Online Appendices for *Nowcasting Euro Area GDP with News Sentiment: A Tale of Two Crises*

Julian Ashwin, Eleni Kalamara and Lorena Saiz

January 2024

A Details on text data and sentiment methods

A.1 Sources

We choose sources with a wide circulation in their respective countries, covering a broad range of political leanings (see Table A.1). The news coverage varies over time (Figure A.1).

Table A.1: News sources

Country	Source	Code	Total articles	Daily Circulation	Political leaning
France	Les Échos[1]	ECHOS	691,287	120,546	economic liberal
	Le Figaro[1]	FIGARO	341,925	313,541	centre-right
	Le Monde[1]	LEMOND	222,260	302,624	centre-left
Germany	Süddeutsche Zeitung[2]	SDDZ	514,284	361,507	centre-left
	Die Welt[2]	DWELT	180,694	165,686	centre-right
	Der Tagesspiegel[2]	TAGSS	71,095	113,716	liberal
	German Collection[3]	GERCOL	67,841	-	-
Italy	Corriere della Sera[4]	CORDES	412,944	258,991	liberal
	La Repubblica[4]	LAREP	263,339	176,010	progressive
	Il Sole 24 Ore[4]	SOLE	605,480	145,685	liberal
	La Stampa[4]	STMA	216,146	115,870	social liberal
Spain	Expansión[5]	EXPNSI	634,659	50,180	liberal conservative
	El Mundo[5]	MUNDO	174,651	248,463	liberal conservative
	El País[5]	PAISN	354,613	359,809	centre-left
	La Vanguardia[5]	VNGDIA	243,611	180,939	liberal

1: Circulation from `https://en.wikipedia.org/wiki/List_of_newspapers_in_France`, 3-Feb-2021.

2: Circulation from `https://en.wikipedia.org/wiki/List_of_newspapers_in_Germany`, 3-Feb-2021.

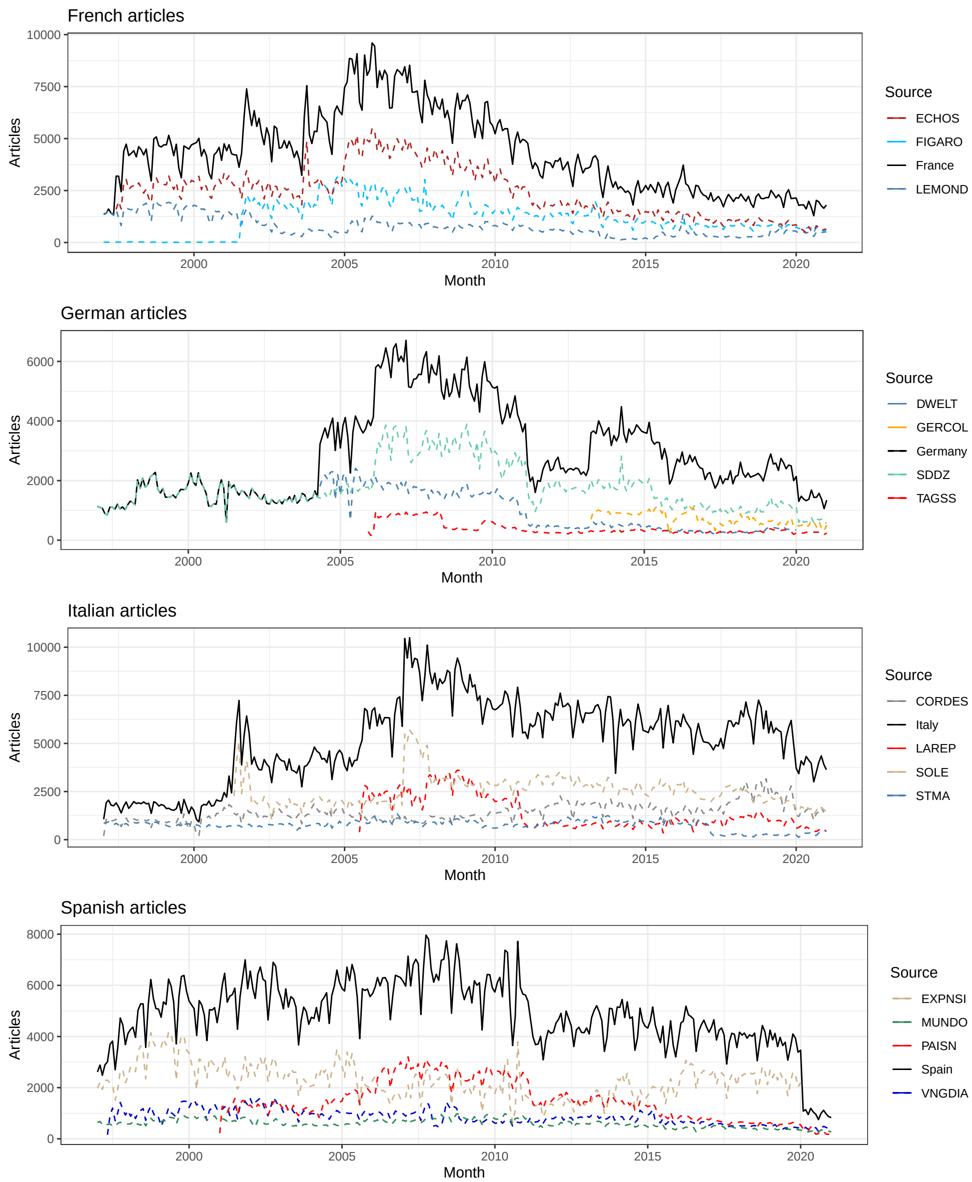
3: Collection of abstracted company, industry and financial news from the leading German general, business and financial newspapers including *Boersen-Zeitung*, *Handelsblatt*, *Süddeutsche Zeitung* and *Frankfurter Allgemeine Zeitung*.

4: Circulation from `https://en.wikipedia.org/wiki/List_of_newspapers_in_Italy`, 3-Feb-2021.

5: Circulation from `https://en.wikipedia.org/wiki/List_of_newspapers_in_Spain`, 3-Feb-2021.

Figure A.1 shows the number of articles taken from each source over time for each country. The coverage is generally fairly consistent over time, with the notable exception of Germany where only the Süddeutsche Zeitung (SDDZ) is available until 2004.

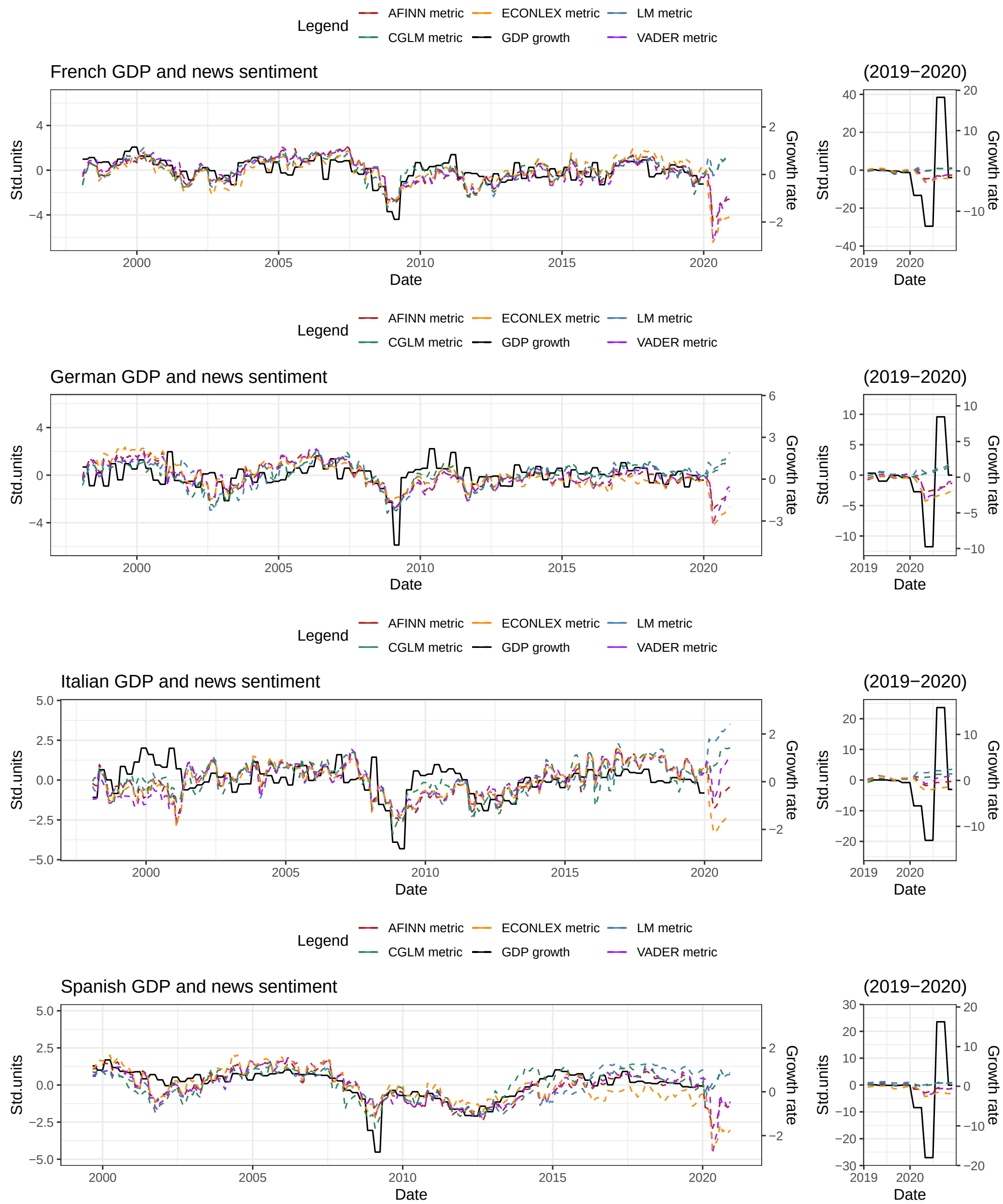
Figure A.1: Articles across time



A.2 Sentiment metric plots

Figure A.2 shows the country-specific sentiment series across the sample period. These charts are the country-specific versions of the Euro Area metrics shown in Figure 1.

Figure A.2: Country-level sentiment series



A.3 Translation methodologies

Figures A.3 to A.6 show some randomly chosen example snippets of articles translated from the original language into English, verifying that the translations are generally high quality enough to preserve the meaning of articles.

Figure A.3: French Translation Example

(a) Original French

Aquoi ressemblerait un groupe français qui résulterait de la fusion de Renault, d'Accor, de Capgemini, de Danone et d'Arcelor? Probablement à un monstre ingouvernable et rapidement promis au déclin. A l'heure où même une entreprise beaucoup moins diversifiée comme Veolia s'interroge sur son avenir, le profil de Tata laisse songeur.

(b) English Translation

What would a French group look like resulting from the merger of Renault, Accor, Capgemini, Danone and Arcelor? Probably to an ungovernable monster and quickly on the verge of decline. At a time when even a much less diversified company like Veolia is wondering about its future, Tata's profile leaves one wondering.

Figure A.4: German Translation Example

(a) Original German

Bundeskanzlerin Angela Merkel lässt offen, ob von deutschen und europäischen Sanktionen gegen Russland wegen der Vergiftung des Oppositionspolitikers Alexej Nawalnyj auch russische Gaslieferungen oder das Pipeline-Projekt Nord Stream 2 betroffen sein könnten. Am Donnerstag äußerte sie sich in Berlin nach einem Treffen mit dem schwedischen Ministerpräsidenten Stefan Löfven. Forderungen, die umstrittene Pipeline Nord Stream 2 nicht fertigzustellen, wurden aus mehreren Parteien laut. Sie kamen von den Grünen, der FDP und der CDU

(b) English Translation

Chancellor Angela Merkel leaves open whether the German and European sanctions against Russia for poisoning opposition politician Alexej Navalnyj could also affect Russian gas deliveries or the Nord Stream 2 pipeline project. On Thursday, she made a statement in Berlin after a meeting with the Swedish Prime Minister Stefan Löfven. Demands not to complete the controversial Nord Stream 2 pipeline were voiced by several parties. They came from the Greens, the FDP and the CDU

Figure A.5: Italian Translation Example

(a) Original Italian

Per migliorare efficienza e distribuzione servirebbero 80 euro per abitante: ora sono soltanto 34; Deficit di investimenti del 60%. Il surplus di polemiche politiche e il deficit di investimenti sono le due caratteristiche strutturali della gestione dell'acqua in Italia, e aprono le falle di una rete idrica che ormai arriva a perdere il 40% dell'acqua immessa nei tubi e di una rete di depurazione che ancora dimentica circa il 20% degli italiani.

(b) English Translation

To improve efficiency and distribution, 80 euros per inhabitant would be needed: now there are only 34; Investment gap of 60%. The surplus of political controversies and the investment deficit are the two structural characteristics of water management in Italy, and open the holes in a water network that is now losing 40% of the water fed into the pipes and in a network of purification that still forgets about 20% of Italians.

Figure A.6: Spanish Translation Example

(a) Original Spanish

El grupo farmacéutico suizo Novartis registró un beneficio neto de 1.477 millones de dólares (1.128 millones de euros) en el primer trimestre, un 16% más. Sus ventas aumentaron un 11% hasta 7.341 millones de dólares. En el mismo periodo, la norteamericana Merck ganó 1.370 millones de dólares, con una caída del 15,4%. Sus ventas se redujeron un 5%, hasta 5.360 millones de dólares.La también estadounidense Schering-Plough obtuvo un beneficio neto de 127 millones de dólares hasta marzo, frente a las pérdidas de 73 millones de dólares del mismo periodo del año anterior.La cifra de negocio fue de 2.369 millones de dólares, un 21% más.

(b) English Translation

The Swiss pharmaceutical group Novartis posted a net profit of 1,477 million dollars (1,128 million euros) in the first quarter, up 16%. Its sales increased 11% to $7,341 million. In the same period, the North American Merck earned 1.37 billion dollars, with a fall of 15.4%. Its sales fell by 5%, reaching 5.36 billion dollars. The also American Schering-Plow obtained a net profit of 127 million of dollars until March, compared to the losses of 73 million dollars in the same period of the previous year. The turnover was 2.369 million dollars, 21% more.

Table A.2 shows the correlations of the daily metrics across the two translation methodologies (using English language dictionaries on translated articles, or translated dictionaries on English language articles).

Table A.2: Correlation of daily metrics across two translation methodologies

Metric	France	Germany	Italy	Spain	Euro area
CGLM	0.670	0.650	0.737	0.845	0.821
LM	0.622	0.576	0.824	0.871	0.813
AFINN	0.691	0.702	0.817	0.814	0.828
HIV	0.654	0.209	0.712	0.573	0.652
NKTGOS	0.482	0.611	0.595	0.715	0.664
HL	0.602	0.792	0.817	0.777	0.817

A.4 VADER details

Unlike polarity-based methods which classify terms or phrases as either positive or negative, valence-based measures take the intensity of the expressed sentiment into account. Valence Aware Dictionary and sEntiment Reasoner (VADER) is based on a lexicon of over 7,500 lexical features, including commonly used abbreviations and emojis. These features are then rated by workers on Amazon Mechanical Turk on a scale from -4 (Extremely Negative) to +4 (Extremely Positive). In addition to this lexicon, VADER also applies general heuristics that affect the intensity or polarity of a sentence. For example, degree modifiers such as "extremely" increase the intensity of sentiment and negation switches the polarity of a sentence. VADER is thus different from pure lexicon based sentiment, as it takes the context of each word into account.

B Comparison to forecasting with text literature

Table B.3: Economic forecasting and nowcasting with text data

Paper	Country	Text methods	Frequency	Target	Benchmark
This paper	Euro Area	English lexicons (see Table 2)	Daily	GDP	ECB projections, PMI DFM, MIDAS, ML macro & financial indicators
Kalamara et al. (2022)	UK	Algorithm-based text metrics	Monthly	GDP, investment, unemployment, inflation	AR(1) model with common factor
Shapiro et al. (2022)	US	VADER with LM, HL and negation rule	Monthly	Consumer sentiment, consumption, ind prod, inflation, employment	LASSO
Larsen and Thorsrud (2019)	Norway	LDA topic model	Quarterly	Output, investment, consumption	AR model, Latent threshold model
Ardia et al. (2019)	US	English lexicons	Monthly	Industrial production	AR model with common factor
Rambaccussing and Kwiatkowski (2020)	UK	LSVM, keyword-based sentiment	Quarterly	Output, inflation unemployment	Model confidence set
Aguilar et al. (2021)	Spain	Spanish lexicon	Monthly	GDP	VAR with economic sentiment
Aprigliano et al. (2022)	Italy	Italian lexicon	Monthly	GDP, investment consumption	Linear model, BMA
Bortoli et al. (2018)	France	French lexicon	Monthly	GDP	AR(1) model with business climate indicator
Barbaglia et al. (2023a)	Germany, France, Italy, Spain, UK	Aspect-based English lexicon	Monthly	GDP	AR model with macro, and financial indicators Double-LASSO
Barbaglia et al. (2023b)	US	Aspect-based English lexicon	Daily	GDP, industrial production, employment, inflation	AR model with macro, and financial indicators, Double-LASSO
Algaba et al. (2023)	Belgium	Dutch and French lexicons	Daily	Consumer confidence	AR(1) model

C Additional Results

C.1 Alternative text aggregation

Figure 5, in Section 3.3 showed the robustness of the value of our news sentiment metrics relative to the PMI are robust to different aggregation approaches for both the text and PMI. Here we provide more detail on the specifications used here as well as showing results for the Great Recession period. The PMI metric are included in three different ways:

1. OLS: Quasi-daily PMI uses the metric described in Section ?? in an OLS regression.

2. OLS: Monthly PMI (-1:3) uses monthly PMI metrics from the third month of the previous quarter and each month of this quarter in an OLS regression, taking the latest vintage and including each month as soon as it becomes available.

3. MIDAS: Monthly PMI (6 lags) uses the 6 latest available values of monthly PMI in each situation and aggregates them using Legendre polynomials with polynomial degree set to two.

The text metrics are included in six different ways:

1. OLS: Daily Cumulative Text is the metric constructed as in Eq 1 and explained in Section 2.2.

2. OLS: Daily Cumulative Text + lag, includes the average sentiment for the third month of the previous quarter along with the daily cumulative metric.

3. OLS: Daily Cumulative Text + diff, includes the difference between the daily average sentiment that day and its value a the end of the previous quarter, along with the daily cumulative metric.

4. OLS: Monthly Text (-1:3) uses monthly average news sentiment from the third month of the previous quarter and each month of this quarter, including each month as soon as it becomes available.

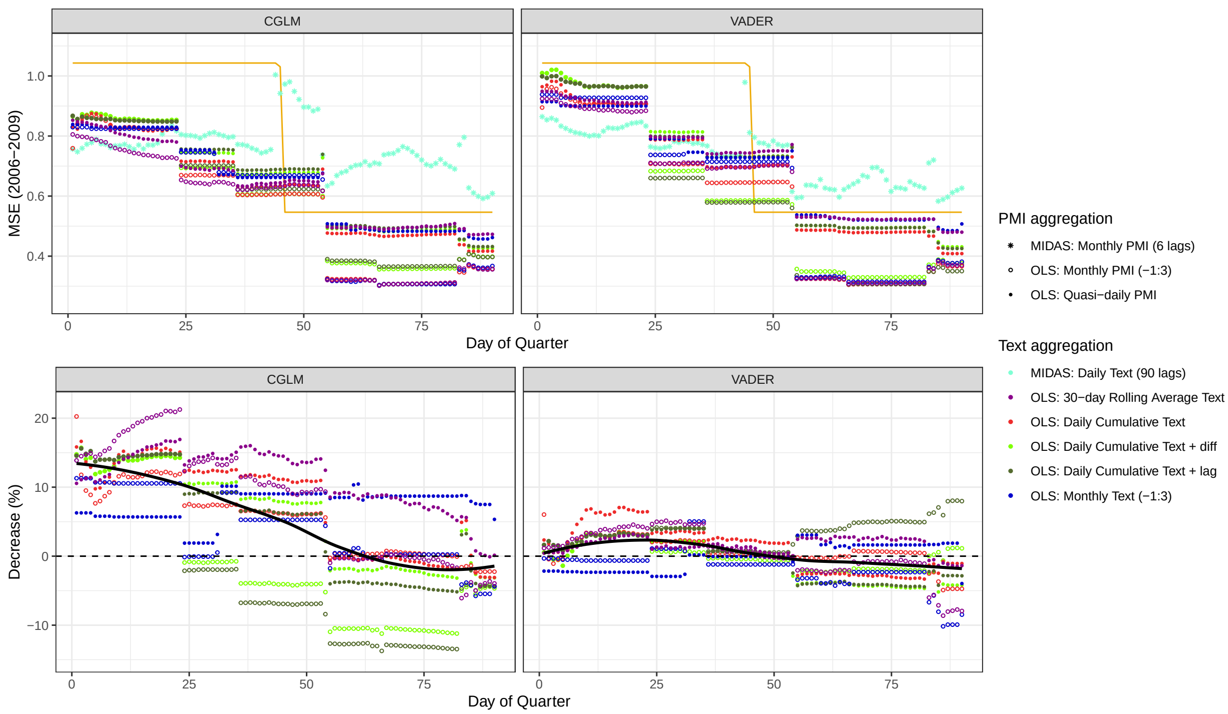
5. OLS: 30-day Rolling Average Text is takes a rolling average over a day window, with the average taken over all words/sentences in that period (so days with very few articles accordingly have a lower weight).

6. MIDAS: Daily Text (90 lags) includes the latest available 90 values of the raw daily text metric and aggregates them using Legendre polynomials with polynomial degree

set to two. This can be seen as an alternative way of aggregating the raw daily text data.

Figure C.7 compares the performance of a model with text to a model with just the PMI across these different aggregation approaches during the Great Recession. This Figure has the same set up as Figure 5, but restricting the sample to April 2006 to December 2009.

Figure C.7: Alternative aggregation approaches for text and PMI (2006-2009)



Notes: This Figure is the same as Figure 5, but restricting the sample to April 2006 to December 2009.

C.2 MIDAS results

Here we give more detail on the results with MIDAS models and explores what might explain their unimpressive performance compared to OLS in our setting. The MIDAS (Mixed Data Sampling) regression (Ghysels et al. (2004)) is an approach in between these two extremes of aggregating daily data to a quarterly indicator and including each day's indicator separately. More specifically,

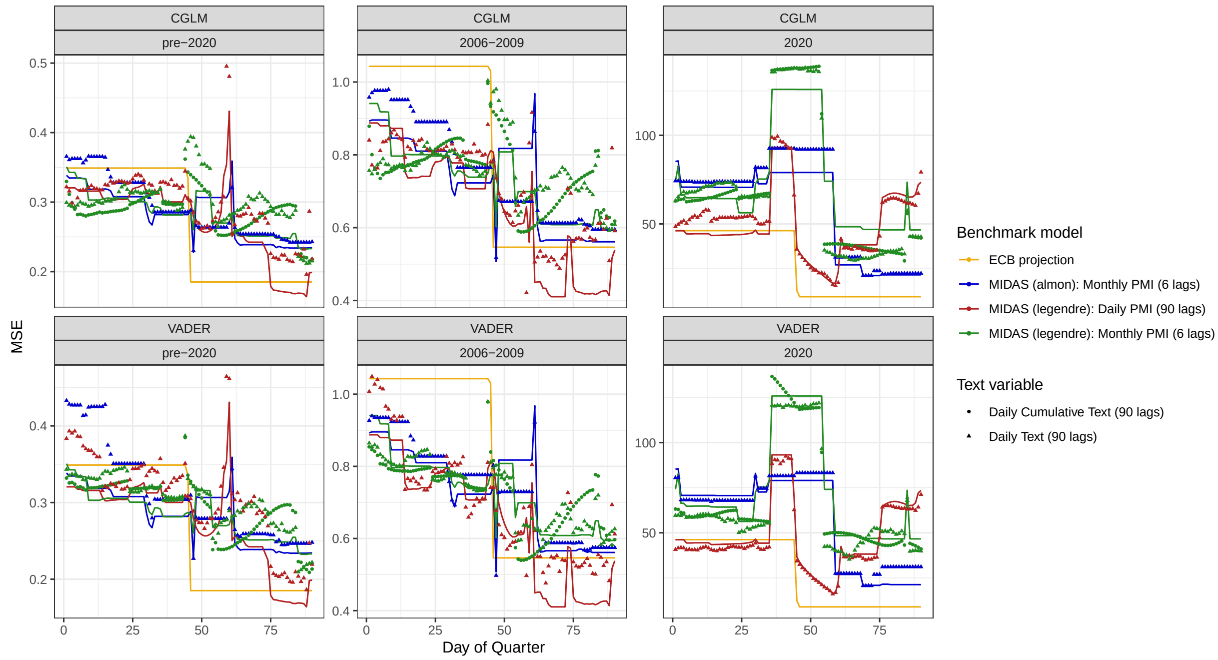
$$y_t^q = \beta_0 + \beta_1 B(L^{1/d};\theta)x_t^d + \varepsilon_t \tag{C.1}$$

where $B(L^{1/d};\theta)$ is a weight function for the K lags of the high frequency indicator x_t^d. Common functions used include Almon lag weighting or polynominal distributed lag, exponential Almon weighting and Legendre polynomial.

Given the better numerical properties, in this paper we have focused on Legendre polynomials up to degree two. The polynomial of degree zero implies all the lags have same weight and it would be equivalent to a simple average. In the polynomial of degree one, each lag has the weight of the lag it refers to (e.g. first lag has a weight of one, second lag a weight of two, etc.). And finally, the polynomial of degree two is 1.5 times the squared lag minus 0.5. We have tested with polynomials of higher order but the results did not improve. We have also considered alternative polynomials (e.g. Almon) and found similar results.

Overall, as can be seen from Figure C.8, the MIDAS PMI benchmarks perform relatively worse than simple OLS regressions from 24 days onwards (i.e. after the first release of PMI data for the first months). We see that across these first 24 days CGLM models substantially outperform the PMI benchmarks both in the pre-2020 sample as a whole and the Great Recession period especially. As in the OLS case, the VADER models show some improvement in the Great Recession period but only marginal benefit across the whole pre-2020 sample. All MIDAS models perform no better than the ECB projection in 2020.

Figure C.8: Comparing text and PMI in MIDAS models

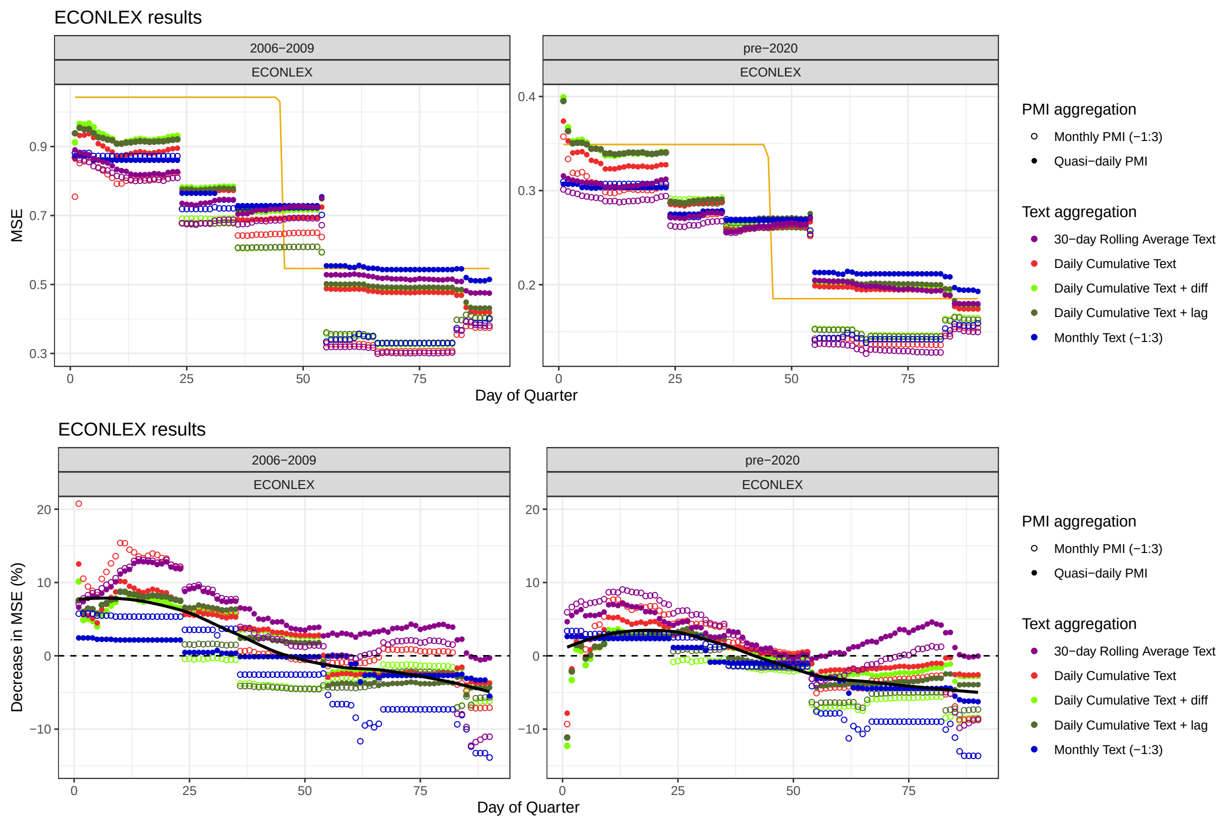


Notes: This Figure compares the performance of a range of MIDAS models with and without text metrics. The upper three panels show results with the CGLM metric and the lower three for the VADER metric. Left-most panels show results across the whole pre-2020 period, central panels focus on April 2006 to December 2019 and the right-most panels show results for 2020. Colors indicate which specification of the MIDAS model is used with the solid lines indicating the performance of the PMI-only benchmark with that MIDAS specification. The points then show performance of the corresponding MIDAS model including text data either as 90 lags of the raw daily metric or 90 lags of the cumulative daily metric.

C.3 ECONLEX results

In general, we find that ECONLEX's performance is between that of CGLM and VADER. Figure C.9 shows the performance of ECONLEX based nowcasts across a range of aggregation approaches and specifications for the PMI based benchmark (mirroring those in Figures 5 and C.7). As in the cases of CGLM and VADER, we see that improvements in performance are greater during the Great Recession period and that results are robust to a range of alternative specifications.

Figure C.9: ECONLEX results with alternative aggregation approaches for text and PMI

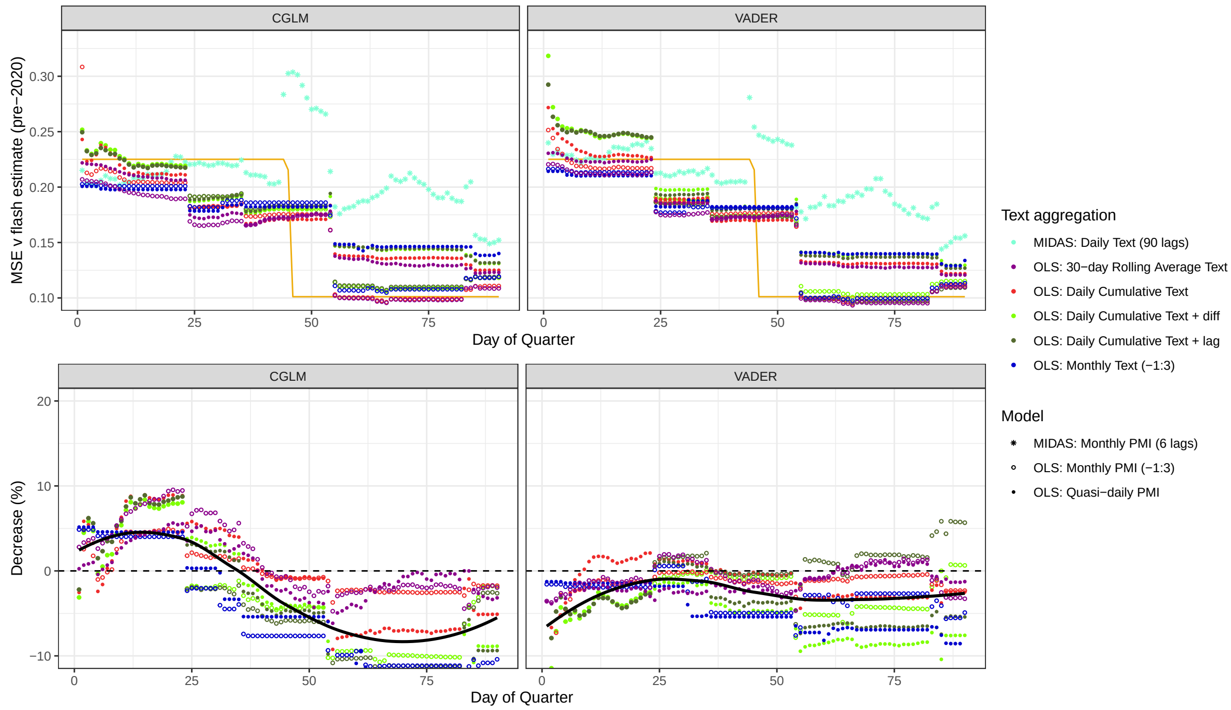


Notes: see Figure 5 for full explanation. The results here use the ECONLEX dictionary to create text metrics. The left panels are for the Great Recession period (April 2006 to December 2009) while the right panels are for the entire pre-2020 sample period.

C.4 Results for flash estimate of GDP

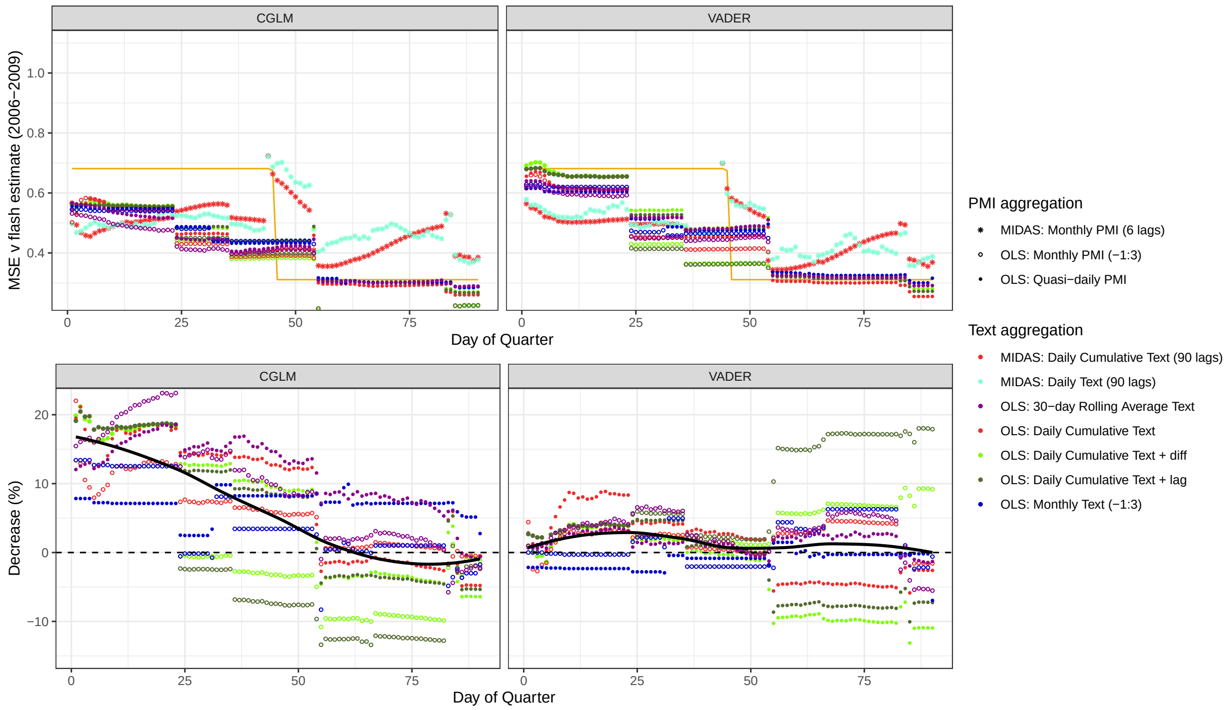
Figures C.10 and C.11 show analogous charts to those in the lower panels of Figures 5 and C.7, but with performance evaluated against the first release of GDP growth rather than the latest available. We see that the results are qualitatively not affected by the vintage of GDP growth used to assess performance.

Figure C.10: Nowcasting performance relative to flash estimate of GDP, pre-2020



Notes: see Figure 5 for full explanation. The results here evaluate performance in forecasting the first available (i.e. flash) vintage of GDP growth. In the interest of space, we only show the relative performance of text to PMI based models and not the MSE compared to the ECB projections. In all cases, both the text and PMI models perform better than the ECB projections in the first half of the quarter.

Figure C.11: Nowcasting performance relative to flash estimate of GDP, 2006-2009



Notes: see Figure C.10 for explanation. These results based only on the sample period April 2006-December 2009.

C.5 Machine Learning models

Below, we present a brief overview of the ML models and their basic properties.

Cross validation strategy. All of our models except OLS, DFM and MIDAS require some form of hyperparameter selection, such as the strength of regularisation parameters, the number of nodes and layers for the NN, or the maximum depth of the leaves for the Forest and Boosting. We take care of not including any future information and perform cross validation only on the in-sample data at each step. In particular, we use Time Series Split from the *sklearn* Python package which is a variation of k-fold targeted to time series (this package is also used to estimate the models themselves). In this approach successive training sets are super-sets of those that come before them. This is consistent with our expanding window evaluation of the out-of-sample test forecasts.

Ridge Regression. Ridge regression is a well-known shrinkage method that combines a linear model with L2 regularization on coefficients.

Random Forest and Boosting. Both Random Forests and Boosting are ensemble methods centred on regression trees. A regression tree is based on consecutively splitting the in-sample dataset until an assignment criterion with respect to the target variable is reached. This allows the fitting of various functional relationships between the target and a set of explanatory variables. A Random Forest combines a large collection of de-correlated trees and then averages across them. Boosting focuses on the predictive power of individual predictors one at a time.[1] The method focuses on the predictive power of individual regressors instead of considering all covariates together. Regressors are chosen sequentially based on their individual ability to explain the target to select the best fitting ones (Friedman, 2001).

Neural Networks. Neural networks (NN) can also incorporate nonlinearity and interaction of variables through a flexible functional form. We use a feed-forward network with ReLu activation functions.[2] The structure of the network can be described by three components: input layer, hidden layer and output layer. In our setting, the input layer corresponds to predictor variables so that the number of neurons in the input layer is the same as the dimension of predictors. The hidden layers converts an output from the preceding layer through a non-linear activation function. Finally, the output layer produces a prediction of the target as a linear combination of the hidden layer's output. A NN becomes more complex and flexible when we increase the number of nodes in a hidden layer (width) or increase the number of hidden layers between input and output layers (depth). Width, depth and penalisation are determined by cross-validation.

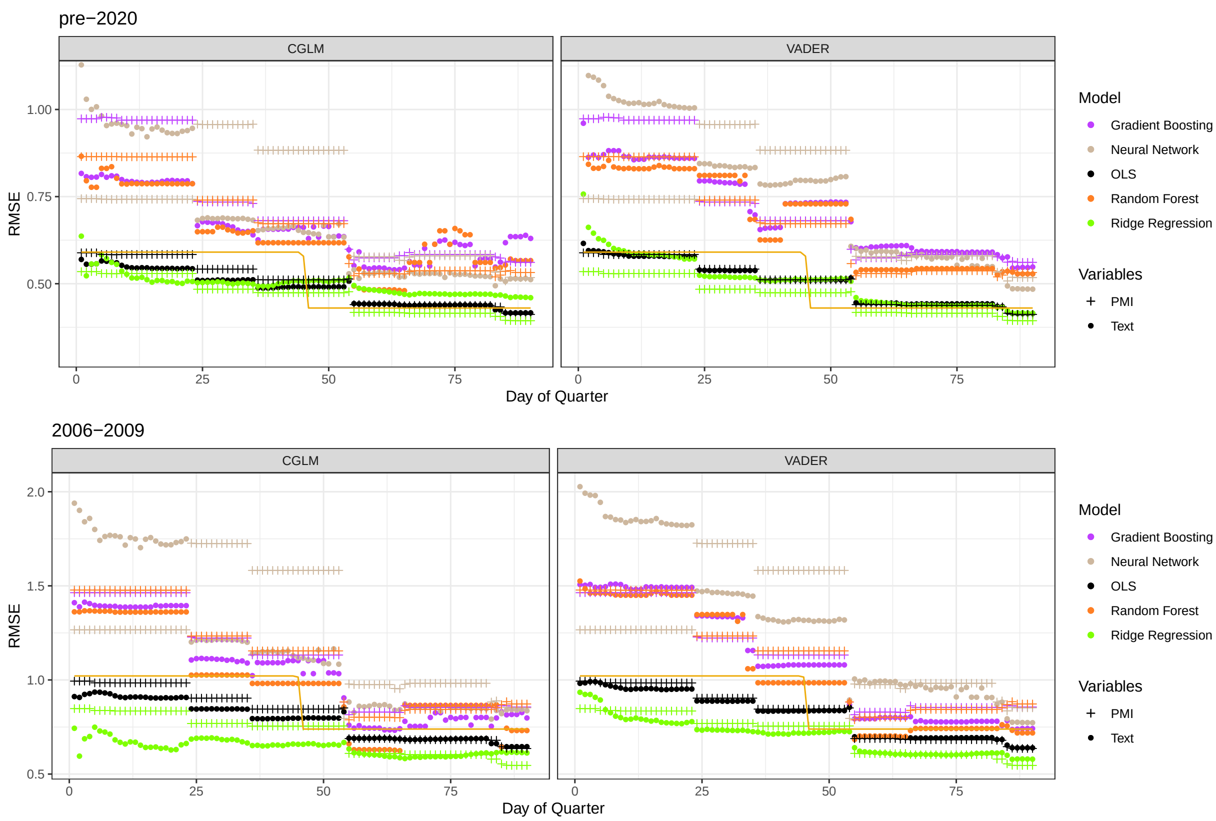
Figure C.12 shows the nowcasting performance of each of these models throughout an

[1] In an economic context, boosting has been applied by Bai and Ng (2009) and Ng (2014)).

[2] While also other activation functions may be considered, *ReLu* remains a preferred choice due to its simple form that facilitates estimation.

average quarter. For each model we show the performance on that day of the quarter of the model with text (circle) and without (cross). For reference the performance of the ECB projections is also shown as a solid yellow line. We see that for both CGLM and VADER, only the ridge regression performs better than the ECB projections across the pre-2020 and Great Recession periods.

Figure C.12: Comparing text and PMI in ML models



Notes: This Figure compares the performance of a range of ML models to OLS and the ECB projections. In each case, we use the daily cumulative text metric and the quasi-daily PMI metric as a benchmark. The horizontal axes in each panel shows the day of the quarter, and the vertical axes show the root mean square error (RMSE) of each model on that day of the quarter. We use the RMSE here as the performance of the ML models varies hugely. The color of each point shows the model used and the shape indicates whether that point corresponds to the performance of the text model or the PMI benchmark. The performance of the ECB projections is shown as the solid yellow line to allow for easy comparison with other figures.

C.6 Macroeconomic and financial indicators

The suite of benchmarks featured in Figure 8 are briefly described here:

1. Dynamic Factor Model: mixed-frequency Dynamic Factor model, using a set of standard monthly macro indicators, business and consumer surveys and financial indicators. These include the following monthly Euro Area indicators: industrial production, construction production, new passenger car registrations, retail trade, external trade, industrial orders, unemployment rate, surveys of the European Commission and the Purchasing Managers' surveys for services, manufacturing and construction, consumer confidence, loans to the private sector (deflated by HICP), monetary aggregate M1 (deflated by HICP), and two financial indicators: Eurostoxx and corporate spreads (i.e. difference between BBB non-financial corporate bond yields and AAA government bond yields).

2. ECB projection: forecast made by ECB/Eurosystem staff using a suite of models and expert judgement.

3. MIDAS: Monthly PMI (6 lags): MIDAS regression using Legendre polynomials up to degree 2 and 6 lags of monthly PMI as high frequency predictor.

4. OLS: Daily Financial: OLS regression between quarterly GDP growth and financial indicators (Eurostoxx and corporate spreads).

5. OLS: Monthly Factors (1:3): OLS regression between quarterly GDP growth and the real-time monthly common factor obtained from the Dynamic Factor model above. As the DFM generates forecasts for future values of the common factor, we use these forecasts until the factor itself is available.

6. OLS: Monthly PMI (-1:3): OLS regression between quarterly GDP growth and monthly PMI indicator from the third month of the previous quarter to the third month of this quarter. Only PMI data that was available in real time is used, so at the beginning of the quarter only the third month of the previous quarter is included

and other months are added (as separate variables) once they become available.

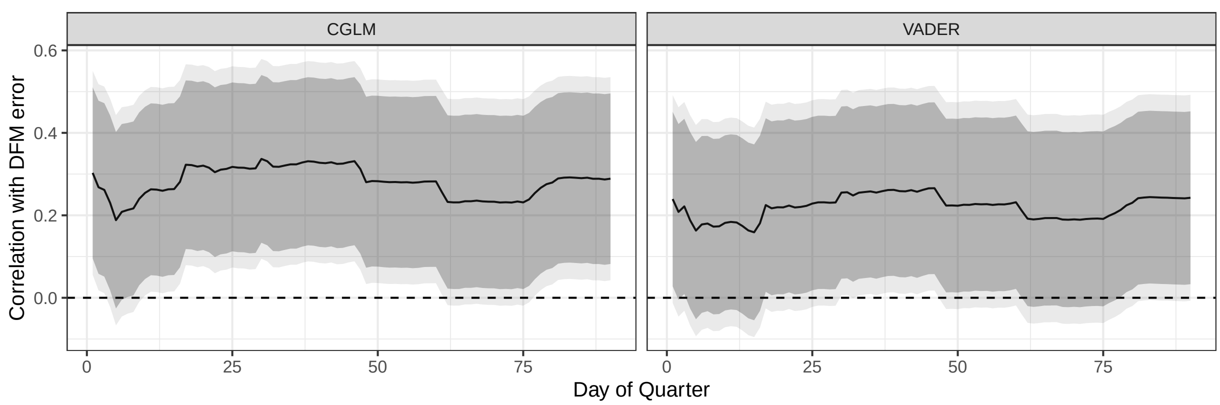
7. OLS: Quarterly Factors: OLS regression between quarterly GDP growth and the quarterly average across real-time monthly common factor obtained from the Dynamic Factor model above.

8. OLS: Quasi-daily PMI: OLS regression between quarterly GDP and the quasi-daily PMI as described in the main text.

9. Ridge: Quasi-daily PMI: Ridge regression between quarterly GDP and the quasi-daily PMI as described in the main text.

Figure C.13 shows the correlation of the DFM nowcast errors with the CGLM and VADER sentiment metrics. These correlations are computed for the whole pre-2020 sample period.

Figure C.13: DFM nowcast error correlation with news sentiment

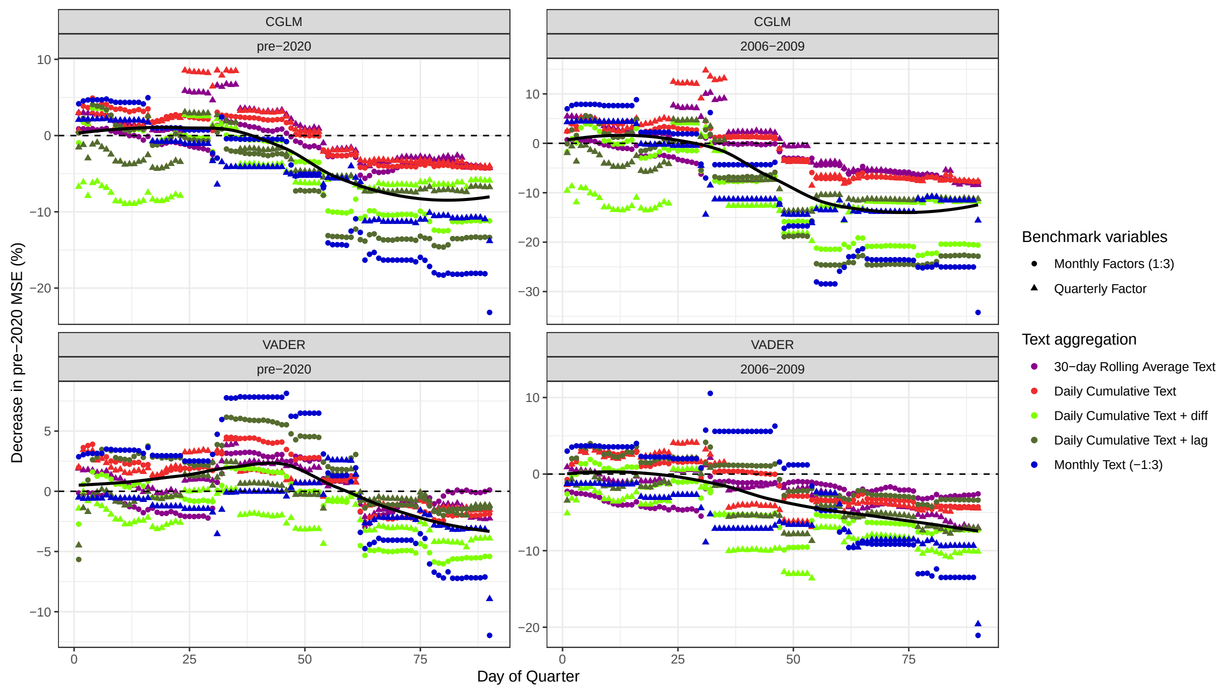


Notes: Correlation of the cumulative daily sentiment metric with the errors made by the DFM throughout the quarter, across the pre-2020 sample period. The day of quarter is shown along the horizontal axes, with the correlation coefficient of daily cumulative sentiment metric at that day with the DFM nowcast error shown on the vertical axes. The shaded bands represent the 90% and 95% confidence intervals. Results with CGLM are shown in the left panel with VADER on the right.

Figure C.14 compares the real-time nowcast performance of models with text indicators to those with the monthly and quarterly factors, across a range of alternative aggregation approaches for the text. We can see that across a range of aggregation approaches both

CGLM and VADER see some improvements in performance during the first half of the quarter, although unsurprisingly these gains are smaller than with respect to just the PMI indicator.

Figure C.14: Comparing text and common factor with alternative text aggregations



Notes: See Figure 5 for full explanation.

C.7 Diebold and Mariano tests

Table C.4: Nowcasting results on a monthly basis for GDP growth using VADER - pre-2020 period

Model	Month	Levels		First differences		Projections		Projections and First Diff.	
		$\frac{RMSE_{text}}{RMSE_{ML\_ben}}$	$\frac{RMSE_{text}}{RMSE_{OLS\_ben}}$	$\frac{RMSE_{text}}{RMSE_{ML\_ben}}$	$\frac{RMSE_{text}}{RMSE_{OLS\_ben}}$	$\frac{RMSE_{text}}{RMSE_{ML\_ben}}$	$\frac{RMSE_{text}}{RMSE_{OLS\_ben}}$	$\frac{RMSE_{text}}{RMSE_{ML\_ben}}$	$\frac{RMSE_{text}}{RMSE_{OLS\_ben}}$
AverageML	1	1.004	1.042	0.891**	0.943*	0.997	1.118	0.991	1.094
AverageML	2	1.033	1.063	0.906**	0.962	0.986	0.987	0.978	0.988
AverageML	3	1.008	1.126	0.883	0.953	0.956*	1.026	0.901	0.965
Boosting	1	1.038	1.068	0.889**	0.965**	1.011	1.123	1.013	1.102
Boosting	2	1.006	1.088	0.856***	0.977	0.983	1.035	0.98	1.016
Boosting	3	0.930**	1.14	0.786***	0.982	0.923	1.087	0.911**	1.044
Forest	1	0.999	1.038	0.881	0.943	0.989	1.111	0.999	1.081
Forest	2	1.005	1.052	0.915	0.996	0.977	0.995	0.974	0.983
Forest	3	0.909*	1.088	0.908	0.999	0.936	1.076	0.917	0.975
NN	1	1.046	1.126	0.985	1.023	0.997	1.159	0.965	1.128
NN	2	1.134	1.131	1.025	1.040	1.017	0.987	0.995	0.998
NN	3	1.384	1.361	1.026	1.034	1.067	1.009	0.898**	0.965
OLS	1	0.981*	1.059	0.918	1.100	1.005	1.005	1.012	1.053
OLS	2	0.997	0.969	0.893	1.132	1.007	1.007	1.007	0.952**
OLS	3	0.984	0.946**	0.81	1.317	1.006	1.006	1.003	0.924
Ridge	1	0.954	0.977	0.943*	0.960	1.036	0.980	0.979	1.008
Ridge	2	1.097	1.130	1.011	1.008	1.027	1.013	0.987	1.049
Ridge	3	1.222	1.294	1.004	1.045	1.053	1.006	0.976	1.127

Notes: Relative RMSEs using the VADER metric and the PMI series as predictors to predict GDP growth compared to only PMI-based model. Results are compared to the relative ML counterpart and the linear model. For the case of the OLS, the columns relative to the 'OLS_ben' correspond to the case where the benchmark is the PMI in levels. Significance of forecast accuracy is assessed via Diebold and Mariano (1995) test statistics with Harvey's adjustment. $***\backslash**\backslash*$ indicates significance at $10\%, 5\%, and 1\%$, respectively.

Table C.5: Nowcasting results on a monthly basis for GDP growth using CGLM - pre-2020 period

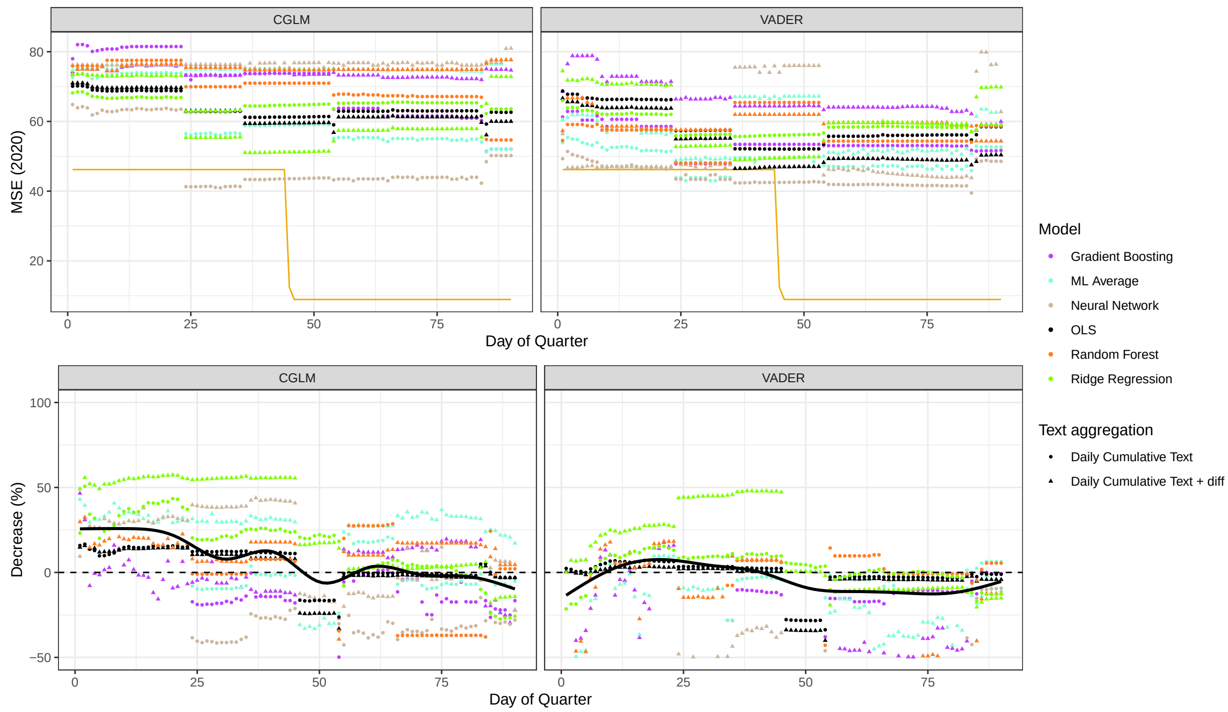
Model	Month	Levels		First differences		Projections		Projections and First Diff.	
		$\frac{RMSE_{text}}{RMSE_{ML\_ben}}$	$\frac{RMSE_{text}}{RMSE_{OLS\_ben}}$	$\frac{RMSE_{text}}{RMSE_{ML\_ben}}$	$\frac{RMSE_{text}}{RMSE_{OLS\_ben}}$	$\frac{RMSE_{text}}{RMSE_{ML\_ben}}$	$\frac{RMSE_{text}}{RMSE_{OLS\_ben}}$	$\frac{RMSE_{text}}{RMSE_{ML\_ben}}$	$\frac{RMSE_{text}}{RMSE_{OLS\_ben}}$
AverageML	1	0.965**	1.082	0.864***	0.914	0.955	1.054	0.918	0.953
AverageML	2	0.987	0.988	0.815	0.866**	0.989*	0.999	0.966	0.994
AverageML	3	1.010	1.085	0.776	0.838**	0.950*	1.017	1.033	1.154
Boosting	1	0.982	1.091	0.857	0.931	0.971	1.056	0.935	0.962
Boosting	2	0.970	1.021	0.760	0.868**	0.975	1.011	0.974	1.054
Boosting	3	1.011	1.191	0.700	0.874**	0.946	1.083	0.998	1.223
Forest	1	0.973	1.092	0.885	0.947*	0.985	1.066	0.918	0.954
Forest	2	0.985	1.003	0.831	0.905**	1.009	1.019	0.932*	0.976
Forest	3	1.079	1.241	0.881	0.970	1.054	1.121	1.108	1.326
NN	1	0.937	1.089	0.883***	0.916**	0.907	1.060	0.903	0.972
NN	2	1.015	0.985	0.892**	0.905*	0.981	0.983	1.010	1.008
NN	3	1.000	0.946*	0.820**	0.826**	0.888**	0.954*	1.048	1.031
OLS	1	0.972	0.972	0.844**	1.011	0.975	1.015	0.921	0.995
OLS	2	1.016	1.016	0.836**	1.060	1.019	0.963	0.970	0.944
OLS	3	1.009	1.009	0.702**	1.143	1.014	0.933*	0.917*	0.881***
Ridge	1	0.990	0.936	0.831***	0.847*	0.903	0.930	0.900*	0.922
Ridge	2	1.074	1.059	0.913*	0.910	1.003	1.066	1.100	1.133
Ridge	3	1.164	1.111	0.880*	0.915*	0.966	1.115	1.323	1.401

Notes: Relative RMSEs using the stability metric and the PMI series as predictors to predict GDP growth compared to only pmi-based model. Results are compared to the relative ML counterpart and the linear model. Significance of forecast accuracy is assessed via Diebold and Mariano (1995) test statistics with Harvey's adjustment. $***\backslash**\backslash*$ indicates significance at $10\%, 5\%, and 1\%$, respectively.

C.8 2020 results

Figure C.15 shows that even when directly predicting GDP growth, both the CGLM and VADER metrics add value over the PMI. However, neither the PMI model nor the PMI plus text perform better than the ECB projections.

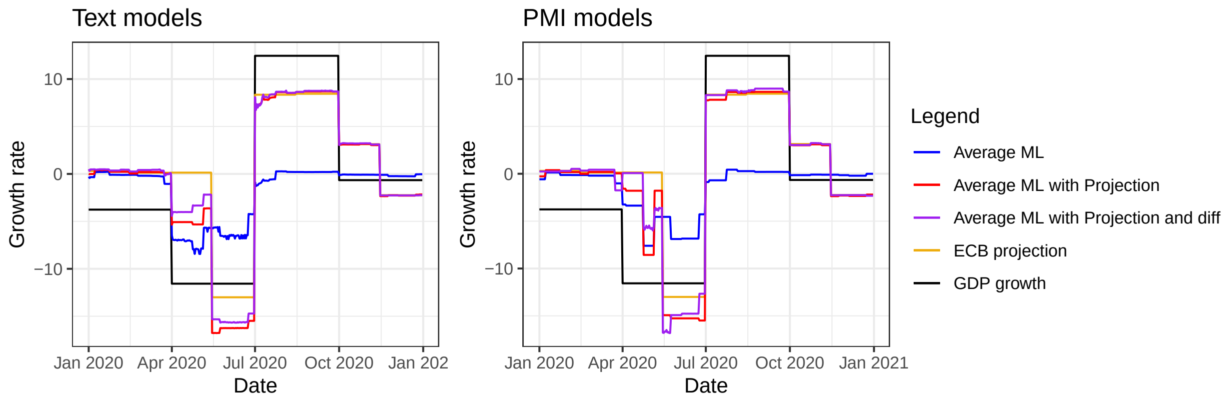
Figure C.15: OLS and ML results in 2020 when targeting GDP growth directly



Notes: This Figure shows MSE across the quarter in 2020 for models that target GDP growth directly, see Figure 5 for a full explanation.

The daily nowcasts with text and PMI models and ML-Averages for alternative specifications are shown in Figure C.16.

Figure C.16: Nowcasts in 2020 (VADER)



Notes: This Figure compares the real-time nowcasts of various text models and PMI models throughout 2020.

References

Aguilar, P., Ghirelli, C., Pacce, M., and Urtasun, A. (2021). Can news help measure economic sentiment? An application in COVID-19 times. *Economics Letters*, 199:109730.

Algaba, A., Borms, S., Boudt, K., and Verbeken, B. (2023). Daily news sentiment and monthly surveys: A mixed-frequency dynamic factor model for nowcasting consumer confidence. *International Journal of Forecasting*, 39(1):266–278.

Aprigliano, V., Emiliozzi, S., Guaitoli, G., Luciani, A., Marcucci, J., and Monteforte, L. (2022). The power of text-based indicators in forecasting italian economic activity. *International Journal of Forecasting*, 39(2):791–808.

Ardia, D., Bluteau, K., and Boudt, K. (2019). Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. *International Journal of Forecasting*, 35(4):1370–1386.

Bai, J. and Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, 24(4):607–629.

Barbaglia, L., Consoli, S., and Manzan, S. (2023a). Forecasting gdp in europe with textual data. *Journal of Applied Econometrics, forthcoming.*

Barbaglia, L., Consoli, S., and Manzan, S. (2023b). Forecasting with economic news. *Journal of Business & Economic Statistics*, 41(3):708–719.

Bortoli, C., Combes, S., and Renault, T. (2018). Nowcasting gdp growth by reading newspapers. *Economie et Statistique*, 505(1):17–33.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

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

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., and Kapadia, S. (2022). Making text count: economic forecasting using newspaper text. *Journal of Applied Econometrics*, 37(5):896–919.

Larsen, V. H. and Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, 210(1):203–218.

Ng, S. (2014). Boosting recessions. *Canadian Journal of Economics/Revue canadienne d'économique*, 47(1):1–34.

Rambaccussing, D. and Kwiatkowski, A. (2020). Forecasting with news sentiment: Evidence with uk newspapers. *International Journal of Forecasting*, 36(4):1501–1516.

Shapiro, A. H., Sudhof, M., and Wilson, D. (2022). Measuring news sentiment. *Journal of Econometrics*, 228(2):221–243.