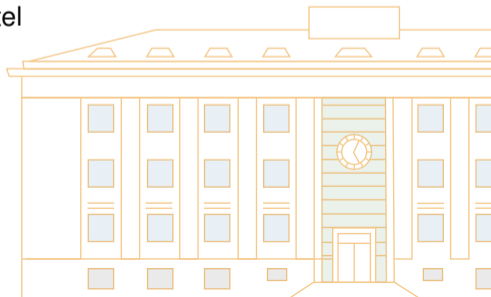


Do daily lead texts help nowcasting GDP growth?

Marc Burri

University of Neuchâtel

August 28, 2023



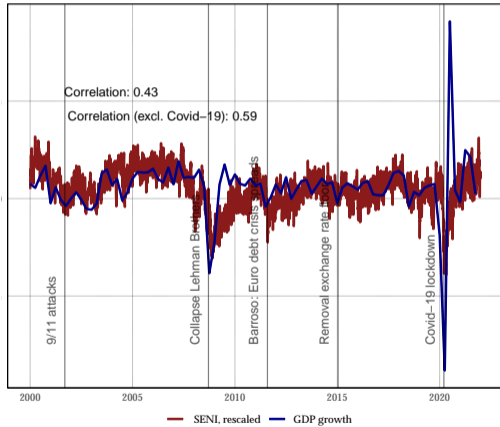
Motivation:

- Need for high-frequency data emerged during Covid-19 pandemic.
- Lead texts from newspaper articles are quickly and freely available online.

Research questions:

- Do they help nowcasting Swiss GDP growth?
- Can they be used to fill the gap when traditional lower frequent indicators are not available?

- A daily composite indicator created from lead texts is highly correlated with Swiss economic activity.
- It outperforms a well-known Swiss business cycle indicator in an out-of-sample evaluation, if one month of information is available.
- Improves nowcasting accuracy, especially during periods of economic distress.



Textual data and news sentiment indicators receive more and more attention

- for measuring and forecasting economic activity (See e.g. Buckman et al. (2020), Thorsrud (2020), Kalamara et al. (2022) or Barbaglia et al. (2022))
- to measure recession perception/economic uncertainty (See e.g. The Economist (2011), Iselin and Siliverstovs (2013), Baker et al. (2016) or Larsen (2021)).

Contribution:

1. Provide a daily measure of Swiss economic activity.
2. Examine the value added of the daily frequency.
3. Evaluate the informational content of publicly available lead texts.
4. Keyword-based method is simple and easy to understand.

- Publicly available titles and lead texts of articles from three Swiss german newspapers.
- Web-scraped from their online archive and filtered for texts containing one of the keywords *Rezession*, *Konjunktur* or *Wirtschaft* (Recession, business cycle, economy).
- Advantage: Data can be updated with a delay of one day and is free of charge.

Economy-related sentiment indicators

- Choose keywords defining economic topics (e.g labor market, inflation etc.).
- Extract keywords, the ten preceding words, and the ten following words.
- Define a list of positive and negative phrases (Remus et al., 2010).
- Calculate a sentiment score by subtracting the counts of negative words from the counts of positive words.
- Daily sentiment indicators for a given topic are calculated as a simple average of the sentiment scores.

Recession indicator

- Choose keywords related to recessions. (e.g recession, crisis)
- Indicator is calculated by counting the appearance of these keywords.

Why creating a single indicator?

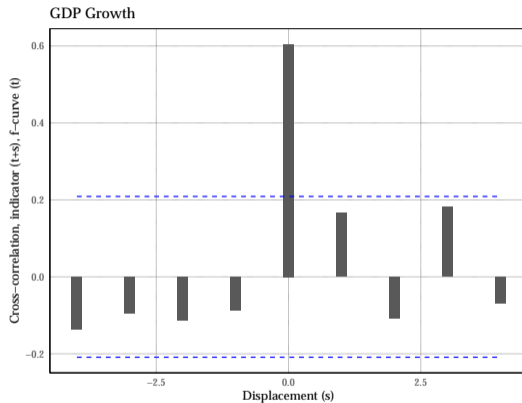
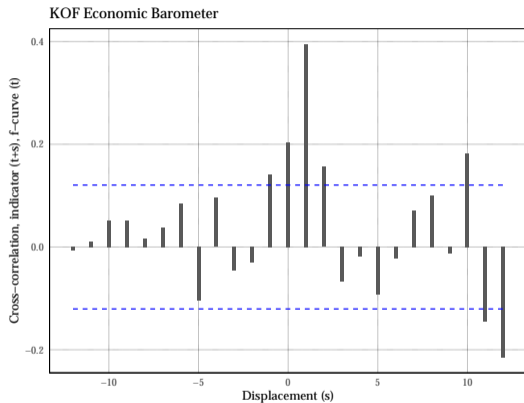
- Need to use models that allow the inclusion of time series of different frequencies (e.g **mixed-data sampling** (MIDAS) and bridge models).
- Problematic to include many explanatory variables in a MIDAS model, as this could lead to parameter proliferation.
- The daily indicators are quite volatile, and they are correlated with each other.

Methodology

- Estimate a factor model in static form

$$X = F\Lambda + e$$

- X = The data matrix
- F = the common factors
- Λ = the factor loadings
- e = the unexplained error term
- First principal component can be interpreted as a coincident business cycle indicator



Notes: Before computing the cross-correlation the series have been pre-whitened with an AR(p) model (see Neusser, 2016, Ch. 12.1).

- Variable of interest is quarterly GDP growth, y_{t_q} .
- Make a daily nowcast for y_{t_q} on an expanding window.
- Real-time data set for quarterly GDP vintages by Indergand and Leist (2014)
- For simplicity I assume every quarter to have $D = 60$ days, reflecting five working days per week and four weeks per month.
- The sample spans from January 1, 2000, to December 31, 2021
- Three different models: MIDAS, Bridge and Midas-IT.
- Benchmarks: AR(1), Bridge model with KOF Barometer and GDP first release.

Three models:

$$\text{MIDAS: } y_{t_q+H+1} = \alpha + \sum_{p=0}^{P-1} \beta_p \sum_{k=0}^{K-1} b(k, \theta) L^{(pD+k)/D} x_{t_d+T_x-T_y} + \varepsilon_{t_q+H+1}$$

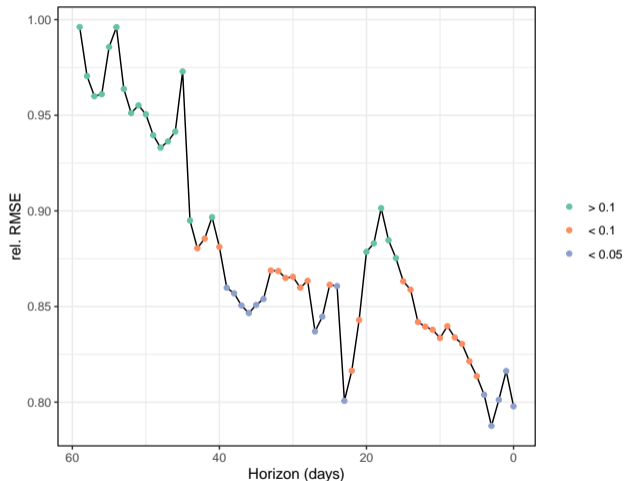
$$b(k, \theta) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{j=0}^{K-1} \exp(\theta_1 j + \theta_2 j^2)}$$

$$\text{BRIDGE: } y_{t_q+H+1} = \alpha + \sum_{p=0}^{P-1} \beta_p L^p x_{t_q+H+1} + \varepsilon_{t_q+H+1} \quad x_{t_q} = \sum_{k=0}^{K-1} \omega(k) L^{k/D} x_{t_d} = \sum_{k=0}^{K-1} \omega_k L^{k/D} x_{t_d}$$

Midas-IT: Mixture of both.

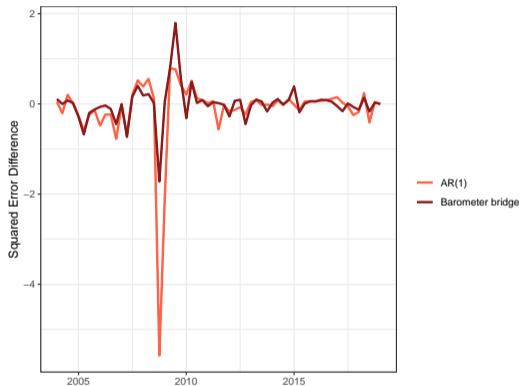
Horizon	Full sample				Excluding Covid crisis				Excluding all crisis periods			
	0	39	79	119	0	39	79	119	0	39	79	119
a) Hypothesis: RMSE Model < RMSE AR(1) model												
Bridge	0.67	0.86**	1.01	1.1	0.87*	0.81*	1.02	1.03	0.86**	0.78*	1.01	0.97
Midas	0.69	0.92	0.99	1.07	0.78**	0.92	1.11	1.01	0.74***	0.9	1.11	0.97
Midas-IT	0.72	0.86*	0.98	1.07	0.8**	0.78**	0.99	0.97	0.77***	0.73**	0.99	0.89
b) Hypothesis: RMSE Model < RMSE Barometer bridge												
Bridge	1.07	0.96	0.96	1.1	0.87*	0.9	1	1.05	0.76**	0.82**	0.94	0.99
Midas	1.12	1.01	0.94	1.07	0.78**	1.01	1.09	1.03	0.66***	0.95	1.03	0.99
Midas-IT	1.17	0.96*	0.93	1.06	0.8**	0.86**	0.98	0.98	0.68***	0.77**	0.92	0.9
c) Hypothesis: RMSE Model > RMSE First release												
Bridge	2.03*				1.2				1.22**			
Midas	2.13*				1.09				1.08			
Midas-IT	2.22*				1.08				1.05			

Notes: Relative Root Mean Squared Errors and Diebold-Mariano-West tests for equal predictive accuracy against alternative hypothesis in row header. Significance levels are given by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$



Notes: RMSE and Diebold-Mariano-West tests for equal predictive accuracy against alternative hypothesis

RMSE Midas-IT < RMSE Barometer bridge. Significance levels are given by: ● $p < 0.1$, ● $p < 0.05$.



Notes: Squared error differences between Midas.IT model for horizon 0 and two benchmarks models.

I examine the informational content of daily news lead texts for nowcasting Swiss GDP growth:

- Using textual analysis, I create various text-based sentiment and recession indicators.
- I use factor analysis to extract a common component of these indicators.
- The indicator is highly correlated with other business cycle indicators and accurately tracks Swiss GDP growth.
- A major strength of the use of publicly available lead texts is that they are free of charge and can be accessed quickly.
- A real-time out-of-sample nowcasting exercise shows that if one month of data is available a model with text data outperforms a model with a well-known economic indicator.

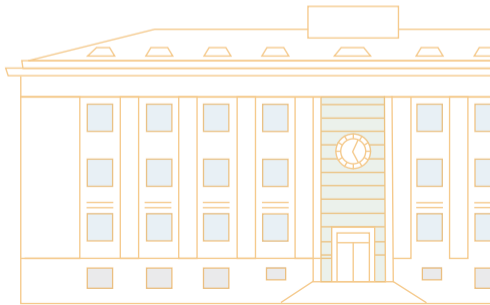
Thank you!



marc.burri@unine.ch



[marcburri.github.io](https://github.com/marcburri)



Appendix

Topic defining keywords

Topic	Keywords	English	Method
Recession	rezession, krise	recession, crisis	Count
Labor market	arbeit, job, beschäftigung	labor, job, employment	KWIC
Financial market	stock, asset, anlage, aktionär, aktie, dividend, börse, finanz, \bsmi\b, dax, \bspi\b, nasdaq, msci	stock, asset, investment, share, dividend, financial,\bsmi\b, dax, \bspi\b, nasdaq, msci	KWIC
Government	regierung, staat, minister, govern, \bbund\b, steuer, politik	government, state, minister, federal, tax, policy	KWIC
Investment	invest	invest	KWIC
Economy	wirtschaft, konjunktur, export, import	economy, business cycle, export, import	KWIC
Inflation	inflation, teuerung, preis	inflation, price	KWIC

Back

Document-level sentiment score

Lead text of FUW from March 6, 2020

Before cleaning

Das Coronavirus trifft die Schweizer Wirtschaft ab dem Frühjahr mit voller Kraft. Volkswirte stimmen auf ein schwaches zweites Quartal ein. Konsum, Tourismus und Exportindustrie leiden bereits. Der Bundesrat sollte die Kurzarbeit ausweiten.

After cleaning

coronavirus trifft schweizer **wirtschaft** frühjahr voller **kraft** volkswirte stimmen **schwaches** zweites quartal konsum tourismus **exportindustrie** **leiden** bereits bundesrat kurzarbeit ausweiten

In English

Before cleaning

The corona virus is hitting the Swiss economy with full force from the spring. Economists are predicting a weak second quarter. Consumption, tourism and the export industry are already suffering. The Federal Council should extend short-time work.

After cleaning

corona virus hitting swiss **economy** spring full **force** economists predicting **weak** second quarter consumption tourism export **industry** already **suffering** federal council extend short time work

Notes: Example of how document-level sentiment scores for two topics are calculated. For the general economy topic that is defined by the keyword (in blue) *wirtschaft*, the number of negative words (in red) is subtracted from the number of positive words (in green) within the ten preceding and following words from the keyword, and this result is divided by the total number of words. In this case, the sentiment score is $S_{t,d,economy} = (1 - 1)/14 = 0$.

References

- Babii, A., Ghysels, E., and Striaukas, J. (2021). Machine Learning Time Series Regressions With an Application to Nowcasting. *Journal of Business & Economic Statistics*, pages 1–23, ISSN: 0735-0015, 1537-2707, DOI: [10.1080/07350015.2021.1899933](https://doi.org/10.1080/07350015.2021.1899933).
- Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636, ISSN: 1531-4650, 0033-5533, DOI: [10.1093/qje/qjw024](https://doi.org/10.1093/qje/qjw024).
- Barbaglia, L., Consoli, S., and Manzan, S. (2022). Forecasting with Economic News. *Journal of Business & Economic Statistics*, pages 1–12, ISSN: 0735-0015, 1537-2707, DOI: [10.1080/07350015.2022.2060988](https://doi.org/10.1080/07350015.2022.2060988).
- Buckman, S. R., Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2020). News sentiment in the time of COVID-19. *FRBSF Economic Letter*, 2020(08):1–05.
- Indergand, R. and Leist, S. (2014). A real-time data set for Switzerland. *Swiss Journal of Economics and Statistics*, 150(IV):331–352, DOI: [10.1007/BF03399410](https://doi.org/10.1007/BF03399410).

References II

- Iselin, D. and Siliverstovs, B. (2013). The R-word index for Switzerland. *Applied Economics Letters*, 20(11):1032–1035, ISSN: 1350-4851, 1466-4291, DOI: [10.1080/13504851.2013.772290](https://doi.org/10.1080/13504851.2013.772290), Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/13504851.2013.772290>.
- Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., and Kapadia, S. (2022). Making text count: Economic forecasting using newspaper text*. *Journal of Applied Econometrics*, page jae.2907, ISSN: 0883-7252, 1099-1255, DOI: [10.1002/jae.2907](https://doi.org/10.1002/jae.2907).
- Larsen, V. H. (2021). Components of Uncertainty. page 20.
- Neusser, K. (2016). *Time Series Econometrics*. Springer Texts in Business and Economics. Springer International Publishing, Cham, ISBN: [978-3-319-32861-4](https://doi.org/10.1007/978-3-319-32861-4), DOI: [10.1007/978-3-319-32862-1](https://doi.org/10.1007/978-3-319-32862-1).
- Remus, R., Quasthoff, U., and Heyer, G. (2010). SentiWS - a publicly available German-language resource for sentiment analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

References III

- Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2020). Measuring news sentiment. *Journal of Econometrics*, page S0304407620303535, ISSN: 03044076, DOI: [10.1016/j.jeconom.2020.07.053](https://doi.org/10.1016/j.jeconom.2020.07.053).
- The Economist (2011). Up means down. The R-word index. The economist's gauge of gloom. Retrieved from <http://www.economist.com/node/21529079>.
- Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, 38(2):393–409, DOI: [10.1080/07350015.2018.1506344](https://doi.org/10.1080/07350015.2018.1506344).