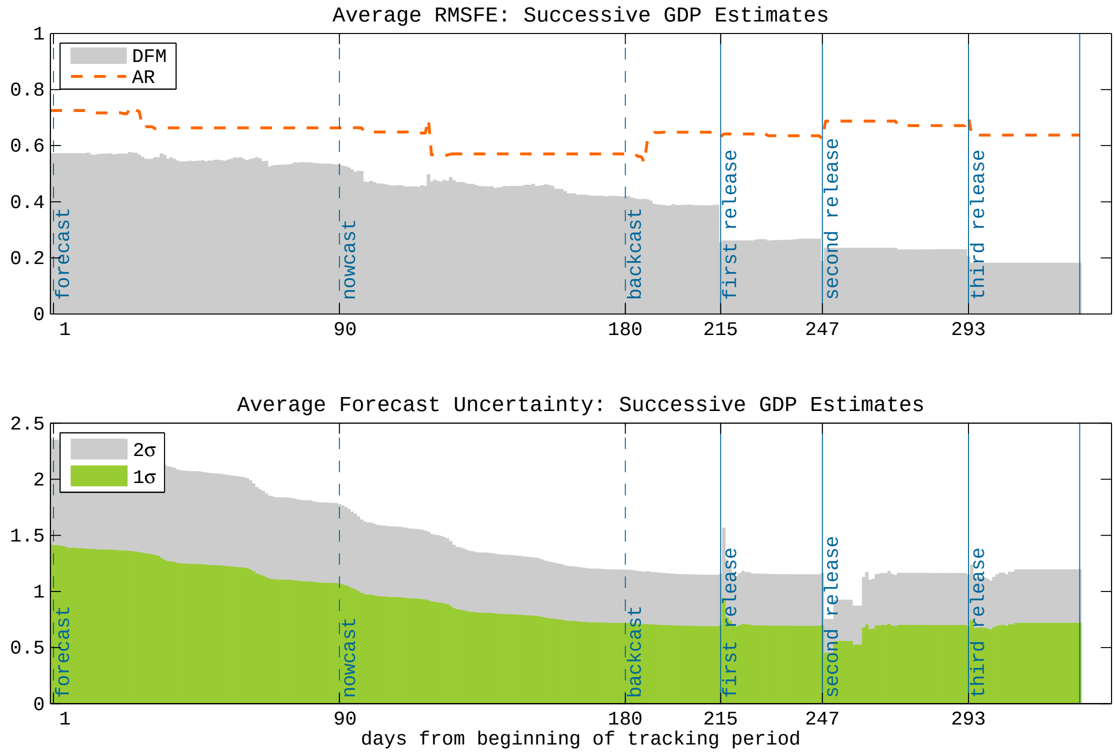


Note: TBA

evaluate the marginal contribution of the nominal and financial blocks of variables by removing them from the benchmark specification. Using US or EA? or both?, cite show that these variables tend to have small relevance compared to real variables and surveys when it comes to nowcasting real GDP growth.

Average performance over the evaluation sample is in Figures 3 and **??**

Figure 4: Average Performance over the Evaluation Sample

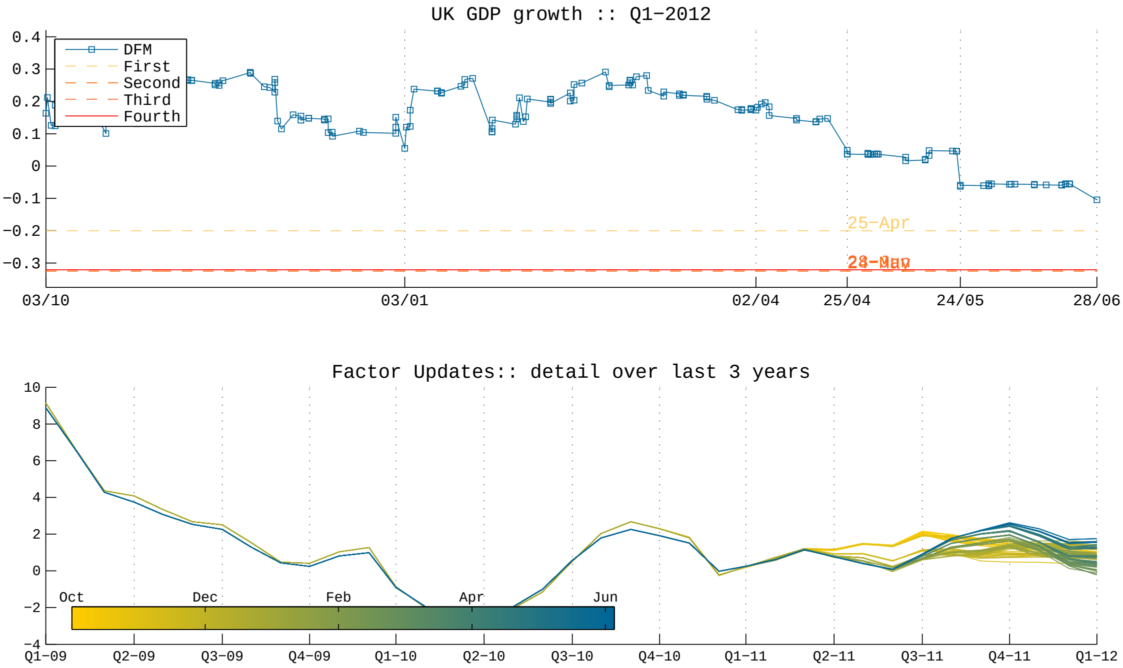


Note: TBA

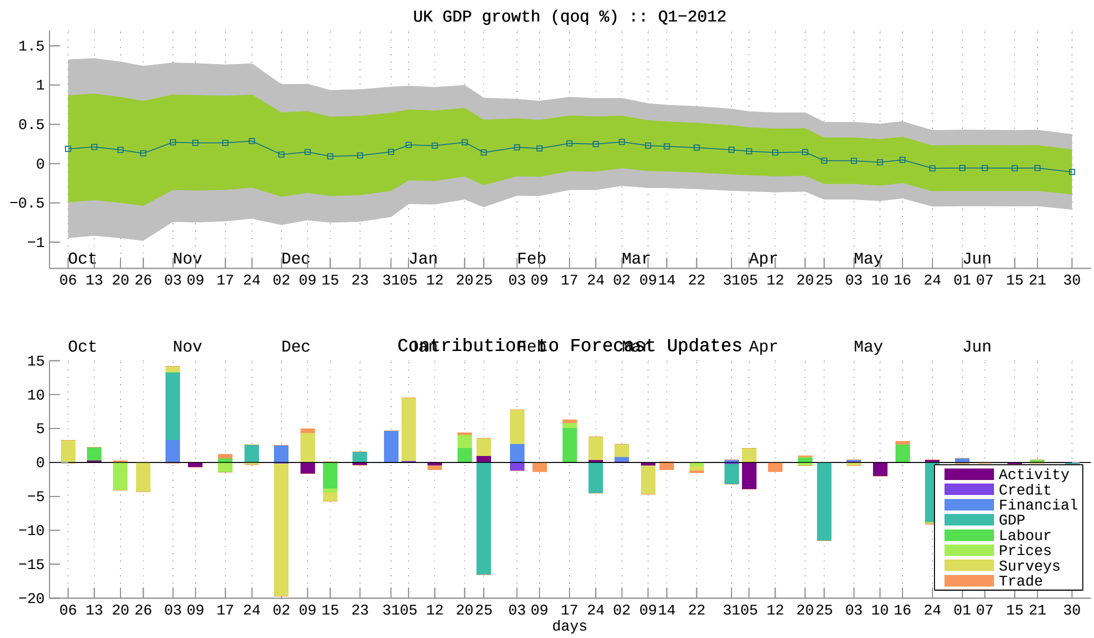
As an illustration, we pick two interesting (for different reasons) quarters to examine the models predictions. The first one is the (non-existing) double dip recession of 2012Q1 (see Figure **??**). Following a major annual revision of historic economic data in June 2013, the ONS revised their GDP figures and showed that output flat-lined in the first three months of 2012 rather than contracting. This meant the economy did not suffer the two consecutive quarters of contraction which is commonly defined a recession. Given the relevance of this event to what the model is designed to produce, its interesting to focus on this quarter and analyse the models predictions. The evolution of the nowcast over time does not suggest growth going into negative territory, rather approaching 0% over time as we move to more mature estimates of GDP, which provides us with some evidence of the model's ability to extract signal and filtering out the noise that might be part of the early statistical releases.

The other one is the EU Referendum in Figure 6. In the aftermath of the EU Referendum in the UK, there has been a significant amount of interest in its impact on the economic activity for both market participants and the central bank. Given that, it would be interesting to see what the model would have predicted at the time. Two main features appear from Figure 6: first, that any weakness of the model prediction during that period was mainly driven by the widely followed Markit/CIPS (PMI) data in July (contributed negatively in the surveys block) but recovered the next month. Also, weak readings of real variables before the Referendum (June) contributed negatively in the nowcast of the quarter, which however recovered very quickly, confirming the ONS preliminary estimate of quarterly GDP growth in that quarter of 0.5%.

Figure 5: The Double-Dip Recession

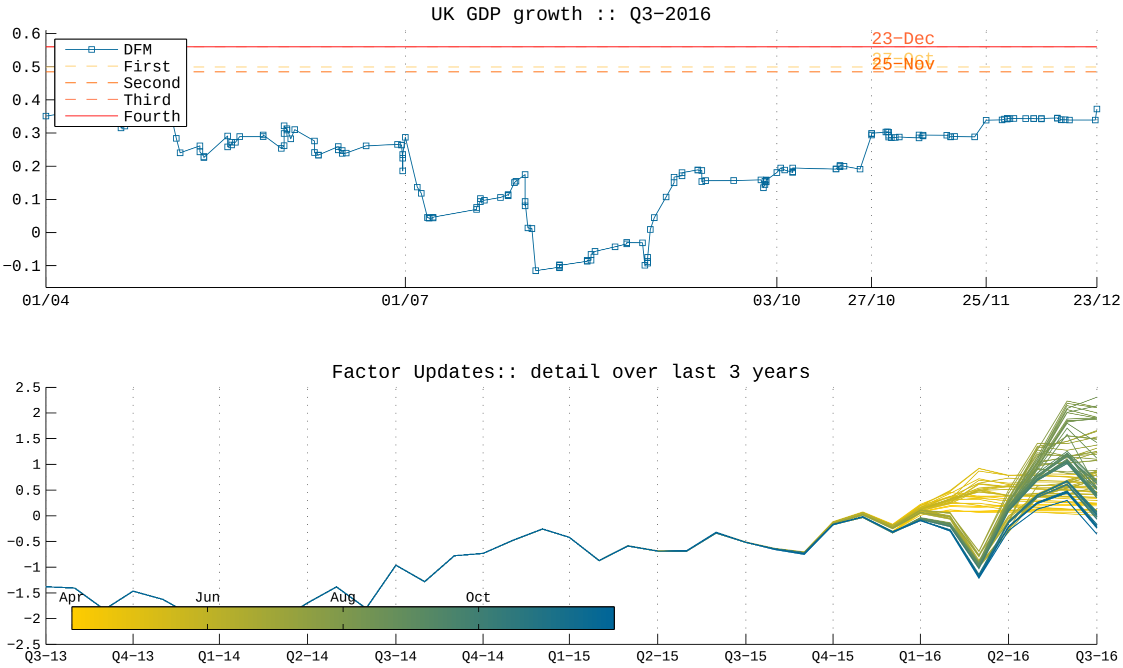


(a) *Note*: TBA

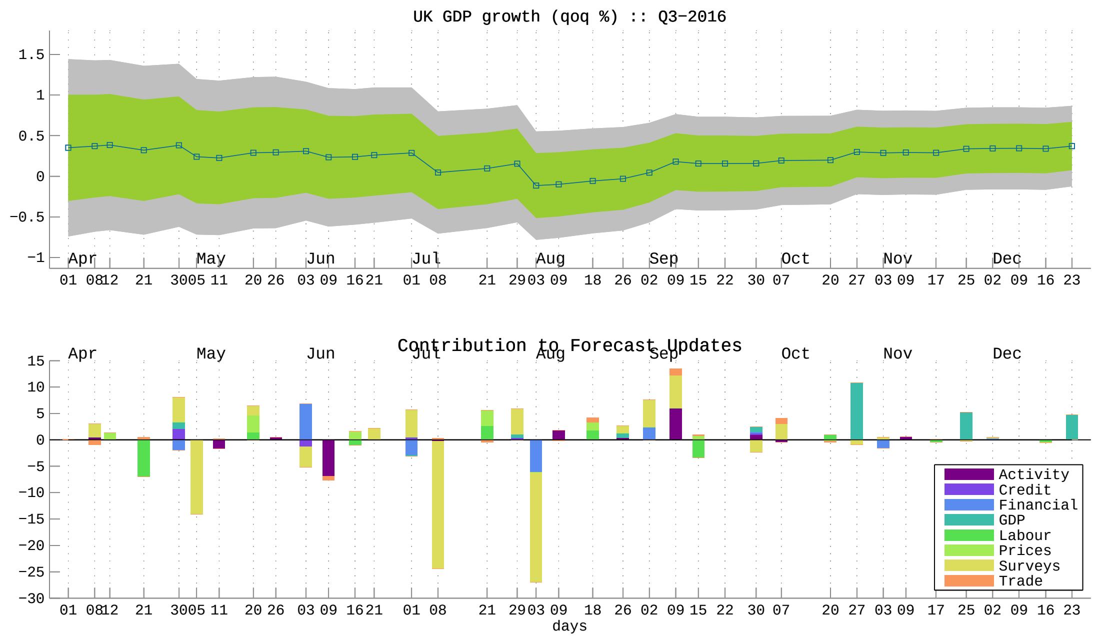


(b) *Note*: TBA

Figure 6: The EU Referendum



(a) *Note*: TBA



(b) *Note*: TBA

4.2 Real-Time Forecasting Exercise

TBA

5 Conclusions

TBA

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.

Ba?bura, Marta and Michele Modugno (2014) "Maximum Likelihood Estimation Of Factor Models On Datasets With Arbitrary Pattern Of Missing Data," *Journal of Applied Econometrics*, Vol. 29, No. 1, pp. 133–160, 01.

Bańbura, Marta and Michele Modugno (2010) "Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data," Working Paper Series 1189, European Central Bank.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010) "Nowcasting," Working Paper Series 1275, European Central Bank.

Bańbura, M., M., 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.

Clements, M. P. (2015) "Assessing macro uncertainty in real-time when data are subject to revision,"Technical report, University of Reading.

Clements, M. P. and A. B. Galvão (2012) "Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models," *Journal of Business and Economic Statistics*, Vol. 30, pp. 554–562.

——— (2013) "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 (2014) "Anticipating Early Data Revisions to US Output Growth and the Effects of Releases on Equity Markets," *ICMA Centre Working Paper Series*, Vol. 2014-6.

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.

Diebold, Francis X., Frank Schorfheide, and Minchul Shin (2016) "Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility," Working Paper 22615, National Bureau of Economic Research.

Galvao, A. B. (2017) "DSGE Models and Data revisions," *Journal of Econometrics*, Vol. 196, pp. 215–232.

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.

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.

A Estimation and State-Space Representation

A.1 State Space

Rewrite Eq. (7), (2), and (3) as:

$$x_t = \mathbf{C}s_t + e_t \qquad e_t \sim \mathcal{N}(0, \mathbf{R}), \tag{A.1}$$

$$s_t = \mathbf{A}s_{t-1} + v_t \qquad v_t \sim \mathcal{N}(0, \mathbf{Q}), \tag{A.2}$$

where x_t and e_t are $(n \times 1)$, $\mathbf{C}$ is $(n \times n_s)$, s_t and v_t are $(n_s \times 1)$, $\mathbf{R}$ is $(n \times n)$ and $\mathbf{A}$ and $\mathbf{Q}$ are $(n_S \times n_s)$. Moreover:

- $n = n_M + n_Q + (k+1)$: number of observables,
- $n_s = n_s^{(f)} + n_s^{(iM)} + n_s^{(iQ)}$: 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,
- $n_s^{(Q)} = n_Q + (k+1)$: number of states for quarterly idiosyncratic. check, q is missing

where q is the number of lags in the polynomial $R(L)$ – see Eq. (4). The full set of observables is:

$$\underset{(n\times 1)}{x_t} = \left(\begin{array}{c:c:c} x_t^{M\prime} & x_t^{Q\prime} & \left(y_t^{(1)}, \ldots, y_t^{(k+1)}\right)' \end{array} \right)'. \tag{A.3}$$

With $p < q$, $q = 4$, $k = 3$, and $nQ = 0$ (i.e. the only quarterly variable is GDP):

$$\underset{(nS\times 1)}{s_t} = \begin{pmatrix} \underset{(n_s^{(f)}\times 1)}{s_t^{(f)}} \\ \hdashline \underset{(n_s^{(M)}\times 1)}{s_t^{(M)}} \\ \hdashline \underset{(n_s^{(Q)}\times 1)}{s_t^{(Q)}} \end{pmatrix} = \begin{pmatrix} \left(f_t' \ \cdots \ f_{t-4}' \right)' \\ \hdashline \left(\xi_{1,t} \ \cdots \ \xi_{n_M,t} \right)' \\ \hdashline \left(\varepsilon_t^{(1)} \ \cdots \ \varepsilon_{t-4}^{(1)} \right)' \\ \vdots \\ \left(\varepsilon_t^{(4)} \ \cdots \ \varepsilon_{t-4}^{(4)} \right)' \end{pmatrix}, \tag{A.4}$$

where the partitions identify (from top to bottom) the states referring to the factors $\left(s_t^{(f)}\right)$, and to the idiosyncratic for the monthly $\left(s_t^{(M)}\right)$ and quarterly $\left(s_t^{(Q)}\right)$ observables.

$$
\underset{(n\times n_s)}{\mathbf{C}} = \left(\begin{array}{cc:c:c}
\Lambda^{(M)} & \mathbf{0}_{n_M\times qr} & \mathbb{I}_{n_M} & \mathbf{0}_{n_M\times n_s^{(Q)}} \\ \hdashline
\Lambda^{(1)}\mathcal{R} & & & \\
\vdots & & \mathbf{0}_{(2k+1)\times n_M} & \mathbb{I}_{2k+1}\otimes\mathcal{R} \\
\Lambda^{(k+1)}\mathcal{R} & & &
\end{array}\right) \tag{A.5}
$$

where $\mathbf{0}_{m\times n}$ denotes an $(m\times n)$ matrix of zeros, $\mathbb{I}_m$ is the identity matrix of dimension m, and $\mathcal{R} = (1\ 2\ 3\ 2\ 1)$.

$$
\underset{(n\times n)}{\mathbf{R}} = \nu\mathbb{I}_n, \tag{A.6}
$$

where ν is a very small number.

$$
\underset{(n_s\times n_s)}{\mathbf{A}} = \left(\begin{array}{c:c:c}
\begin{matrix} A_1 & \dots & A_p & \mathbf{0}_{r\times(q-p+1)} \\ & \mathbb{I}_{r(q-1)} & & \mathbf{0}_{r(q-1)\times r}\end{matrix} & \mathbf{0}_{n_s^{(f)}\times n_s^{(M)}} & \mathbf{0}_{n_s^{(f)}\times n_s^{(Q)}} \\ \hdashline
\mathbf{0}_{n_M\times n_s^{(f)}} & \mathbf{A}^{(M)} & \mathbf{0}_{n_M\times n_s^{(Q)}} \\ \hdashline
\mathbf{0}_{n_s^{(Q)}\times n_s^{(f)}} & \mathbf{0}_{n_s^{(Q)}\times n_s^{(M)}} & \mathbf{A}^{(Q)}
\end{array}\right), \tag{A.7}
$$

where

$$
\underset{(n_s^{(M)}\times n_s^{(M)})}{\mathbf{A}^{(M)}} = \text{diag}\left([\rho_1,\dots,\rho_{n_M}]\right),
$$

and

$$
\underset{(n_s^{(Q)}\times n_s^{(Q)})}{\mathbf{A}^{(Q)}} = \text{blkdiag}\left(\begin{matrix}
\begin{pmatrix} \rho^{(1)} & \mathbf{0}_{1\times(q-1)} \\ \mathbb{I}_{q-1} & \mathbf{0}_{(q-1)\times 1}\end{pmatrix}, \\
\vdots \\
\begin{pmatrix} \rho^{(k+1)} & \mathbf{0}_{1\times(q-1)} \\ \mathbb{I}_{q-1} & \mathbf{0}_{(q-1)\times 1}\end{pmatrix},
\end{matrix}\right).
$$

ρ_i $i = 1,\dots,n_M$ denotes the the autoregressive coefficients for the monthly idiosyncratic, and $\rho^{(i)}$ $i = 1, \ldots, k+1$ those for the $k+1$ GDP releases. Finally,

$$\underset{(n_s \times n_s)}{\mathbf{Q}} = \left(\begin{array}{c:c:c} \begin{matrix} \Sigma_u & \mathbf{0}_{r \times rq} \\ \multicolumn{2}{c}{\mathbf{0}_{r(q-1) \times n_s^{(f)}}} \end{matrix} & \mathbf{0}_{n_s^{(f)} \times n_s^{(M)}} & \mathbf{0}_{n_s^{(f)} \times n_s^{(Q)}} \\ \hdashline \mathbf{0}_{n_s^{(M)} \times n_s^{(f)}} & \mathbf{Q}^{(M)} & \mathbf{0}_{n_s^{(M)} \times n_s^{(Q)}} \\ \hdashline \mathbf{0}_{n_s^{(Q)} \times n_s^{(f)}} & \mathbf{0}_{n_s^{(Q)} \times n_s^{(M)}} & \mathbf{Q}^{(Q)} \end{array}\right), \tag{A.8}$$

where

$$\underset{(n_s^{(iM)} \times n_s^{(M)})}{\mathbf{Q}^{(M)}} = \text{diag}\left(\left[\sigma^2_{\epsilon_1}, \ldots, \sigma^2_{\epsilon_{n_M}}\right]\right),$$

and

$$\underset{(n_s^{(Q)} \times n_s^{(Q)})}{\mathbf{Q}^{(Q)}} = \text{blkdiag}\left(\begin{array}{c} \left(\begin{array}{cc} \sigma^2_{\epsilon^{(1)}} & \mathbf{0}_{1 \times (q-1)} \\ \multicolumn{2}{c}{\mathbf{0}_{(q-1) \times q}} \end{array}\right), \\ \vdots \\ \left(\begin{array}{cc} \sigma^2_{\epsilon^{(k+1)}} & \mathbf{0}_{1 \times (q-1)} \\ \multicolumn{2}{c}{\mathbf{0}_{(q-1) \times q}} \end{array}\right) \end{array}\right).$$

Σ_u is the variance of the residuals of the factors VAR (see Eq. 1 to 3). $\sigma^2_{\epsilon_i}$ $i = 1, \ldots, n_M$ is used to denote the variance of the monthly idiosyncratic, and $\sigma^2_{\epsilon^{(i)}}$ $i, \ldots, k+1$ that of the idiosyncratic of the release processes.

The structure in Eq. (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

Conditional on feasibility, Maximum Likelihood estimation of the MF-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. 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 Ba?bura and Modugno, 2014).