Université de Neuchâtel
Faculté des sciences économiques

Three Essays in Macroeconomics
with Applications in Textual Analysis

PhD thesis submitted to the Faculty of Economics and Business

Institute of Economic Research

University of Neuchâtel

For the PhD degree in Economics

by

Marc Burri

Approved by the dissertation committee:

Prof. Daniel Kaufmann, University of Neuchâtel, thesis director

Prof. Jean-Marie Grether, University of Neuchâtel

Prof. Leif Anders Thorsrud, Norwegian Business School BI, Norway

Prof. Mark W. Watson, Princeton University, USA

Defended on July 5, 2024

University of Neuchâtel

Ph.D. in Economics

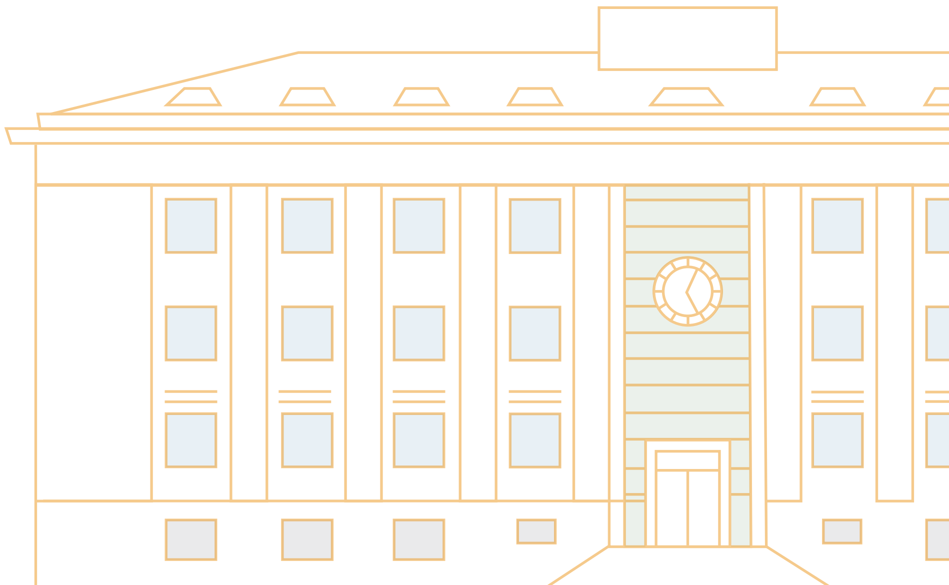
Three Essays in Macroeconomics

with Applications in Textual Analysis

Author
Marc Burri

Advisor
Prof. Daniel Kaufmann

Members of the Commitee
Prof. Jean-Marie Grether
Prof. Daniel Kaufmann
Prof. Leif Anders Thorsrud
Prof. Mark W. Watson



Bern,
July 5, 2024

Nöistart-Tröim ha doch när nid erwache...

— Migo

Preface

First and foremost, my deepest gratitude goes to my supervisor and co-author, Prof. Daniel Kaufmann, for providing me with the opportunity to pursue this Ph.D. His close guidance and relentless support were instrumental in enabling me to realize my potential as a researcher.

My sincere thanks are extended to the members of the thesis committee. To Prof. Leif Anders Thorsrud, for his warm welcome at the Norwegian Business School in Oslo and for his constructive feedback, which significantly improved this thesis. His encouragement to embrace the sunny days by occasionally stepping out of the office for cross-country skiing in the picturesque Oslo forests has been truly invigorating. To Prof. Mark W. Watson, for his invaluable feedback. His passion and drive have enriched our discussions, and I have benefited greatly from his guidance and teaching.

I also wish to convey my heartfelt appreciation to all the people I had the privilege of working with at the University of Neuchâtel, in Gerzensee, and at the Norwegian Business School. They made this academic journey incredibly enjoyable.

Despite, or precisely because of their ongoing attempts to grasp the specifics of my research, my endless gratitude goes out to my family and friends. They provided constant encouragement and support from my earliest memories.

Last but certainly not least, my gratitude to Sarah knows no bounds. It is only with her that attempting to solve math problems deep into the night becomes an enjoyable adventure. Her unwavering love has been my greatest support.

Marc Burri *Bern, May 2024*

Summary

Textual analysis in economics involves examining unstructured text data to extract meaningful insights about economic variables. This approach differs from traditional reading in its ability to handle extensive amounts of text and to generate reproducible results, often uncovering patterns that human readers might miss. With the rise of digital data and large language models, textual analysis has become increasingly relevant in economics, where it helps measure hard-to-quantify variables, track business cycles, and identify causal relationships.

This dissertation contributes to this growing literature with three essays that utilize textual analysis to address questions in macroeconomics.

The first chapter develops a daily business cycle indicator for Switzerland. Using financial market and news data, a timely "fever curve" is created, offering more accurate nowcasts of economic activity than existing indicators.

The second chapter focuses on disentangling the effects of different U.S. monetary policy shocks on the exchange rate based on a new methodology that identifies shocks based on changes in the variance-covariance matrix of financial variables.

The third chapter constructs a historical business cycle indicator for Switzerland from 1821 to the present. Using textual data from company records and newspapers, a chronology of Swiss economic fluctuations is provided, aligning with well-known events and offering new insights into historical business cycles.

Keywords: Textual analysis; News sentiment; Nowcasting; Business cycles; Monetary policy; Causal identification; Economic history

Résumé

L’analyse textuelle en économie consiste à examiner des données textuelles non structurées afin d’en extraire des informations sur les variables économiques. Cette approche se distingue de la lecture traditionnelle par sa capacité à traiter des volumes importants de texte et à générer des résultats reproductibles, révélant souvent des modèles que les lecteurs humains pourraient manquer. Avec l’essor des données numériques et des “large language models”, l’analyse textuelle est devenue pertinente en économie, où elle aide à mesurer des variables difficiles à quantifier, à suivre les cycles économiques et à identifier des relations causales.

Cette thèse contribue à cette littérature en pleine expansion avec trois essais qui utilisent l’analyse textuelle pour répondre à des questions de macroéconomie.

Le premier chapitre développe un indicateur quotidien des cycles économiques pour la Suisse. En utilisant des données des marchés financiers et des articles de journaux, une “courbe de fièvre” est créée, offrant des prévisions plus précises de l’activité économique que les indicateurs existants.

Le deuxième chapitre se concentre sur le démêlage des effets des différents chocs de politique monétaire américaine sur le taux de change, en s’appuyant sur une nouvelle méthodologie qui identifie les chocs à partir des changements dans la matrice de variance-covariance des variables financières.

Le troisième chapitre construit un indicateur historique des cycles économiques pour la Suisse de 1821 à aujourd’hui. En utilisant des données textuelles telles que des archives d’entreprises et des journaux, une chronologie des fluctuations économiques est proposée, en phase avec les événements bien connus.

Mots-clés : Analyse textuelle ; News sentiment ; Nowcasting ; Cycles économiques ; Politique monétaire ; Identification causale ; Histoire économique

Contents

List of Figures

List of Tables

Introduction

Textual analysis involves systematically examining and interpreting the meanings of various texts. It aims to uncover patterns, themes, and relationships within the texts, often through coding schemes and qualitative analytic techniques. Textual analysis includes various computational tools and statistical techniques to extract meaning from unstructured textual data. In principle, it is similar to reading in that both seek to extract useful information from text sources by identifying and exploring interesting patterns. However, textual analysis differs from reading in at least three critical respects. First, using a computer allows processing more texts than a human can ever read. Second, the extracted meaning can be consistently reproduced with structured approaches, whereas two people may arrive at different conclusions reading the same text. Third, textual analysis may be able to extract meanings from texts that may be overlooked by human readers, who miss specific patterns because they are inconsistent with prior beliefs and expectations (Bholat et al., 2015).

In the wake of the advance of the internet as well as the exponential growth of computing and storage capacity, there has been a surge in digitized textual data during the last two decades: Products are purchased and rated online; Governments interact with the people online; Discussions are held in online forums; People meet and interact on social media; Newspapers are read digitally. Texts generated by these activities are a natural way of human communication. Hence, textual documents reflect many human, including economic, activities. Therefore, analyzing texts has become important in many fields of research. The recent rise of large language models (LLM) has further increased the interest and the possibilities of textual analysis.

Although widely applied for a longer time in other fields, such as finance, political science, and marketing, textual analysis as a technique has only recently become more extensively used in economics.[1] Broadly speaking, three strands of the economic literature using textual analysis can be distinguished. First, it is used to extract hard-to-quantify and unobserved economic variables, such as inflation expectations or economic policy uncertainty. Second, textual data is used to track and forecast economic statistics that are published with a delay, such as GDP growth. Finally, textual data can be used in causal analysis. For example, narratives and economic news reflected in textual data may themselves be causal drivers of the business cycle (Shiller, 2017, 2019). In the following, I briefly review the literature in these three areas and position my thesis within it.

Textual analysis and hard-to-quantify variables

Unobserved variables are often of interest in economics. For example, inflation expectations are a key variable in monetary policy. However, they are not directly observable. Binder (2016) uses textual analysis to obtain a measure of inflation expectations during the Great Depression. Where studies of more recent periods use surveys or financial market expectations, similar data are not available for the 1930s. Therefore, Binder (2016) constructs a measure of inflation expectations by calculating a "positive" and "negative" news index based on articles in the New York Times. To validate the method, the author replicates the measure for the period 1978-97 and compares it with Michigan Survey of Consumers expectations over the same period. In a related study, Angelico et al. (2022) use machine learning techniques to extract inflation expectations from Twitter (now X) tweets. They show that the extracted expectations are highly correlated with survey-based measures of inflation expectations. Larsen et al. (2021) show that many topics the news media writes about have predictive power for inflation expectations based on surveys.

A second example deals with measuring economic and policy uncertainty. Baker et al. (2016) develop an index of economic policy uncertainty in the U.S. using newspaper archives.[2] Their primary index covers the period 1985-2015 and is based on searches for word combinations in the archives of 10 newspapers. The authors can match the spikes in the resulting uncertainty index to events such as the Gulf Wars, Presidential election results, and important events during the Global Financial Crisis.

Missing hard data for historical periods presents a challenge in quantitative analysis.

[1]See Ash and Hansen (2023) or Gentzkow et al. (2019) for an overview.

[2]Following Baker et al. (2016) several studies used a similar methodology to estimate economic policy uncertainty. See e.g. Ardia et al. (2021), Azqueta-Gavaldón (2017), Cieslak et al. (2021), and Larsen (2021).

To address this, researchers have turned to textual analysis to measure related concepts, such as economic sentiment, and explore its impact on the economy. Van Binsbergen et al. (2024) utilized machine learning techniques to construct a 170-year-long measure of economic sentiment from historical newspapers, demonstrating its predictive power on a range of economic variables. Similarly, Kabiri et al. (2023) delved into the role of sentiment during the 1920s boom and the 1930s depression, revealing significant effects of sentiment shocks on both economic and financial market indicators. Hanna et al. (2020) uncovered a statistically significant relationship between newspaper sentiment and stock market returns from 1899 to 2010. Furthermore, Hirshleifer et al. (2023) show that a war factor derived from New York Times articles predicts stock market returns.

Textual analysis and the business cycle

Techniques from textual analysis have also been used to obtain business cycle indicators, that is, indicators that track or forecast booms and recessions. For instance, Larsen and Thorsrud (2018), Shapiro et al. (2022), Thorsrud (2020), and Bybee et al. (2023) apply text mining techniques to news sources to obtain leading indicators and measures of the business cycle.

Larsen and Thorsrud (2018) and Thorsrud (2020) classified the collected news articles into 80 possible news topics or narratives. Then, they identified the tone of each article using the "lexical" approach. [3] To convert the sentiment indicators into a daily coincident index of the business cycle, they use a Dynamic Factor Model (DFM).

Shapiro et al. (2022) manually classified part of their articles into positive or negative sentiments. They then show that a novel sentiment-scoring model, which combines existing lexicons with a new lexicon that they construct specifically to capture the sentiment in economic news articles, predicts their manual classification more accurately than any of the existing lexicons. They then develop a monthly news sentiment index by estimating month-fixed effects from a regression over the articles' sentiment score calculated using their sentiment-scoring model.

Textual data are also increasingly used for forecasting economic variables. Ardia et al. (2019) show that news sentiments improve forecasts of U.S. industrial production growth. Additionally, Ellingsen et al. (2022) find that news data offer valuable insights for predicting consumption developments. Furthermore, Barbaglia et al. (2023) and Kalamara et al. (2022) provide evidence that leveraging textual data enhances forecasts

[3] This approach counts "positive" and "negative" words from a predefined lexicon.

for macroeconomic indicators including GDP, inflation, and unemployment.

Textual analysis and causal identification

Larsen and Thorsrud (2019) use newspaper articles to identify news and sentiment shocks for the Norwegian economy. They classify the articles into topics and identify the news shock as the innovations to a news index, derived as a weighted average of the topics with the highest predictive scores for key economic variables. The sentiment shock is identified as the innovations to asset prices, orthogonal to news shocks.

Ter Ellen et al. (2022) adapt an event study framework and propose a method to quantify narratives from news articles. They identify "narrative monetary policy surprises" as the change in economic media coverage explained by central bank communication accompanying interest rate meetings. Similarly, Aruoba and Drechsel (2024) apply methods from textual analysis to documents that economists at the Federal Reserve Board (FED) prepare for Federal Open Market Committee (FOMC) meetings, thereby capturing all the information available to the FOMC at the time of their interest rate decision. Following Romer and Romer (2004), they then estimate a novel monetary policy shock series by predicting changes in the target interest rate conditional on this information and obtain a measure of monetary policy shocks as the residual.

The release of textual data such as transcripts and minutes of central bank meetings can potentially affect the economy. Hansen and McMahon (2016) and Hansen et al. (2019) provide evidence that central bank communication indeed has a causal impact on the economy. Nonetheless, the effect on financial market variables is stronger than on real economic variables. This is confirmed by Swanson and Jayawickrema (2023), who find that speeches by the FOMC Chair and Vice-Chair are even more important than FOMC meeting announcements.

Finally, the use of archival text material is common for shock identification. For instance, narrative methods, such as Friedman and Schwartz (1993), Romer and Romer (1989) and Romer and Romer (2004), are well established. Researchers have also used their reading of documents to make quantitative judgments. For example, Naef (2019) uses archival information from the Bank of England to identify the success rate of central bank interventions in foreign exchange markets from 1952 to 1993.

Contribution of thesis

This thesis aims to contribute to the literature with three essays in macroeconomics using tools from textual analysis. With chapter 1 and chapter 3 of my dissertation, I primarily contribute to the literature on measuring and predicting business cycle fluctuations. In addition, chapter 3 shows that textual data can be used to measure a hard-to-quantify variable over historical episodes. In chapter 2, I contribute to the literature on the causal identification of monetary policy shocks.

Chapter 1 uses textual analysis to measure and nowcast the business cycle on a daily basis. It is based on work co-authored with Daniel Kaufmann. We develop a fever curve for the Swiss economy using publicly available daily financial market and news data. The indicator is highly correlated with macroeconomic data and survey indicators of Swiss economic activity. Therefore, it provides timely and reliable warning signals if the health of the economy takes a turn for the worse. In a nowcasting exercise, we show that our indicator provides more accurate forecasts than a prominent business cycle indicator once one month of daily information of the current quarter is available.

Chapter 2 aims to measure the causal impact of monetary policy shocks on the exchange rate for the United States. It is joint work with Daniel Kaufmann. We propose a novel methodology to disentangle an interest rate target, path, and term premium shock. The target shock affects interest rates of all maturities, the path shock affects interest rates in the medium-term and long-term, and the term premium shock only affects interest rates in the long-term. Our approach uses the difference of the variance-covariance matrix of financial market variables on policy event and control days for identification. Therefore, it does not require high-frequency data nor exact knowledge of the intraday timing of the event. Nevertheless, our shocks are correlated with existing high-frequency surprises. Our application shows that a target shock temporarily appreciates the USD. But the effect vanishes after about ten working days. We find similar effects for a path shock. The term premium shock appreciates the USD with a small delay of ten working days. Finally, we use our shocks to estimate the monthly macroeconomic effects in an Structural Vector Autoregression (SVAR) identified using external instruments. Although the exchange rate response is persistent, there is no evidence of a delayed overshooting puzzle. In this chapter, textual analysis has been used to identify relevant speeches and Testimony by the FOMC Chair and Vice-Chair to include them in the event study. This is a more data-driven approach compared to Swanson and Jayawickrema (2023), who read the newspaper the next day to check whether the speech was relevant.

Chapter 3 uses textual data, such as historical company records, newspapers, and business association reports, to develop a quarterly business cycle indicator for Switzerland starting in 1821. The methodology is strongly related to Chapter 1.

However, the data collection is more extensive, surpassing existing datasets in scope and historical coverage. The business cycle indicator strongly correlates with real economic activity, effectively capturing historical downturns and expansions. Further, I use the indicator to provide a business cycle chronology for Switzerland. The identified recessionary phases for the late 20th century are in line with the dating of a private business cycle dating committee. For the 19th and early 20th centuries, my business cycle dating is consistent with narratives of wars and economic crises.

References

Angelico, C., Marcucci, J., Miccoli, M., & Quarta, F. (2022). Can we measure inflation expectations using twitter? *Journal of Econometrics*, *228*(2), 259–277. https://doi.org/10.1016/j.jeconom.2021.12.008

Ardia, D., Bluteau, K., & 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. https://doi.org/10.1016/j.ijforecast.2018.10.010

Ardia, D., Bluteau, K., & Kassem, A. (2021). A century of economic policy uncertainty through the french–canadian lens. *Economics Letters*, *205*, 109938. https://doi.org/10.1016/j.econlet.2021.109938

Aruoba, S. B., & Drechsel, T. (2024). Identifying monetary policy shocks: A natural language approach. *NBER Working Paper No. 32417*. https://doi.org/10.3386/w32417

Ash, E., & Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, *15*(1), 659–688. https://doi.org/10.1146/annurev-economics-082222-074352

Azqueta-Gavaldón, A. (2017). Developing news-based economic policy uncertainty index with unsupervised machine learning. *Economics Letters*, *158*, 47–50. https://doi.org/10.1016/j.econlet.2017.06.032

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Barbaglia, L., Consoli, S., & Manzan, S. (2023). Forecasting with economic news. *Journal of Business & Economic Statistics*, *41*(3), 708–719. https://doi.org/10.1080/07350015.2022.2060988

Bholat, D. M., Hansen, S., Santos, P. M., & Schonhardt-Bailey, C. (2015). Text mining for central banks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2624811

Binder, C. C. (2016). Estimation of historical inflation expectations. *Explorations in Economic History*, *61*, 1–31. https://doi.org/10.1016/j.eeh.2016.01.002

Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2023). Business news and business cycles. *Journal of Finance*, *Forthcoming*. http://dx.doi.org/10.2139/ssrn.3446225

Cieslak, A., Hansen, S., McMahon, M., & Xiao, S. (2021). Policymakers' uncertainty. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3936999

Ellingsen, J., Larsen, V. H., & Thorsrud, L. A. (2022). News media versus FRED-MD for macroeconomic forecasting. *Journal of Applied Econometrics*, *37*(1), 63–81. https://doi.org/10.1002/jae.2859

Friedman, M., & Schwartz, A. J. (1993). *A monetary history of the united states: 1867 - 1960* (9. paperback print). Princton Univ. Press.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Hanna, A. J., Turner, J. D., & Walker, C. B. (2020). News media and investor sentiment during bull and bear markets. *The European Journal of Finance*, *26*(14), 1377–1395. https://doi.org/10.1080/1351847X.2020.1743734

Hansen, S., & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, *99*, S114–S133. https://doi.org/10.1016/j.jinteco.2015.12.008

Hansen, S., McMahon, M., & Tong, M. (2019). The long-run information effect of central bank communication. *Journal of Monetary Economics*, *108*, 185–202. https://doi.org/10.1016/j.jmoneco.2019.09.002

Hirshleifer, D., Mai, D., & Pukthuanthong, K. (2023). War discourse and the cross section of expected stock returns. *NBER Working Paper No. 31348*. https://doi.org/10.3386/w31348

Kabiri, A., James, H., Landon-Lane, J., Tuckett, D., & Nyman, R. (2023). The role of sentiment in the US economy: 1920 to 1934. *The Economic History Review*, 3–30. https://doi.org/10.1111/ehr.13160

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2022). Making text count: Economic forecasting using newspaper text. *Journal of Applied Econometrics*, *37*(5), 896–919. https://doi.org/10.1002/jae.2907

Larsen, V. H., & Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, *210*(1), 203–218. https://doi.org/10.1016/j.jeconom.2018.11.013

Larsen, V. H., & Thorsrud, L. A. (2018). Business cycle narratives. *CESifo Working Paper No. 7468*.

Larsen, V. H., Thorsrud, L. A., & Zhulanova, J. (2021). News-driven inflation expectations and information rigidities. *Journal of Monetary Economics*, *117*, 507–520. https://doi.org/10.1016/j.jmoneco.2020.03.004

Larsen, V. H. (2021). Components of uncertainty. *International Economic Review*, *62*(2), 769–788. https://doi.org/10.1111/iere.12499

Naef, A. (2019). Blowing against the wind? a narrative approach to central bank foreign exchange intervention. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3483895

Romer, C. D., & Romer, D. H. (1989, July). Does monetary policy matter? a new test in the spirit of friedman and schwartz. *NBER macroeconomics annual 1989, volume 4* (pp. 121–184). National Bureau of Economic Research, Inc. https://ideas.repec.org/h/nbr/nberch/10964.html

Romer, C. D., & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, *94*(4), 1055–1084. https://doi.org/10.1257/0002828042002651

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, *228*(2), 221–243. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.07.053

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004. https://doi.org/10.1257/aer.107.4.967

Shiller, R. J. (2019). *Narrative economics: How stories go viral & drive major economic events*. Princeton University Press.

Swanson, E. T., & Jayawickrema, V. (2023). Speeches by the fed chair are more important than FOMC announcements: An improved high-frequency measure of u.s. monetary policy shocks. *mimeo*.

Ter Ellen, S., Larsen, V. H., & Thorsrud, L. A. (2022). Narrative monetary policy surprises and the media. *Journal of Money, Credit and Banking*, *54*(5), 1525–1549. https://doi.org/10.1111/jmcb.12868

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. https://doi.org/10.1080/07350015.2018.1506344

Van Binsbergen, J., Bryzgalova, S., Mukhopadhyay, M., & Sharma, V. (2024). (almost) 200 years of news-based economic sentiment. *NBER Working Paper No. 32026*. https://doi.org/10.3386/w32026

1 A daily fever curve for the Swiss economy[†]

1.1 Introduction

Because macroeconomic data is published with a substantial delay, assessing the health of the economy during the rapidly evolving coronavirus disease of 2019 (Covid–19) crisis was challenging. Usually, policymakers and researchers make predictions and take decisions based on early information from surveys, financial markets, and rapidly available statistics (see, e.g., Abberger et al., 2014; Galli, 2018; Kaufmann & Scheufele, 2017; OECD, 2010; Stuart, 2020; Wegmüller & Glocker, 2019, for Swiss applications). These indicators and forecasts are published with a delay of at least one to two months.[1] Although this approach works well in stable periods, the Covid–19 crisis has shown that these indicators are not sufficient to make important decisions in a rapidly changing situation. Consequently, there is a pressing need for high-frequency information to accurately evaluate the current economic conditions and assess the impact of dynamic factors such as health restrictions and economic stimulus measures.

[†]This chapter is partly co-authored with Daniel Kaufmann. The co-authored part has been published in the Swiss Journal of Economics and Statistics (Burri & Kaufmann, 2020). We thank David Ardia, Ronald Indergand, Alexander Rathke, Maximilian Schröder, Jonas Striaukas, Jan-Egbert Sturm, Leif Anders Thorsrud, Mark W. Watson, Philipp Wegmüller, and Vaidotas Zemlys-Balevičius for helpful comments and discussions and participants at the IRENE PhD meeting, the IAAE Annual Conference in Oslo, the EEA-ESEM Conference in Barcelona, and the Money-Macro-Finance Conference in Warsaw for valuable comments.

[1]See table 1.A.1 in the Appendix for publication lags of some important macroeconomic data and leading indicators.

We introduce a novel approach to assess the well-being of the Swiss economy, proposing a daily fever curve (f–curve) based on publicly available financial market and news data. To construct the f–curve, we utilize risk premia on corporate bonds, term spreads, and stock market volatility indices from 2000 onwards. Additionally, we gather short economic news lead texts from online newspaper archives. Using these news articles, we generate sentiment indicators for various hand-selected economic concepts (topics), such as the labor market or inflation.[2] Subsequently, we use factor analysis to estimate a composite indicator that can be interpreted as a fever curve. Analogous to monitoring a patient's tempereature, an upward trend in the fever curve serves as a reliable and timely warning signal indicating a deterioration in the economy's health.

Panel (a) of Figure 1.1 shows the f–curve (on an inverted scale) jointly with real gross domestic product (GDP) growth: the indicator closely tracks economic crises. It presages the downturn during the Global Financial Crisis, responds to the removal of the minimum exchange rate in 2015, and to the euro area debt crisis. The f–curve also responds strongly to the Covid–19 crisis (see panel (b)). The indicator starts to rise in late February. By then, it became evident that the Covid–19 crisis would hit most European countries; in Switzerland, the first public events were canceled. It reaches a peak shortly after the lockdown. Afterwards, the fever curve gradually declines with news about economic stimulus packages and the gradual loosening of the lockdown. The peak during the Covid–19 crisis is comparable to the Global Financial Crisis. But the speed of the downturn is considerably higher. In addition, the crisis is less persistent. By the end of July 2020, the f–curve improved to 1/4 of its peak value during the lockdown. The indicator also reflects the health situation very well. It improves continuously as measures are relaxed one after the other and starts to rise just before tougher restrictions are introduced.

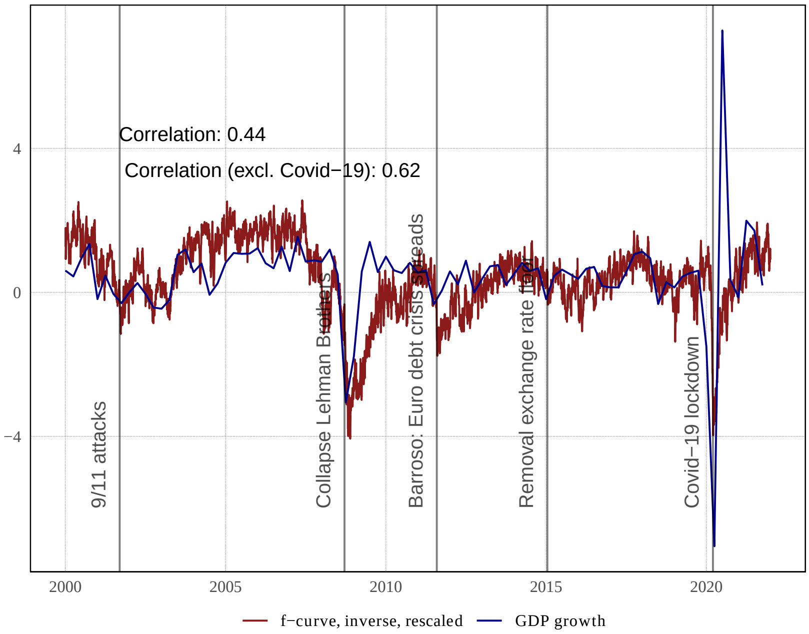
To evaluate the informational content of the f–curve more formally, we conduct an in-sample assessment, as well as a pseudo-real-time out-of-sample nowcasting exercise using **mi**xed-**da**ta **s**ampling (MIDAS) and bridge models (Baffigi et al., 2004; Ghysels et al., 2004).[3] In-sample, the indicator has a coincident or leading relationship with many business cycle indicators. In addition, it Granger causes many of them. Therefore, it allows us to track business cycle fluctuations in a timely and cost-effective manner. The out-of-sample analysis shows that nowcasts of GDP growth using the f–curve are

[2] In this context, "hand-selected" implies that we carefully choose keywords to define each topic by reviewing various lead texts.

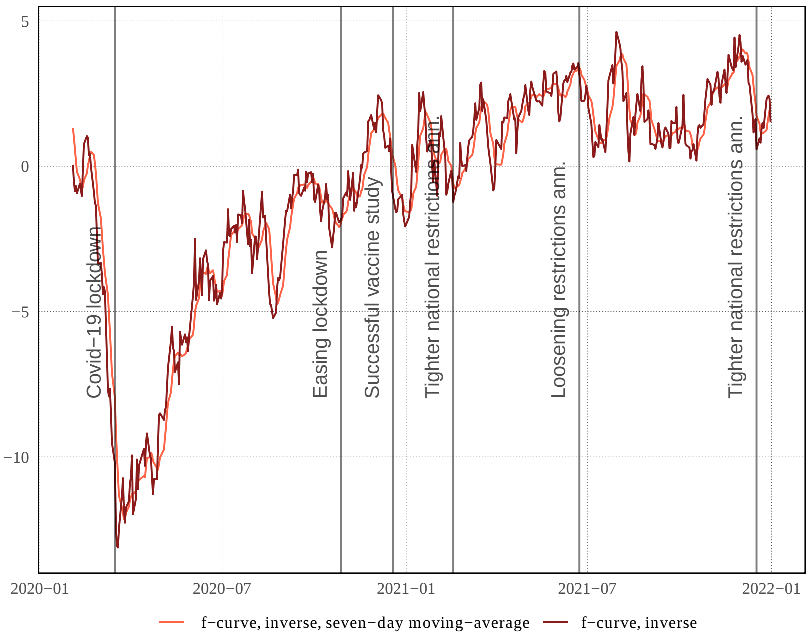
[3] Nowcasting refers to the problem of predicting the present, the very near future, and the very recent past. See Bańbura et al. (2012) and Bańbura et al. (2013) for extensive surveys.

Figure 1.1 — A fever curve for the Swiss economy

(a) Correlation with real GDP growth



(b) Evolution during the Covid–19 crisis



Notes: Panel (a) compares the *f*–curve (inverted and rescaled) to quarterly GDP growth. To calculate correlations, the *f*–curve is aggregated to quarterly. Periods of Covid–19 are defined as quarters 2 and 3 of 2020. Panel (b) panel gives daily values of the fever curve along with important policy decisions.

more accurate than those using existing business cycle indicators. Therefore, the f–curve provides accurate and timely information about the state of the economy. We then investigate the role of the state of information and the predictive performance over the business cycle. The f–curve provides more accurate forecasts once one month of daily information of the current quarter is available. Considering that most monthly survey data for the current month are released at the start of the following month, the f–curve contains valuable information even after accounting for realistic publication lags. Furthermore, the findings emphasize that nowcasting performance improvements mostly occur during business cycle turning points, underscoring the f–curve's effectiveness in detecting recessions in real-time..

Various initiatives in Switzerland and abroad exist to satisfy the demand for reliable high-frequency information that emerged during the Covid–19 crisis. Becerra et al. (2020) and Eichenauer et al. (2021) develop sentiment indicators using internet search engine data. Brown and Fengler (2020) provide information on Swiss consumption behavior based on debit and credit card payment data. Eckert et al. (2020) develop a daily mobility index using data on traffic, payments, and cash withdrawals. Using data on real economic activity, weekly indices to track economic activity have been developed for various countries (see, e.g. Lewis et al., 2022; Wegmüller et al., 2023). Moreover, Shapiro et al. (2022) create a daily news sentiment indicator that leads U.S. traditional consumer sentiment based on surveys.[4] News texts are also increasingly used for forecasting economic variables.[5] Ardia et al. (2019) show that news sentiments improve forecasts of U.S. industrial production growth. Ellingsen et al. (2022) find that news data are particularly informative for forecasting consumption developments. Moreover, Barbaglia et al. (2023) and Kalamara et al. (2022) demonstrate that it improves forecasts of macroeconomic variables such as GDP, inflation, and unemployment.

This chapter contributes to this literature in several ways. First, it investigates the information content of the daily financial market and news data for measuring business cycle fluctuations. So far, most studies in this field evaluated the information content of monthly indicators. Although daily business cycle indicators exist, there is no analysis of how many daily observations are needed to forecast GDP growth accurately. Second, unlike previous studies that relied on expensive or confidential data sources, we use

[4]Research on the use of textual data for measuring and forecasting economic activity has already been conducted before the Covid–19 pandemic, but usually at monthly frequency (See Bybee et al., 2023; Larsen & Thorsrud, 2018; Shiller, 2019; Thorsrud, 2020, for a few examples).

[5]Short news texts are also used to predict asset prices. Li et al. (2022) and Bai et al. (2022) show that news headlines can effectively be used to predict future prices.

publicly available financial market and news data, making it more accessible and cost-effective. Third, an innovation of this study lies in integrating financial market and text data. While studies evaluating the predictive value for both separately exist, their combined value has not been examined before. Moreover, the methodology to create news-based indicators for hand-selected economic concepts and the aggregation process is simple and robust to data leakage.[6]

The chapter proceeds as follows. In the next section, we describe the data and explain how they relate to the business cycle. Section 1.3 presents the methodology for creating text-based indicators and aggregating the data into a composite business cycle indicator. Moreover, it explains the mixed frequency methods for evaluating the indicators' out-of-sample performance. Section 1.4 provides a descriptive analysis and evaluates the f–curve in- and out-of-sample. The last section concludes.

1.2 Data

When selecting variables for the indicator, special attention is given to various properties. On the one hand, the data selection process is based on economic theory and intuition. On the other hand, daily data going back to at least the year 2000 is used, and the time series must be freely and quickly available. The selected variables can be updated with a delay of one day, provided that the data providers' websites are accessible. The following description outlines the two data types underlying the f–curve.

1.2.1 Financial market data

We use publicly available bond yields underlying the SIX Swiss Bond Indices® (SBI) (SIX, 2020a). These data are accessible and updated daily, albeit with a one-day delay. To account for the fact that many bond yields only began around 2007, the series is extended by closely matching government and corporate bond yields from the Swiss National Bank. The detailed information can be found in Figure 1.A.2 in the Appendix.[7]

Various spreads are then calculated, which are expected to correlate with economic activity. These include the government bond term spread (8Y - 2Y), the interest rate differential relative to the euro area (1Y), and the risk premia of short- and long-term corporate debt. In addition to interest rate spreads for Switzerland, the risk premia

[6] Data leakage is a situation where future information is accidentally incorporated into the model. Topic models (see, e.g. Thorsrud, 2020) are particularly prone to data leakage because they use the entire text to classify articles into topics.

[7] Data from the Swiss National Bank are published with a longer delay. Therefore, these bond yields cannot be used to track the economy on a daily basis.

of foreign companies that issue short- and long-term debt in Swiss francs are also computed. Term spreads for the U.S. and the euro area are included as well. For the latter, short-term interest rates in euro (European Central Bank, 2020) and long-term yields of German government debt (Deutsche Bundesbank, 2020) are utilized.

Furthermore, two implied volatility measures of the Swiss and U.S. stock markets are incorporated. The Swiss data is sourced from SIX (2020b) and is published with a one-day delay, while the U.S. data is obtained from the Chicago Board Options Exchange (2020).

These financial market data should be related to the Swiss business cycle. Stuart (2020) shows that the term spread exhibits a lead on the Swiss business cycle.[8] Kaufmann (2020) argues that a narrowing of the interest rate differential appreciates the Swiss franc and thereby dampens economic activity. Finally, risk premia are correlated with the default risk of companies, which should increase during economic crises. Finally, recent research documents increased uncertainty during economic downturns (Baker et al., 2016; Scotti, 2016). There are various ways to measure uncertainty (see, e.g. Dibiasi & Iselin, 2016). A measure of stock market volatility is preferred because the aim is to exploit quickly and freely available financial market data.

1.2.2 News lead texts

We complement the financial market data with publicly and quickly (usually with a delay of one day) available news data to create sentiment and recession indicators based on three Swiss newspapers. The newspaper data stem from the online archives of the *Tages-Anzeiger* (TA),[9] the *Neue Zürcher Zeitung* (NZZ)[10], and the *Finanz und Wirtschaft* (FUW).[11] These three newspapers are among the most relevant German-language newspapers reporting on economic affairs in Switzerland and abroad. Their archives cover the period from 2000 until today.[12] For this chapter, we use data from January 1, 2000 to December 31, 2021.

One of the main benefits of using news data is its immediate availability and longer time coverage compared to other high-frequency data sources. Additionally, extracting information from the data is straightforward, and since the data is publicly accessible, it

[8]Therefore, all term spreads are moved forward by half a year.
[9]See tagesanzeiger.ch/zeitungsarchiv-930530868737.
[10]See zeitungsarchiv.nzz.ch/archive.
[11]See fuw.ch/archiv.
[12]Sometimes the *Tages-Anzeiger* updates its archive with a relevant delay or not at all. Therefore, we additionally use lead texts from the *Tages-Anzeiger* website: tagesanzeiger.ch/wirtschaft.

can be used by anyone. While the public availability of news data is a positive aspect, it should be noted that only the titles, lead texts, or specific passages of articles are typically available instead of full articles. However, this should not be considered a significant drawback. These lead texts often succinctly convey the article's main message and tend to have less extraneous information, making signal extraction more precise.

We further reduce the signal-to-noise ratio by utilizing solely texts about the economy. Articles about subjects unrelated to the economy, like sports, may also express a sentiment, but it does not necessarily have any meaning for the economy. To filter out the most relevant articles, we focus on those that include specific German keywords related to the economy such as *Wirtschaft*, *Konjunktur* and *Rezession* (which translate to economy, business cycle, and recession respectively).[13] As a small open economy, Switzerland is greatly affected by economic developments in other countries. To account for this, we create indicators that measure sentiments and recession prevalence for both Switzerland and foreign countries by using location-specific keywords.[14] For more information on the search queries used to filter out relevant articles, see Table 1.A.3 in the Appendix.

Table 1.1 presents an overview of the web-scraped text data. Roughly 900'000 lead texts, text passages, or titles were collected. It is important to note that these 900'000 texts stem not from unique articles but are the results of multiple search queries. The average number of words is higher than 20 only for the *NZZ*, as it is the only newspaper that provides short passages from articles instead of just titles and lead texts. An interesting observation is that the more liberal *NZZ* and *FUW* have a more positive sentiment towards the economy than the *TA*, which is known to be rather left-leaning liberal. The lower sentiment found in the *Tages Anzeiger* webpage (TAW) is likely due to its different time coverage.

[13] Why not using keywords such as *Wirtschaftsaufschwung* (economic recovery) as well? The research conducted by Becerra et al. (2020) using Google Trends data suggests that terms associated with positive sentiment do not align with changes in economic activity. This finding highlights that people's interest in the economy is not symmetrical. It is also reflected in the behavior of journalists, who tend to focus more on recessions than on periods of growth. This phenomenon, known as "negativity bias" is not exclusive to journalists and is well-documented in the literature, where it has been shown that people tend to pay more attention to and remember negative information over positive information (see, e.g. Baumeister et al., 2001).

[14] We use specific keywords to identify articles related to the Euro area, Germany, and the USA, as these countries are major trading partners of Switzerland. For example, *Wirtschaft Schweiz* or *Rezession Deutschland* (economy switzerland, recession germany).

Table 1.1 — Descriptive statistics of the news data

Journal	#Texts	Avg. #Words	Avg. Sentiment	Coverage
Finanz und Wirtschaft	100’084	15.8	0.043	2000 - 2021
Neue Zürcher Zeitung	720’530	62.7	0.038	2000 - 2021
Tages Anzeiger	29’359	19.2	0.015	2000 - 2021
Tages Anzeiger Webpage	54’288	16.2	0.003	2008 - 2021

Notes: The total number of texts is not a unique count of articles. It is the total count of all articles satisfying the search queries represented in Table 1.A.3 in the appendix. The average number of words is calculated from the cleaned texts as outlined in Section 1.3. The average sentiment is calculated as the total number of positive minus the total number of negative words as defined by Remus et al. (2010), divided by the total number of words.

1.3 Methodology

This section describes the method of extracting information from the textual data, creating various sub-indices covering different areas of the economy, and aggregating all indicators into a business cycle indicator. Moreover, the models used for the out-of-sample nowcasting exercise are explained.

1.3.1 Creating text-based indicators

To convert the high-dimensional and unstructured newspaper texts into time series, they must be preprocessed (cleaned). We, therefore, filter out irrelevant information, as is standard in the natural language processing (NLP) literature. We remove Hyper Text Markup Language (HTML) tags, punctuation, numbers, and stopwords, which are words that are not informative, typically conjunctions such as “or” and “if”. The stop words are provided by Feinerer and Hornik (2019). Finally, we transform all letters to lowercase. Many NLP applications then stem the words, which is a process of removing and replacing word suffixes to arrive at a common root form of the word. However, this is unnecessary because we use a sentiment lexicon that is not stemmed.

The collected data is used to create text-based indicators that capture the different contexts (topics) of the economy in Switzerland and abroad. We use different sets of keywords denoted by $\mathcal{K}$ to define these contexts. A detailed list of the topic-defining keywords is shown in Table 1.2. The indicators are then created using two different methods.[15]

The first method creates recession indicators by counting the occurrence of keywords related to the recession topic and summing them up to a daily time series. The recession index, also known as the R-word index, was invented by The Economist (2011) in the early 90's. Iselin and Siliverstovs (2013) create an R-word index for Switzerland and find that it has predictive power to forecast Swiss GDP growth.[16] These indicators measure economic uncertainty and recession fears and are negatively correlated with the business cycle.

Table 1.2 — Keywords for economic topics

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, wechselkurs	stock, asset, investment, share, dividend, financial,\bsmi\b, dax, \bspi\b, nasdaq, msci, exchange rate	KWIC
Government	regierung, staat, minister, govern, \bbund\b, steuer, politik	government, state, minister, federal, tax, policy	KWIC
Investment	invest	invest	KWIC
Economy	wirtschaft, konjunktur, industrie, handel, import, export	economy, business cycle, industry, trade, import, export	KWIC
Inflation	inflation, teuerung, preis	inflation, price	KWIC

Notes: The column 'English' lists contextual translation of the German words. The queries use wildcard operators (i.e **krise** which also matches *krisengeschwächt*). The symbol \b reverses the wildcard operator (i.e \bspi\b doesn't match *spillovers*).

The second approach, the keyword-in-context (KWIC) method, utilizes word co-occurrences to calculate topic-specific sentiment indicators (Luhn, 1960). For this, we create new documents by screening the texts for keywords in $\mathcal{K}$. Whenever a keyword appears in a text, the keyword, along with the ten preceding and ten following words, is extracted into a new document.[17] A sentiment score is then calculated for each of these documents. Thus, the sentiment score is local because it considers only the text related

[15]The procedure is documented in detail in algorithm 1 in the Appendix.

[16]The recession indicators are highly correlated with uncertainty indicators invented by Baker et al. (2016). An evaluation has shown that creating uncertainty indicators provides no added value.

[17]This means we use a context window of ten words. We have also tested context windows of five or fifteen words. However, this does not significantly change the results.

to a topic of interest. Denote by $\mathcal{P}$ and $\mathcal{N}$ the list of phrases identified as positive and negative sentiment derived by Remus et al. (2010). The sentiment score subtracts the counts of words in $\mathcal{N}$ from the counts of terms in $\mathcal{P}$ in document d, and scales it by the number of total terms in document d. This is also referred to as the lexical methodology (see, e.g., Ardia et al., 2019; Shapiro et al., 2022; Thorsrud, 2020). More formally, let $w_{t,d,i,j} = (w_{t,d,i,j,1}, w_{t,d,i,j,2}, ...w_{t,d,i,j,N_{t,d,i,j}})$ be the list of terms in document d at date t for topic j. i is either "domestic" or "foreign". For simplicity, we drop the subscript i in what follows. The document-level news sentiment is hence given by

$$S_{t,d,j} = \frac{\sum_n \mathbb{1}(w_{t,d,j,n} \in \mathcal{P}) - \sum_n \mathbb{1}(w_{t,d,j,n} \in \mathcal{N})}{N_{t,d,j}} \tag{1.1}$$

where $N_{t,d,j}$ is the number of terms in the document. Figure 1.2 provides a more intuitive example of how the document-level sentiment score is being calculated. Finally, daily news sentiment indicators, $S_{t,j}$, for the domestic and foreign economy and for a given topic j are calculated as a simple average of the sentiment scores.

Several studies have used a probabilistic topic model to classify articles into topics (see, e.g., Ellingsen et al., 2022; Hansen et al., 2018; Thorsrud, 2020). With a topic model, articles can be classified into distinct topics based on their word content, enabling the assignment of similar-worded articles to the same topics. However, assigning the article-level sentiment score to a specific topic becomes challenging due to the possibility of articles discussing multiple topics. The KWIC approach offers greater specificity by focusing solely on the topic-specifying keyword and calculating the sentiment score based on a few surrounding words. Moreover, the brevity of the texts and the occasional absence of complete passages pose difficulties in accurately estimating a topic model (Yan et al., 2013). Finally, employing the proposed approach, using keywords not tied to any economic events, helps mitigate potential data leakage issues, which could adversely impact the nowcasting exercise (see Kalamara et al., 2022).[18]

[18] Nevertheless, we have tested two topic models specifically designed for short texts. The first is an algorithm that models co-occurrences of bi-terms (bi-terms are pairs of words appearing together in a text), and the second is a structural topic model, which is a general framework for topic modeling with document-level covariate information (see Roberts et al., 2014; Yan et al., 2013). The results were significantly worse than those of the KWIC method used here.

Figure 1.2 — 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 based on a article lead text from FUW. 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 $S_{t,d,\text{economy}} = (1-1)/14 = 0$. Note that there are only 14 words in the denominator because the keyword is close to the beginning of the text. The same method is applied to calculate the sentiment score for other topics, such as the industry topic, which in this example is given by $S_{t,d,\text{industry}} = (1-2)/19 = -0.05$.

1.3.2 Estimation of indicator

This chapter aims to provide a daily indicator of the business cycle and evaluate the value added of the daily frequency for nowcasting quarterly Swiss GDP growth. To accomplish this, we use models that can handle time series of different frequencies within the same regression without needing to transform or aggregate them. MIDAS and bridge models are commonly used in the literature for this purpose. However, including many explanatory variables in a MIDAS model can lead to parameter proliferation.[19] Additionally, financial market data and news indicators are quite volatile and correlated.

[19] Parameter proliferation in the context of MIDAS models refers to the issue that arises when too many parameters are included in the model.

To effectively summarize the information content of the data and eliminate idiosyncratic noise while avoiding parameter proliferation, we estimate a factor model in static form:[20]

$$x_{i,t} = \lambda_i f_t + e_{i,t} \tag{1.2}$$

where $x_{i,t}$ denotes one of the $n = 24$ indicators $(i = 1, ..., n)$, λ_i is the factor loading, f_t are the common factors at time t, and e_t is the idiosyncratic component. The model comprises T daily observations $(t = 1, ..., T)$. The advantage of using a factor model is that it allows for summarizing the information in a large data matrix with a small number of common factors. Factors and loadings can be estimated through principal components assuming that the idiosyncratic components are only weakly serially and cross-sectionally correlated (Bai & Ng, 2013; Stock & Watson, 2002).[21]

Given that the construction of the indicators is based on economic reasoning, the first principal component of the static factor model can be interpreted as a coincident business cycle indicator. Moreover, an information criterion to determine the number of factors in approximate factor models proposed by Bai and Ng (2002) confirms that one

[20]The news indicators are much more volatile than the financial market data (see Figures 1.A.1, 1.A.3 and 1.A.4 in the Appendix). We, therefore, compute a one-sided ten-day moving average before including them in the factor model. Comparable studies smooth their news sentiments with a sixty-day or higher moving average (see, e.g. Shapiro et al., 2022; Thorsrud, 2020). The moving average time window choice is a trade-off between less volatility and more timeliness. We tried different time windows and found a good compromise with the ten-day window.

[21]We exclude weekends and holidays. Then, we interpolate additional missing values using an EM algorithm (Stock & Watson, 2002) after standardizing the data to have zero mean and unit variance. We choose a relatively large number of factors for interpolating the data ($r = 4$). Finally, we use the first principal component of the interpolated data set. As a robustness test, we additionally estimate the factor model with missing values using least squares instead of the EM algorithm. The results are almost identical. Consider the model

$$x_{i,t} = \lambda_i f_t + e_{i,t}$$

and let $\delta_{i,t} = 1$ if $x_{i,t}$ is observed and zero otherwise. Then the principal components estimates of λ_i and f_t solve

$$\min_{\lambda_i, f_t} \sum_{t=1}^{T} \sum_{i=1}^{n} (x_{i,t} - \lambda_i f_t)^2 \delta_{i,t}$$

subject to the normalization

$$\sum_{i=1}^{n} \lambda_i' \lambda_i = I_r.$$

The problem can then be solved by using least squares iteratively.

factor is representing the data well enough.[22] We use the information criterion BIC_3, which is recommended for $n > 18$.

1.3.3 Out-of-sample evaluation models

How reliable is the *f*–curve and what is the informational content of the daily frequency? We perform a daily pseudo-real-time forecast evaluation using mixed frequency methods to answer these questions.

The variable of interest is quarterly GDP growth, which is denoted as y_{t_q}, where t_q is the quarterly time index $t_q = 1, 2, ..., T_y$, with T_y being the last quarter for which GDP figures are available. We use the real-time data set for quarterly GDP vintages by Indergand and Leist (2014), accounting for the ragged-edge structure due to the different publication dates of official quarterly GDP figures. The aim is to now- and forecast quarterly GDP growth, y_{Ty+H+1} with a horizon of $H = 0, 1$ quarters. We use this notation to emphasize that a horizon of $H = 0$ corresponds to a nowcast, whereas $H = 1$ is a forecast.

Similarly to Kuzin et al. (2011) and Schumacher (2016), we assume that the information set for now- and forecasting includes one stationary daily indicator x_{t_d} in addition to the available GDP observations. For simplicity, we assume every quarter to have $D = 60$ days, reflecting approximately five working days per week and four weeks per month. Hence, the time index for the daily observations is defined as a fraction of the low-frequency quarter according to $t_d = 1 - 59/60, 1 - 58/60, ..., 1, 2 - 59/60, ..., T_x - 1/60, T_x$, where T_x is the last day for which the daily indicator is available. Nowcasts are predictions for horizons of $h = 0, ..., 59$ days and forecasts for horizons of $h = 60, ...119$ days. The now- or forecast for GDP is conditional on information available in T_x, including all observations until T_x and the GDP observations up to T_y. The latter is because of $T_x \geq T_y$. The sample spans from January 1, 2000, to December 31, 2021.

To determine the informational content of the *f*–curve we forecast GDP growth using three models that exploit the information in the high-frequency indicator and link it to the low-frequency GDP. First, we estimate a MIDAS model introduced by Ghysels et al. (2004) and Ghysels et al. (2007). Second, we employ bridge equations following Baffigi et al. (2004). Third, we consider an iterative MIDAS model, a mixture of both, as discussed by Schumacher (2016).

[22]Nevertheless, an interesting extension would be to examine whether more than one factor comprises relevant information for Swiss economic activity. We leave this extension for future research.

MIDAS model

The MIDAS approach is a direct multi-step forecasting tool. We use the following model for a forecast horizon of H quarters (using the terminology of Schumacher (2016))

$$y_{t_q+H+1} = \alpha + \sum_{p=0}^{P-1} \beta_p \sum_{k=0}^{K-1} b(k,\theta) L^{(pD+k)/D} x_{t_d+T_x-T_y} + \varepsilon_{t+H+1} \tag{1.3}$$

where α is a constant, P denotes the number of low-frequency lags, and K is the number of high-frequency lags per low-frequency lag (both including zero). This modeling strategy is very flexible, allowing for different lag structures. We set $P = 2$ and $K = 60$, meaning the dependent variable depends on all 60 high-frequency values of the current and the last quarter. The daily lag operator is defined as $L^{1/60} x_{t_d} = x_{t_d - 1/60}$.

We determine the effect of the daily indicator $x_{t_d+T_x-T_y}$ on y_{t_q+H+1} by estimating a regression coefficient β_p for every low-frequency lag included.[23] Because x_{t_d} is sampled at a much higher frequency than y_{t_q}, we potentially have to include many high-frequency lags to achieve adequate modeling. This can easily lead to overparameterization in the unrestricted linear case. We use a non-linear weighting scheme given by the polynomial $b(k,\theta)$ to avoid parameter proliferation. The same polynomial specification is applied to all low-frequency lags included in the model.[24]

We use an exponential Almon lag of order two for the polynomial specification. This polynomial is extensively discussed in Ghysels et al. (2007) and has the following form:[25]

$$b(k,\theta) = \frac{exp(\theta_1 k + \theta_2 k^2)}{\sum_{j=0}^{K} exp(\theta_1 j + \theta_2 j^2)}. \tag{1.4}$$

As shown by Ghysels et al. (2007), this functional form allows for many different shapes. The weighting scheme can, for instance, be hump-shaped, declining, or flat. By

[23]We also estimate a model with only one regression coefficient for all included low-frequency lags. This, however, deteriorates forecasting performance. Results can be requested from the author.

[24]We also estimate a model with different polynomial specifications for every included low-frequency lag. However, this led to converging issues in the non-linear least squares (NLS) estimation for some periods, which deteriorates forecasting performance. Results can be requested from the author.

[25]For robustness, we also use a Legendre polynomial proposed by Babii et al. (2021). The results, however, are less promising and shown in Table 1.A.4 in the Appendix.

definition, they sum to one. Moreover, it parsimoniously represents the large number of predictors – with $P = 2$ we only have to estimate five parameters. The parameters are estimated by non-linear least squares (NLS). Since MIDAS models are a direct forecasting tool and depend on the forecast horizon H, we have to estimate a model for every H and re-estimate them whenever new information becomes available (here every day).

Bridge equation

Another common approach in the literature is the use of bridge equations that link the low-frequency variable and time-aggregated high-frequency indicators (See, e.g. Baffigi et al., 2004; Diron, 2008; Foroni & Marcellino, 2013). This approach is a two-step procedure. In the first step, the high-frequency variable has to be forecasted to the end of the desired quarter and then aggregated over time to obtain values corresponding to the low-frequency. In the second step, the aggregated values are used in the bridge equation to forecast the low-frequency variable. We estimate a bridge model for a forecast horizon H of the following form:

$$y_{t_q+H+1} = \alpha + \sum_{p=0}^{P-1} \beta_p L^p x_{t_q+H+1} + \varepsilon_{t_q+H+1} \tag{1.5}$$

where α is a constant, P is the number of lags, and the lag operator is defined as $L^1 x_{t_q} = x_{t_q-1}$. Note that

$$x_{t_q} = \sum_{k=0}^{K-1} \omega(k) L^{k/D} x_{t_d} = \sum_{k=0}^{K-1} \omega_k L^{k/D} x_{t_d} \tag{1.6}$$

is the time-aggregated high-frequency variable. The aggregation function, $\omega(k)$, depends on the nature of the indicator. Here it is a simple, equal-weighted average (i.e., $\omega_k = 1/D \quad \forall k$). The bridge equation in (1.5) can be estimated by OLS only on sample periods where all the high-frequency variables are available. We first need to forecast the high-frequency indicator to the end of the desired quarters to get a forecast of the low-frequency variable. To do so, we use an AR(p) model where the Bayesian Information Criterion (BIC) determines the lag order. These predictions are then aggregated according to equation (1.6) and plugged into the estimated equation (1.5).

Iterative MIDAS

The iterative MIDAS (MIDAS-IT) model was introduced by Schumacher (2016). It is an intermediate model between bridge and MIDAS. In principle, it is a bridge model where the aggregation function $\omega(k)$ is replaced with the restricted weighting polynomial $b(k, \theta)$. As for the bridge model, we use an AR(p) model to forecast the indicator variable until the end of the desired quarters. We use the same polynomial specification as for the MIDAS model. Using these three model types allows us to identify the advantages of selected aspects of MIDAS and bridge models.

Benchmarks

The forecasts are compared to three benchmarks. First, we use an autoregressive model (AR) of order one estimated on the corresponding real-time vintage for GDP growth. Second, using bridge equations, we forecast GDP growth using the KOF Economic Barometer, a well-known monthly composite leading indicator for Switzerland (Abberger et al., 2014). Because there is no real-time vintage of the KOF Barometer available, we use the release from March 2022. For the out-of-sample exercise, we assume that the Barometer value for the current month is available three days before the month ends. This is a reasonable assumption since the Barometer is usually published towards the end of each month. Third, we compare the forecasts to the preliminary quarterly GDP growth release for the respective quarter. Given that the quarterly GDP figures are revised after the initial release, we consider the initial quarterly GDP release to be a forecast of the final GDP outcome. To compute the forecast errors, we use the release of quarterly GDP from December 2021.

1.4 Evaluation of the *f*–curve

This section first provides a descriptive analysis of the *f*–curve and its underlying indicators. It then demonstrates the in-sample information content of the *f*–curve, highlighting that it is available earlier than most leading indicators. In addition, it provides an evaluation of its pseudo-out-of-sample performance for nowcasting (forecasting) real GDP growth. While the focus stays on real GDP growth, it should be noted that the *f*–curve is correlated with many key macroeconomic variables (See Figure 1.A.6 in the Appendix).

1.4.1 Descriptive analysis

Most of the indicators underlying the *f*–curve are substantially correlated with GDP growth (after pre-whitening with an AR(p) model (see Neusser, 2016, Ch. 12.1)). Figure

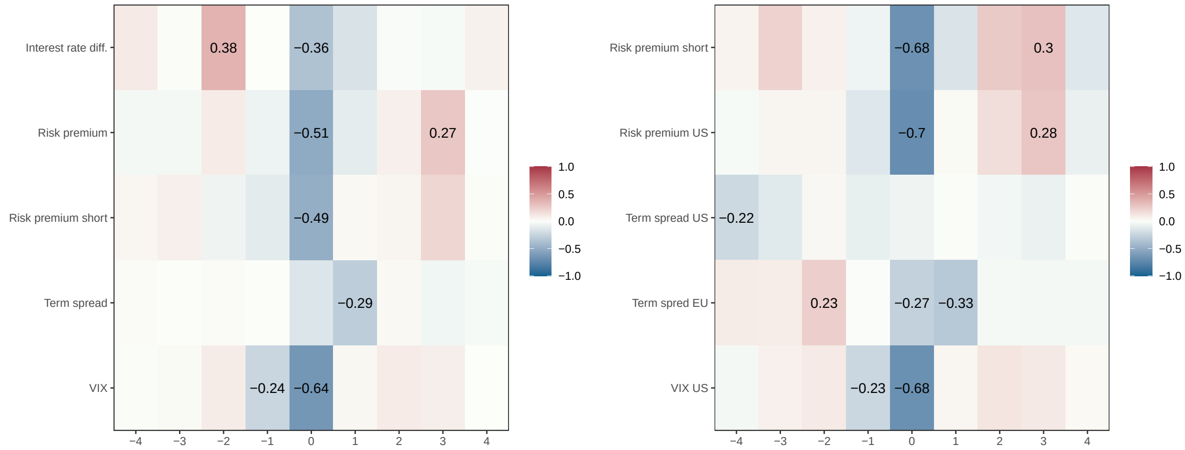
1.3 provides an overview of the cross-correlations. The absolute correlation coefficients range from 0.26 to 0.75. The coefficients for financial market data (panel a) and the text-based recession indices (panel b) are negative and significant. The term spreads are somewhat less strongly correlated. Moreover, term spreads for Switzerland and the euro area lead GDP growth by about half a year. This confirms that uncertainty measures such as risk premia and volatility indicators are highly correlated with the business cycle. The news sentiment indices in panel (b) feature a strong positive correlation with GDP growth. This correlation indicates that journalists' discussion of the economy holds valuable information about economic activity.

The f–curve is determined by computing the first principal component using all these variables. It is, therefore, valuable to assess the individual contributions of each variable to the factor. These contributions are represented by the factor loadings, which indicate how well the factor explains each original variable. Variables with higher absolute loadings are more strongly associated with the factor. Table 1.3 presents the factor loadings for the input variables, indicating their relationship with the f–curve. Positive loadings indicate a positive (negative) association with the business cycle (f–curve), while negative loadings suggest a negative (positive) association. The factor loadings show that the Swiss business cycle is well represented by news sentiment on the general economy, the financial markets, investment, and the labor markets. Risk premia, volatility indices, and text-based recession indices also contribute substantially to the factor. By contrast, news sentiments on inflation and politics, as well as term spreads and the interest rates differential, have lower loadings and contribute only a little. Finally, foreign variables tend to have higher factor loadings than domestic variables, reflecting Switzerland's strong dependence on foreign countries.

After updating the f–curve on daily basis over one year since May 2020, it is possible to judge the actual real-time performance of the indicator. Figure 1.4 provides preliminary results on how many input variables of the f–curve are available. The indicator is revised because not all data series are available in real-time (ragged edge problem). It shows results over the first one and a half year of daily updating the indicator. On average, more than 12 out of 24 series are available with a delay of one day. After three days, almost all indicators are available. The main reason why the average lies below 24 is that the archive of the *NZZ* was not accessible between November 2020 and March 2021. This led to a rather large revision. Moreover, the *Tages-Anzeiger* has not been updated for a couple of months in 2020. For this reason, we augmented the indicator with information from this newspaper's online edition. Finally, on rare occasions, the

Figure 1.3 — Cross-correlations of data underlying indicator with real GDP growth

(a) Financial market data



(b) Text-based indicators

Notes: Cross-correlation between financial market data (Panel a) and news-based sentiment indicators (Panel b) underlying the f–curve and real Swiss GDP growth. On the left, indicators for domestic data are displayed, while indicators for foreign data are on the right. The sample ends in 2020 Q1. We aggregate all data to quarterly frequency. Only statistically significant correlations at displacement s given on the x-axis are labeled. A significant correlation at $s > 0$ means the series is leading. Before computing the cross-correlation, the series have been pre-whitened with an AR(p) model (see Neusser, 2016, Ch. 12.1). The lag order has been determined using the Bayesian Information Criterion.

websites of financial sources were not available. All revisions to the indicator are solely because specific sources are unavailable at the time of the update. The inputs of the f–curve are not revised. The availability of inputs within one to two days and relatively small revisions if a data source was inaccessible confirms the indicator's excellent suitability for real-time tracking of the Swiss economy.

Table 1.3 — Factor loadings

Variable	Country	Loading
News Economy	Foreign	0.29
Long term Risk Premium	Foreign	-0.27
News Financial Market	Foreign	0.27
News Economy	Switzerland	0.26
Long term Risk Premium	Switzerland	-0.25
News Recession	Foreign	-0.24
News Recession	Switzerland	-0.24
VIX	USA	-0.23
News Investment	Foreign	0.23
Short term Risk Premium	Switzerland	-0.23
News Labor Market	Foreign	0.23
VIX	Switzerland	-0.22
News Financial Market	Switzerland	0.21
Short term Risk Premium	Foreign	-0.20
News Labor Market	Switzerland	0.19
News Investment	Switzerland	0.18
News Inflation	Foreign	0.18
News Politics	Foreign	0.18
News Inflation	Switzerland	0.14
News Politics	Switzerland	0.07
Term Spread	USA	0.07
Interest rate differential	Switzerland	-0.07
Term Spread	Euro Area	-0.03
Term Spread	Switzerland	-0.02

Notes: Factor loadings of the first principal component (f–curve). Positive loadings indicate a negative association with the factor, while negative loadings suggest a positive association.

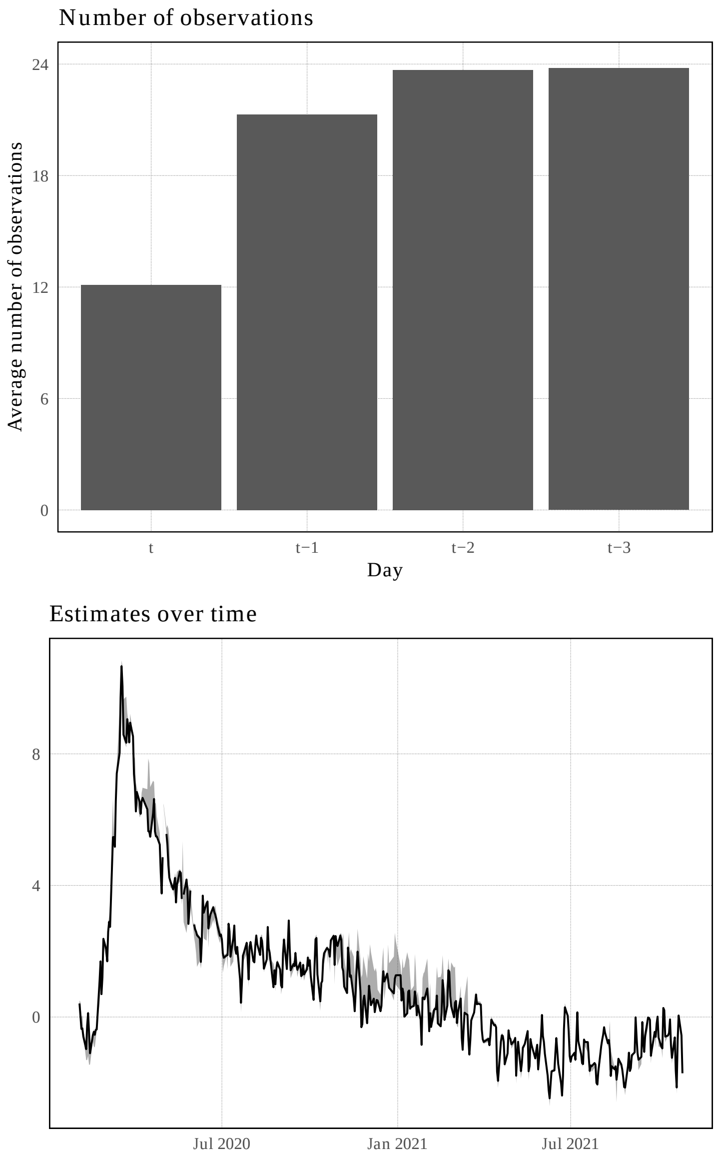
1.4.2 In-sample analysis

Because the f–curve is a combination of sentiment indicators covering several economic topics and financial market indicators, it is correlated with many key macroeconomic variables (see Figure 1.A.6 in the Appendix). Since it is not optimized to track any particular measure of economic activity in its current form, we evaluate the in-sample information content of the f–curve.

To compare the in-sample information content of the f–curve to other leading indicators, we perform a cross-correlation test (see Neusser, 2016, Ch. 12.1).[26] Figure 1.5 shows a

[26]It is noteworthy that other indicators are estimated or smoothed such that they undergo substantial revisions over time. Moreover, some of the indicators are published with significant delays (see Table 1.A.1 in the Appendix); finally, some are based on lagged data (see, e.g., OECD, 2010).

Figure 1.4 — Real-time results since initial version of the *f*–curve



Notes: Average number of variables available for calculation of the *f*–curve (top figure). The gray shaded area represents minimum and maximum estimates from 5 May 2020 to 27 October 2021 (bottom figure). Estimates over time are based on the published version (Burri & Kaufmann, 2020). See https://github.com/dankaufmann/f-curve for more information.

substantial correlation between the *f*–curve and many prominent leading indicators. There is a coincident or a leading relationship with the KOF Economic Barometer, SECO's Swiss Economic Confidence (SEC), consumer confidence, and trendEcon's perceived economic situation.[27] There is a leading, a coincident, and a lagging

[27] All data sources are given in Table 1.A.1 in the Appendix.

relationship with the Organisation for Economic Co-operation and Development composite leading indicator (OECD CLI) and with the SNB's Business Cycle Index (BCI). However, the OECD CLI is a smoothed indicator subject to substantial revisions. TrendEcon's perceived economic situation starts only in 2006 and the BCI is published with a relevant delay.

Another way of assessing the in-sample information of the f–curve is to test whether it is Granger causing other indicators (Granger, 1969). The f–curve, f_t, is said to Granger cause another indicator, I_t, if it contains statistically significant information about the future values of the other indicator. Therefore, the f–curve should Granger cause the indicator. To test this condition, the following model is estimated

$$I_t = \sum_i \alpha_i f_{t-i} + \sum_i \beta_i I_{t-i} + \varepsilon_t. \tag{1.7}$$

The joint hypothesis that $\alpha_i = 0$ is then tested using a Wald test. The rejection of the hypothesis suggests that the f–curve Granger causes the indicator. The other indicator should not Granger cause the f–curve. To test this condition, the following model is estimated

$$f_t = \sum_i \alpha_i f_{t-i} + \sum_i \beta_i I_{t-i} + \varepsilon_t \tag{1.8}$$

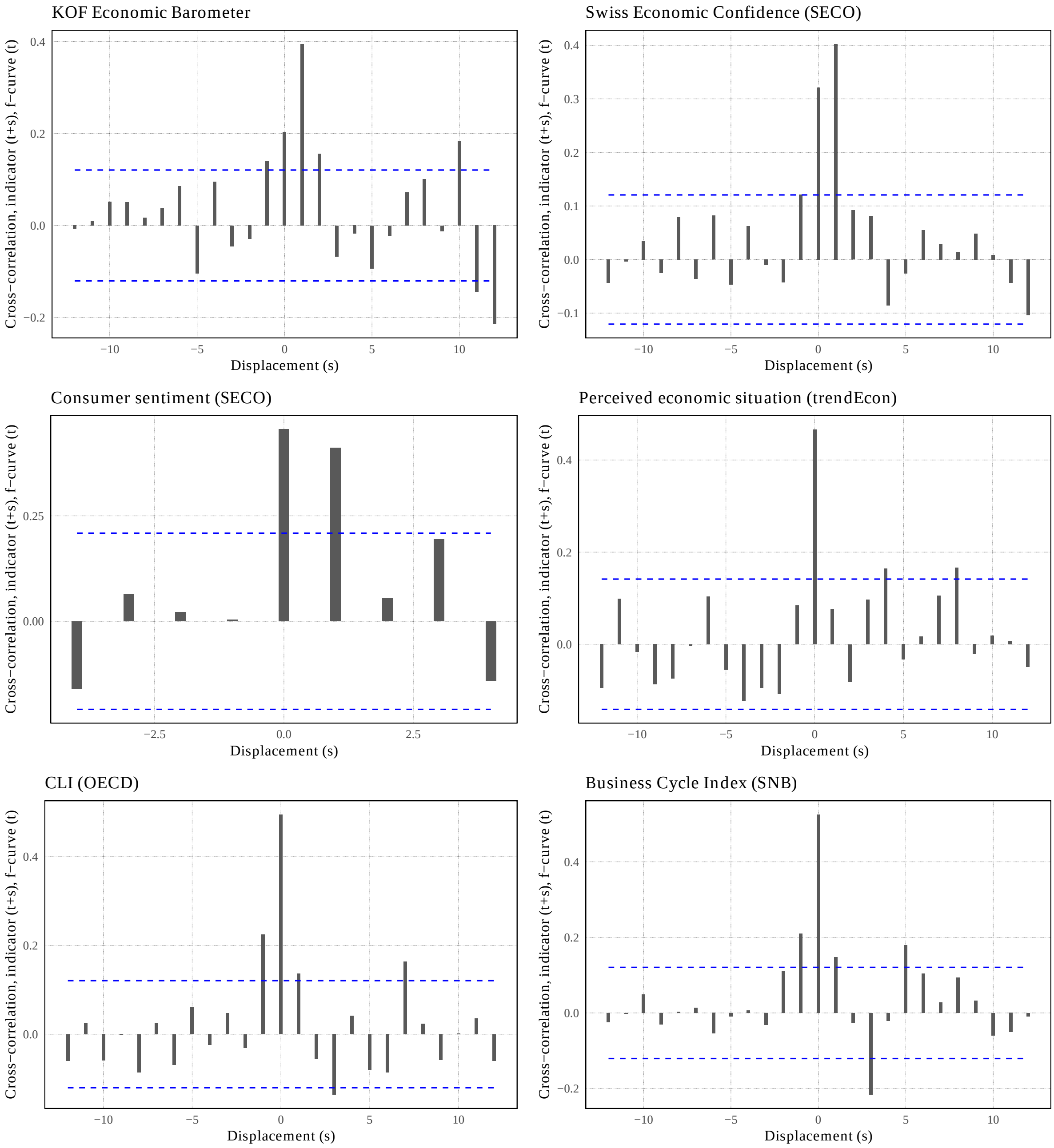
and the joint hypothesis that $\beta_i = 0$ is tested. To satisfy the condition, this hypothesis should not be rejected.

Table 1.4 — Testing Granger causality properties of f–curve

Indicator	f–curve granger causes Ind.	Ind. doesn't granger cause f–curve
KOF Barometer	✓	✓
SECO SEC	✓	✓
Consumer Sentiment	✓	✓
Trendecon	✓	✓
OECD CLI	✓	✗
SNB BCI	✓	✗
SECO WEA	✓	✓

Notes: All tests are conducted on the 5% significance level.

Figure 1.5 — Cross-correlation with other indicators



Notes: Cross-correlation between the *f*–curve and other prominent leading and sentiment indicators. We aggregate all data either to quarterly frequency (consumer sentiment) or monthly frequency (remaining indicators). The dashed lines give 95% confidence intervals. A bar outside of the interval suggests a statistically significant correlation between the indicators at a lead/lag of s. Before computing the cross-correlation, the series have been pre-whitened with an AR(p) model (see Neusser, 2016, Ch. 12.1). The lag order has been determined using the Bayesian Information Criterion. The only exception is the OECD CLI, for which an AR(4) model is used.

If both conditions are satisfied, there is a one-directional Granger causality. As shown in Table 1.4, the f–curve Granger causes all indicators under consideration. However, the f–curve is Granger caused by the OECD CLI and the SNB BCI. The reason for this might be technical and due to smoothing or estimation procedures to calculate the indicators. The tests for Granger causality confirm the conclusions from the cross-correlation tests above.

Overall, these results suggest the f–curve provides sensible information comparable with other existing indicators. The key advantage of the f–curve is its prompt availability and availability over a longer period.

Table 1.5 — Real-time evaluation: Relative RMSE and DMW tests

	Full sample				**Excluding Covid crisis**				**Excluding all crisis periods**			
Horizon	0	40	80	120	0	40	80	120	0	40	80	120
a) Hypothesis: Model > AR(1) model												
Bridge	0.64	0.84**	1.02	1.1	0.82*	0.78*	1.03	1.03	0.83***	0.76*	1	0.93
Midas	0.73	0.88*	0.98	1.08	0.77**	0.85	1.07	1.01	0.73***	0.82	1.07	0.92
Midas-IT	0.73	0.88*	0.98	1.06	0.78**	0.81*	1.01	0.98	0.75***	0.74**	0.98	0.88
b) Hypothesis: Model > Barometer bridge												
Bridge	1.04	0.93*	0.97	1.1	0.82**	0.86*	1.02	1.04	0.73**	0.81**	0.93	0.95
Midas	1.17	0.98	0.93	1.07	0.77**	0.94	1.05	1.02	0.65***	0.87	0.99	0.94*
Midas-IT	1.17	0.97	0.93	1.06	0.78**	0.89*	1	0.99	0.66***	0.78**	0.91	0.9*
c) Hypothesis: Model < First Release												
Bridge	1.97*				1.13				1.17*			
Midas	2.23*				1.06				1.03			
Midas-IT	2.23*				1.06				1.04			

Notes: Relative Root-mean-squared errors (RMSE) for forecasts with selected forecast horizons. A lower RMSE implies higher predictive accuracy compared to the benchmark. Horizons from 0 to 59 (60 - 119) denote forecasts for the current (next) quarter. The full sample spans from January 1st 2002 to December 31st 2021. The sample exluding the Covid-19 crisis excludes quarters 2-3 of 2020. The sample excluding all crisis periods additionally excludes the GFC from 2008 -2009. I use three benchmarks. First, I use an AR(1) model (panel a). Second, I use a bridge model with the KOF Economic Barometer (panel b). Third, I use the first quarterly release of the corresponding quarter (panel c). Note, I restrict the sample in panel (c) to years where the GDP figures already include official annual figures by the SFSO (up to 2020). The Diebold-Mariano-West (DMW) test provides a p-value for the null hypothesis of equal predictive accuracy against the alternative (where $A > B$ means A has lower RMSE than B) written in the row header (Diebold & Mariano, 2002; West, 1996). I assume a quadratic loss function. Significance levels are given by: *** $p < 0.01$, ** $p < 0.05$, *$p < 0.1$

1.4.3 Out-of-sample evaluation

Table 1.5 presents the relative root-mean-squared errors (RMSE) of the pseudo-real-time out-of-sample nowcasting exercise.[28] To better understand the results, we conduct a sub-sample analysis by excluding quarters two and three of 2020. Further, we exclude the Great Financial Crisis from 2008 to 2009 as a robustness check. The models used in the analysis do not significantly outperform the AR(1) model in forecasting GDP growth for horizons of 60 days and beyond (Table 1.5, panel (a)). This observation holds true across all sub-samples. However, when examining nowcasts for the current quarter's GDP growth (horizons of 0 to 59 days), the models consistently exhibit significantly lower RMSE. The full-sample nowcast (horizon of 0) is slightly lower than the benchmark, but this difference is not statistically significant due to large forecasting errors during the Covid–19 crisis.

Similar patterns are observed when employing a bridge equation model with the KOF Economic Barometer as the benchmark. Panel (b) of Table 1.5 shows that the f–curve does not exhibit superior performance compared to the KOF Barometer in forecasting GDP growth for the next quarter. However, focusing on the current quarter nowcast for both sub-samples, the f–curve significantly outperforms the Barometer bridge model. The relative RMSEs for the sample excluding the Covid–19 crisis are slightly higher than those for the sample also excluding the GFC.

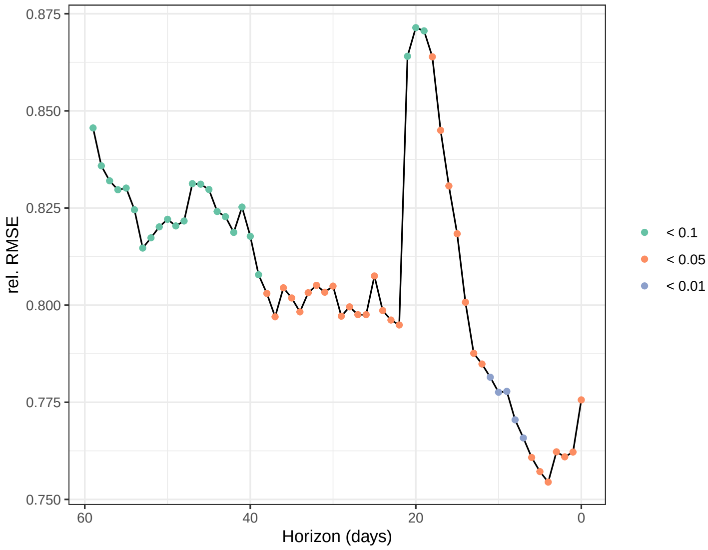
Moving to panel (c), the RMSE of the f–curve is higher than that of the first official GDP release, but the difference is mostly not statistically significant in either sub-sample. The key advantage of the f–curve is its ability to provide a full quarter's forecast approximately two months earlier than the first GDP release. Note that as the current vintage of GDP is subject to future revisions, particularly with the inclusion of annual GDP estimates based on comprehensive firm surveys by the SFSO, we restrict the sample in panel (c) to years where the GDP figures already incorporate these annual figures (up to 2020).

The results show that the MIDAS-IT model outperforms the other two models across most nowcasting horizons.[29] Additionally, the bridge model and the MIDAS model perform comparably well. These findings have important implications for utilizing mixed-frequency methods with daily data for nowcasting. Firstly, the bridge model's direct forecasting approach shows more promise than the MIDAS model's iterative

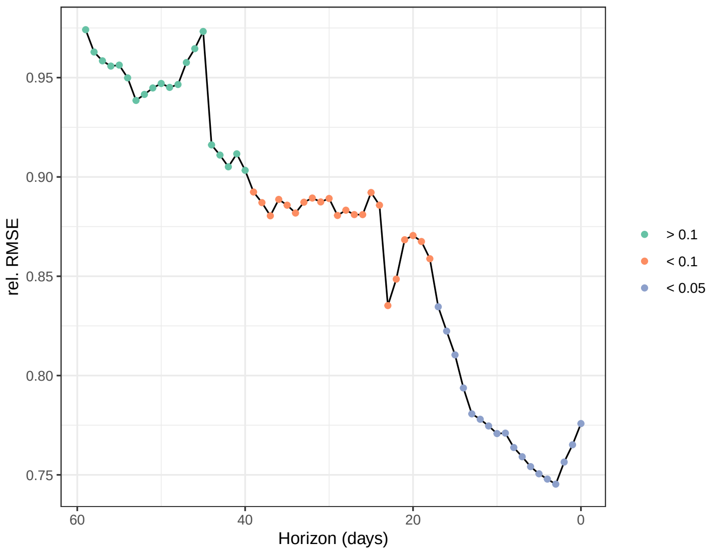
[28] For comparison, the results using the Legendre polynomial are shown in Table 1.A.4 in the Appendix.
[29] Figure 1.A.9 in the Appendix presents the absolute RMSE of the out-of-sample nowcasting exercise

Figure 1.6 — Relative RMSE of daily current quarter nowcasts

(a) Midas-IT vs. AR(1)



(b) Midas-IT vs. KOF Barometer bridge



Notes: Relative Root-mean-squared errors (RMSE) for forecasts with forecast horizons from 119 to 0 days. Periods of the Covid–19 crisis are excluded. A lower RMSE implies higher predictive accuracy compared to the benchmark. Horizons from 0 to 59 (60 - 119) denote forecasts for the current (next) quarter. We use two benchmarks. First, we use an AR(1) model (panel a). Second, we use a bridge model with the KOF Economic Barometer (panel b). The Diebold-Mariano-West (DMW) test provides a p-value for the null hypothesis of equal predictive accuracy against the alternative that the MIDAS-IT model is more accurate (Diebold & Mariano, 2002; West, 1996). We assume a quadratic loss function. Significance levels are given by: ● $p > 0.1$, ● $p < 0.1$, ● $p < 0.05$, ● $p < 0.01$

approach. Secondly, the employment of a nonlinear polynomial specification, despite requiring the estimation of more parameters, proves to be beneficial.

The performance of the f–curve compared to the benchmark AR(1) model and the bridge model with the KOF Barometer demonstrates its superiority for shorter horizons. The question arises: at what horizon does this difference become significant? Figure 1.6 illustrates the relative RMSE of the MIDAS-IT model compared to these benchmark models for the current quarter nowcast. To prevent significant nowcasting errors during the pandemic from distorting the analysis, the errors are calculated on the sample excluding the Covid-19 crisis.[30] The figure also presents p-values for Diebold-Mariano-West (DMW) tests, assessing the null hypothesis of equal predictive accuracy against the alternative hypothesis that the MIDAS-IT model is more accurate.

Examining panel (a) of Figure 1.6, which displays the relative RMSE against the AR(1) model, two key observations emerge. First, starting from a horizon of 40 days, the MIDAS-IT model significantly outperforms the AR(1) model in nowcasting the current quarter's GDP growth. Second, as indicated by spikes in the relative RMSE at a horizon of around 20 days, the AR(1) model demonstrates increased accuracy relative to the MIDAS-IT model when new GDP vintages are released. Panel (b) shows a similar behavior of the relative RMSE against the KOF Barometer bridge model. The f–curve starts surpassing the KOF Barometer at a horizon of around 40 days, approximately after the first month of the current quarter has passed.

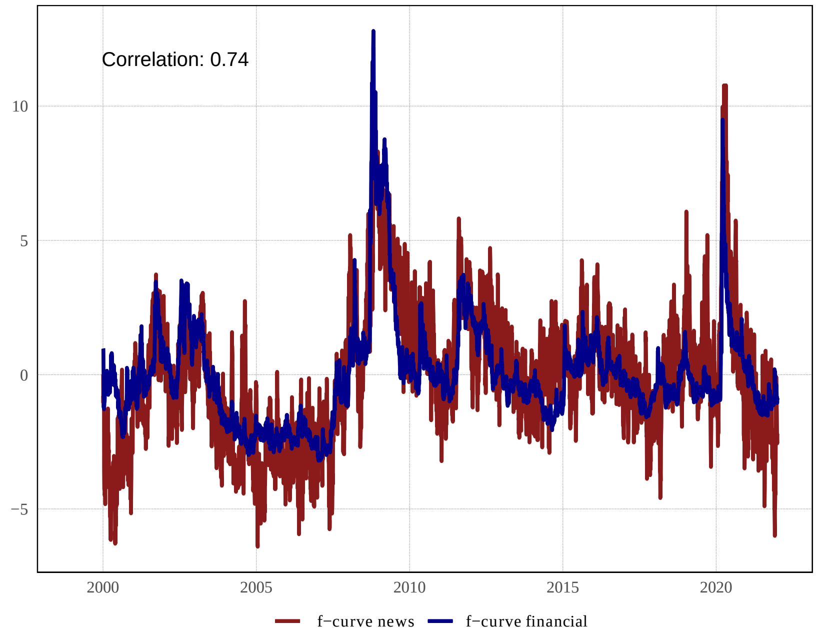
Are the financial market or news data more important for the forecasting performance of the f–curve? Figure 1.7 shows two indicators only calculated with financial market and news data, respectively. Although the indicators are highly positively correlated, there is one key difference. The news data are more volatile.[31] This suggests that financial market data provide a more accurate signal of the business cycle than the news data. Looking at relative RMSEs in tables 1.A.5 and 1.A.6 in the Appendix, this cannot be confirmed. For different models and different time horizons, sometimes the news-based f–curve and at other times the financial f–curve perform better. For larger time horizons, the financial f–curve outperforms the news-based f–curve. Hence, the more volatile news data does not worsen the f–curve because the factor model, including financial

[30]For comparison Figure 1.A.7 in the Appendix shows the evolution of the relative RMSE for the sample excluding all crisis periods.

[31]This is also because we smooth the news indicators with a moving average of only ten days. Comparable studies smooth over a much longer period. For example, Thorsrud (2020) uses a moving average of 60 days. On the one hand, this reduces the volatility of the news sentiment. On the other hand, this renders the indicator less helpful in detecting rapid daily changes.

market data, removes the idiosyncratic fluctuations. The overall f–curve therefore benefits more from the financial market data for some nowcasting horizons and more from the news data for others.[32]

Figure 1.7 — Comparison news and financial market data



Notes: The graph shows two indicators estimated only on financial market and news data, respectively.

Having established the usefulness of daily data for nowcasting, an important question arises: when does this data exert the most significant influence on forecast accuracy? To explore this, Figure 1.A.8 in the Appendix provides a comprehensive breakdown of squared error differences between the MIDAS-IT model (horizon 0) and the benchmark models. The lines in the graph represent the squared error differences in nowcasts compared to the benchmark model, with values below zero indicating superior performance of the f–curve-based model. The analysis reveals that most performance improvements relative to the benchmarks are concentrated during the financial crisis. In a broader context, forecast enhancements are often observed during turning points. These findings are consistent with recent literature on the use of textual data for predicting economic activity (see Barbaglia et al., 2023; Ellingsen et al., 2022; Kalamara et al., 2022). However, it is essential to acknowledge that the f–curve-based model exhibits certain weaknesses in nowcasting recovery periods. The Barometer bridge model, for example, was more accurate during the recovery from the GFC and the

[32] An interesting avenue for future research would be to conduct forecast encompassing tests (Harvey, 1989).

Covid–19 crisis. Nonetheless, it is crucial to note that the overall strong performance of the f–curve cannot be solely attributed to recession periods.

1.5 Concluding remarks

We develop a daily indicator of Swiss economic activity. The indicator consists of publicly available financial market and news data. It has the interpretation of a fever curve: As for monitoring the temperature of a patient, an increase in the fever curve provides a reliable and timely warning signal if health takes a turn for the worse.

A major strength of the indicator is that it can be updated with a delay of only one day and it is hardly revised ex-post. An evaluation of the indicator shows that it correlates with other business cycle indicators and accurately tracks Swiss GDP growth. In-sample assessments demonstrate a strong correlation between this indicator and several existing business cycle indicators. Furthermore, the analysis reveals that once one month of data becomes available within a given quarter, a model using the indicator outperforms a widely-used Swiss business cycle indicator in nowcasting GDP growth. In line with recent literature on the subject, the analysis demonstrates that the indicator significantly improves nowcasting performance compared to the benchmark models, particularly during turning points. Therefore, the f–curve provides an accurate and flexible framework to track Swiss economic activity at high frequency.

There is still room for improvement. We see six promising avenues for future research. First, the news sentiment indicators could exploit other publicly available news sources, particularly newspapers from Switzerland's French- and Italian-speaking parts. Second, instead of publicly available news lead texts, full texts of these articles could be incorporated. Third, customizing the lexicon specifically for economic news, as exemplified by Shapiro et al. (2022), has the potential to augment the accuracy of sentiment analysis in capturing economic trends. Fourth, we could examine the predictive ability of multiple factors and for other macroeconomic data. Fifth, the information could be used to disaggregate quarterly GDP and industrial production into monthly or weekly series. Finally, it would be valuable to investigate the potential gains obtained by incorporating all individual indicators in a nowcasting model, with the sparse-group LASSO-MIDAS model proposed by Babii et al. (2021) offering a promising avenue for exploration in this context. Exploiting all this new information will likely further improve our understanding of the health of the Swiss economy at high frequency.

References

Abberger, K., Graff, M., Siliverstovs, B., & Sturm, J.-E. (2014). The KOF economic barometer, version 2014. a composite leading indicator for the swiss business cycle. *KOF working papers No. 353*. https://doi.org/10.3929/ethz-a-010102658

Ardia, D., Bluteau, K., & 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. https://doi.org/10.1016/j.ijforecast.2018.10.010

Babii, A., Ghysels, E., & Striaukas, J. (2021). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 1–23. https://doi.org/10.1080/07350015.2021.1899933

Baffigi, A., Golinelli, R., & Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting*, *20*(3), 447–460. https://doi.org/10.1016/S0169-2070(03)00067-0

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221. https://doi.org/10.1111/1468-0262.00273

Bai, J., & Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, *176*(1), 18–29. https://doi.org/10.1016/j.jeconom.2013.03.007

Bai, Y., Li, X., Yu, H., & Jia, S. (2022). Crude oil price forecasting incorporating news text. *International Journal of Forecasting*, *38*(1), 367–383. https://doi.org/10.1016/j.ijforecast.2021.06.006

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. *Handbook of economic forecasting* (pp. 195–237). https://doi.org/10.1016/B978-0-444-53683-9.00004-9

Bańbura, M., Giannone, D., & Reichlin, L. (2012, September 18). Nowcasting. In M. P. Clements & D. F. Hendry (Eds.), *The oxford handbook of economic forecasting* (1st ed., pp. 193–224). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398649.013.0008

Barbaglia, L., Consoli, S., & Manzan, S. (2023). Forecasting with economic news. *Journal of Business & Economic Statistics*, *41*(3), 708–719. https://doi.org/10.1080/07350015.2022.2060988

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Becerra, A., Eichenauer, V. Z., Indergand, R., Legge, S., Martinez, I., Mühlebach, N., Oguz, F., Sax, C., Schuepbach, K., & Thöni, S. (2020). trendEcon. https://www.trendecon.org

Brown, M., & Fengler, M. (2020). MonitoringConsumptionSwitzerland. https://public.tableau.com/profile/monitoringconsumptionswitzerland

Burri, M., & Kaufmann, D. (2020). A daily fever curve for the swiss economy. *Swiss Journal of Economics and Statistics*, *156*(1), 6. https://doi.org/10.1186/s41937-020-00051-z

Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2023). Business news and business cycles. *Journal of Finance, Forthcoming*. http://dx.doi.org/10.2139/ssrn.3446225

Chicago Board Options Exchange. (2020). CBOE volatility index: VIX. https://fred.stlouisfed.org/series/VIXCLS

Deutsche Bundesbank. (2020). Zeitreihe BBK01.WT1010: Rendite der jeweils jüngsten bundesanleihe mit einer vereinbarten laufzeit von 10 jahren. https://www.bundesbank.de/dynamic/action/de/statistiken/zeitreihen-datenbanken/zeitreihen-datenbank/723452/723452?tsId=BBK01.WT1010

Dibiasi, A., & Iselin, D. (2016). *Measuring uncertainty* (KOF bulletin No. 101). KOF Swiss Economic Institute, ETH Zurich. https://ethz.ch/content/dam/ethz/special-interest/dual/kof-dam/documents/KOF_Bulletin/kof_bulletin_2016_11_en.pdf

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), 134–144. https://doi.org/10.1198/073500102753410444

Diron, M. (2008). Short-term forecasts of euro area real GDP growth: An assessment of real-time performance based on vintage data. *Journal of Forecasting*, *27*(5), 371–390. https://doi.org/10.1002/for.1067

Eckert, F., Kronenberg, P., Mikosch, H., & Neuwirth, S. (2020). Tracking economic activity with alternative high-frequency data. *KOF Working Papers No. 488*. https://doi.org/10.3929/ETHZ-B-000458723

Eichenauer, V. Z., Indergand, R., Martínez, I. Z., & Sax, C. (2021). Obtaining consistent time series from google trends. *Economic Inquiry*, ecin.13049. https://doi.org/10.1111/ecin.13049

Ellingsen, J., Larsen, V. H., & Thorsrud, L. A. (2022). News media versus FRED-MD for macroeconomic forecasting. *Journal of Applied Econometrics*, *37*(1), 63–81. https://doi.org/10.1002/jae.2859

European Central Bank. (2020). Yield curve spot rate, 1-year maturity - government bond, nominal, all issuers whose rating is triple. https://sdw.ecb.europa.eu/browseExplanation.do?node=qview&SERIES_KEY=165.YC.B.U2.EUR.4F.G_N_A.SV_C_YM.SR_1Y

Feinerer, I., & Hornik, K. (2019). *Tm: Text mining package*. https://CRAN.R-project.org/package=tm

Foroni, C., & Marcellino, M. G. (2013). A survey of econometric methods for mixed-frequency data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2268912

Galli, A. (2018). Which indicators matter? analyzing the swiss business cycle using a large-scale mixed-frequency dynamic factor model. *Journal of Business Cycle Research*, *14*(2), 179–218. https://doi.org/10.1007/s41549-018-0030-4

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. *CIRANO working papers 2004s-20*.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, *26*(1), 53–90. https://doi.org/10.1080/07474930600972467

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424. https://doi.org/10.2307/1912791

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the FOMC: A computational linguistics approach*. *The Quarterly Journal of Economics*, *133*(2), 801–870. https://doi.org/10.1093/qje/qjx045

Harvey, C. R. (1989). Forecasts of economic growth from the bond and stock markets. *Financial Analysts Journal*, *45*(5), 38–45. https://doi.org/10.2469/faj.v45.n5.38

Indergand, R., & Leist, S. (2014). A real-time data set for switzerland. *Swiss Journal of Economics and Statistics*, *150*, 331–352. https://doi.org/10.1007/BF03399410

Iselin, D., & Siliverstovs, B. (2013). The r-word index for switzerland. *Applied Economics Letters*, *20*(11), 1032–1035. https://doi.org/10.1080/13504851.2013.772290

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2022). Making text count: Economic forecasting using newspaper text. *Journal of Applied Econometrics*, *37*(5), 896–919. https://doi.org/10.1002/jae.2907

Kaufmann, D. (2020, April). *Wie weiter mit der tiefzinspolitik? szenarien und alternativen* (IRENE policy reports No. 20-01). IRENE Institute of Economic Research. https://ideas.repec.org/p/irn/polrep/20-01.html

Kaufmann, D., & Scheufele, R. (2017). Business tendency surveys and macroeconomic fluctuations. *International Journal of Forecasting*, *33*(4), 878–893. https://doi.org/10.1016/j.ijforecast.2017

Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, *27*(2), 529–542. https://doi.org/10.1016/j.ijforecast.2010.02.006

Larsen, V. H., & Thorsrud, L. A. (2018). Business cycle narratives. *CESifo Working Paper No. 7468*.

Lewis, D. J., Mertens, K., Stock, J. H., & Trivedi, M. (2022). Measuring real activity using a weekly economic index. *Journal of Applied Econometrics*, *37*(4), 667–687. https://doi.org/10.1002/jae.2873

Li, J., Li, G., Liu, M., Zhu, X., & Wei, L. (2022). A novel text-based framework for forecasting agricultural futures using massive online news headlines. *International Journal of Forecasting*, *38*(1), 35–50. https://doi.org/10.1016/j.ijforecast.2020.02.002

Luhn, H. P. (1960). Key word-in-context index for technical literature. *American Documentation*, *11*(4), 288–295. https://doi.org/10.1002/asi.5090110403

Neusser, K. (2016). *Time series econometrics*. Springer International Publishing. https://doi.org/10.1007/978-3-319-32862-1

OECD. (2010). Review of the CLI for 8 countries. *OECD Composite Indicators*. https://www.oecd.org/fr/sdd/indicateurs-avances/44556466.pdf

Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS - a publicly available german-language resource for sentiment analysis. *Proceedings of the seventh international conference on language resources and evaluation (LREC'10)*.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Schumacher, C. (2016). A comparison of MIDAS and bridge equations. *International Journal of Forecasting*, *32*(2), 257–270. https://doi.org/10.1016/j.ijforecast.2015.07.004

Scotti, C. (2016). Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises. *Journal of Monetary Economics*, *82*, 1–19. https://doi.org/10.1016/j.jmoneco.2016.06.002

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, *228*(2), 221–243. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.07.053

Shiller, R. J. (2019). *Narrative economics: How stories go viral & drive major economic events*. Princeton University Press.

SIX. (2020a). SBI–swiss bond indices. https://www.six-group.com/exchanges/indices/data_centre/bonds/sbi_en.html

SIX. (2020b). VSMI–volatility index on the SMI. https://www.six-group.com/exchanges/indices/data_centre/strategy_indices/vsmi_en.html

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162. https://doi.org/10.1198/073500102317351921

Stuart, R. (2020). The term structure, leading indicators, and recessions: Evidence from switzerland, 1974–2017. *Swiss Journal of Economics and Statistics*, *156*(1), 2–17. https://doi.org/10.1186/s41937-019-0044-4

The Economist. (2011). Up means down. the economist's gauge of gloom. 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. https://doi.org/10.1080/07350015.2018.1506344

Wegmüller, P., & Glocker, C. (2019). 30 indikatoren auf einen schlag. *Die Volkswirtschaft*, 19–22. https://dievolkswirtschaft.ch/de/2019/10/30-indikatoren-auf-einen-schlag/

Wegmüller, P., Glocker, C., & Guggia, V. (2023). Weekly economic activity: Measurement and informational content. *International Journal of Forecasting*, *39*, 228–243. https://doi.org/10.1016/j.ijforecast.2021.10.010

West, K. (1996). Asymptotic inference about predictive ability. *Econometrica*, *64*(5), 1067–84. https://doi.org/10.2307/2171956

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. *Proceedings of the 22nd international conference on World Wide Web*, 1445–1456. https://doi.org/10.1145/2488388.2488514

Appendix 1.A Supplementary material

Table 1.A.1 — Macroeconomic data and leading indicators

	Type	Publication	Frequency	Source	Comments
GDP	Target	+9 weeks	Quarter	SECO	First publication subject to further revisions
Employment	Target	+9 weeks	Quarter	SFSO	
Registered unemployment	Target	+1 week	Month	SECO	
ILO unemployment	Target	+6 weeks	Month	SFSO	
Output gap	Target	> +4 months	Quarter	SNB	
SNB Business Cycle Index	Indicator	> +2 months	Month	SNB	
Internet search sentiment	Indicator	+1 day	Day	trendEcon	Indicator based on internet search engine
KOF Economic Barometer	Indicator	+0 days	Month	KOF	Some underlying data probably missing at the end of the sample
Consumer sentiment	Indicator	+4 weeks	Quarter	SECO	Survey during first month of quarter. Indicator published at beginning of second month
OECD CLI	Indicator	> +1 week	Month	OECD	Many underlying data are lagged two months

Notes: Publication lags between the last day of the variable frequency (i.e. last day of the quarter or last day of the month) and the publication date of a recent release. Therefore, all publication lags are approximate and may change over time.

Figure 1.A.1 — Daily financial market indicators for *f*–curve

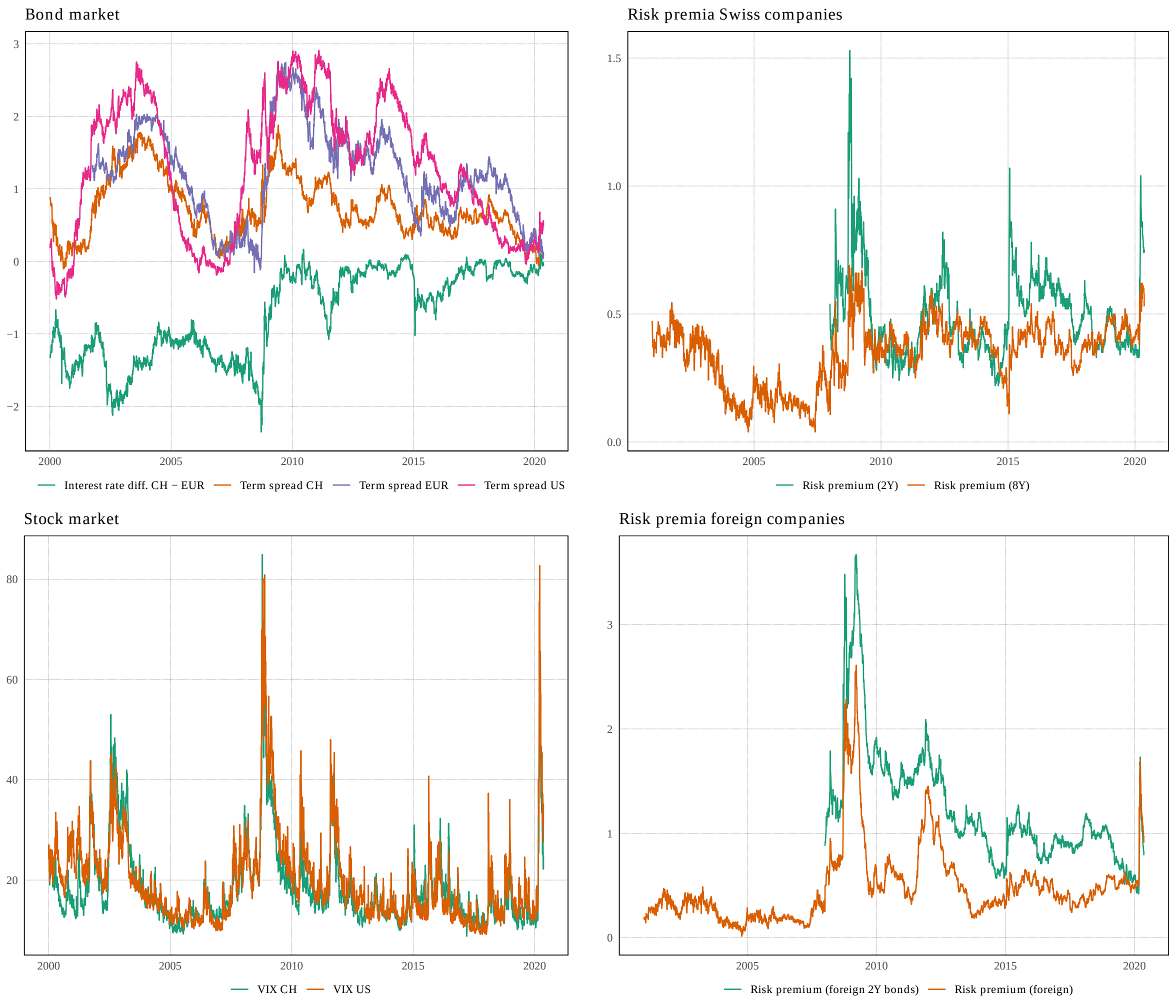


Table 1.A.2 — Data underlying f–curve

	Type	Publication	Frequency	Source	Comments
Term spread CH	f–curve	+1 day	Day	SIX, SNB	8Y – 2Y. SNB data used before SIX data available. Maturity of SIX data is approximate
Term spread USA	f–curve	+1 day	Day	Fed Board	10Y – 2Y
Term spread Europe	f–curve	+1 day	Day	Buba, ECB	10Y Germany – 1Y euro area. 1Y EUR Libor used before 2004
Risk premium CH	f–curve	+1 day	Day	SIX, SNB	8Y AAA-AA – 8Y government. SNB data for debt issues by banks used before SIX data available. Maturity of SIX data is approximate
Short-term risk premium CH	f–curve	+1 day	Day	SIX	1-3Y AAA-BBB – 1-3Y government. Start in 2008
Risk premium foreign	f–curve	+1 day	Day	SIX, SNB	8Y Foreign corp. – 8Y government. SNB data used before SIX data available. Average of various credit ratings
Short-term risk premium foreign	f–curve	+1 day	Day	SIX	1-3Y AAA-AA – 8Y government. Start in 2008
Stock market volatility CH	f–curve	+1 day	Day	SIX	
Stock market volatility USA	f–curve	+1 day	Day	CBOE	
Interest rate differential	f–curve	+1 day	Day	SIX, SNB, ECB	1-3 year government bonds CH – 1 year government bond yields euro area. 1Y EUR Libor used before 2004
Domestic news sentiment	f–curve	+1 day	Day	FuW, NZZ, TA	More details in Table 1.A.3
Foreign news sentiment	f–curve	+1 day	Day	FuW, NZZ, TA	More details in Table 1.A.3

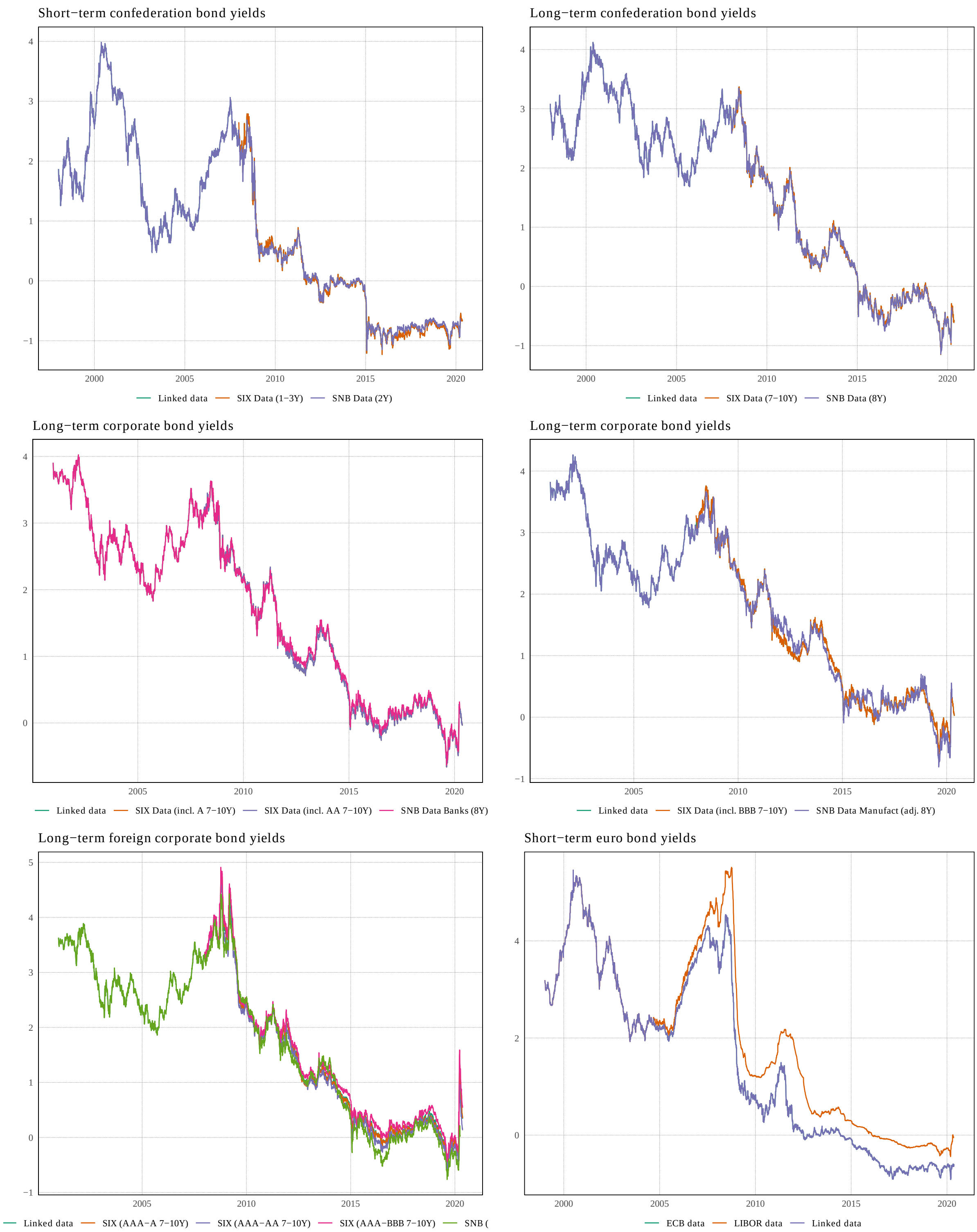
Notes: The SIX Swiss Exchange AG disclaimer applies to the SIX data: https://www.six-group.com/exchanges/download/market/data_services/six_disclaimer.pdf

Table 1.A.3 — Queries underlying news indicators

	URL	Keywords
	Domestic news sentiment	
FuW	fuw.ch/unternehmen/ fuw.ch/makro/	We use all articles listed in *Makro* and *Unternehmen* and select those containing the word *schweiz** in either lead text, tag or category.
NZZ	zeitungsarchiv.nzz.ch	[*konjunktur** OR *wirtschaft** OR *rezession**] AND *schweiz**
TA	tagesanzeiger.ch/zeitungsarchiv-930530868737	[*konjunktur* OR *wirtschaft* OR *rezession*] AND *schweiz*
TA Web	tagesanzeiger.ch	[*konjunktur* OR *wirtschaft* OR *rezession*] AND *schweiz*
	Foreign news sentiment	
FuW	fuw.ch/unternehmen/ fuw.ch/makro/	We use all articles listed in *Makro* and *Unternehmen* and select those containing [*ausland* OR *eu* OR *euro** OR *deutsch** OR *us** OR *amerika**] in either lead text, tag or cateogry.
NZZ	zeitungsarchiv.nzz.ch	[*konjunktur** OR *wirtschaft** OR *rezession**] AND [*ausland* OR *eu* OR *euro** OR *deutsch** OR *us** OR *amerika**]
TA	tagesanzeiger.ch/zeitungsarchiv-930530868737	[*konjunktur* OR *wirtschaft* OR *rezession*] AND [*ausland* OR *eu* OR *euro* OR *europa* OR *deutschland* OR *us* OR *usa* OR *amerika*]
TA Web	tagesanzeiger.ch	[*konjunktur* OR *wirtschaft* OR *rezession*] AND [*ausland* OR *eu* OR *euro* OR *europa* OR *deutschland* OR *us* OR *usa* OR *amerika*]

Notes: Since the *Finanz und Wirtschaft* is a business newspaper, we do not restrict the search with keywords related to the economy. The asterisk (*) represents a wildcard search operator. E.g. the query *schweiz** matches also *schweizerische*. Wildcards are allowed only in the NZZ archive.

Figure 1.A.2 — Spliced data underlying f–curve



Algorithm 1: Keyword in context for economic sentiment analysis

1. Define sets of keywords $\mathcal{K}$ describing the j topics.
2. Define context window size ws.
3. **for** *each set of keywords* $\mathcal{K}_j$ *in* $\mathcal{K}$ **do**

if $\mathcal{K}_j$ *is recession topic* **then**

a. **for** *each article a in each location i* **do**

i. Identify phrase $p = w_{t,a,i,n,j} \in \mathcal{K}_j$ that matches the recession topic (per article multiple phrases can match).

b. Calculate daily recession indicators, $S_{t,i,j}$, about the domestic and foreign economy by simply counting the matched phrases

$$S_{t,i,j} = P_{t,i,j}$$

else

a. **for** *each article a in each location i* **do**

i. Identify phrase $p = w_{t,a,i,n,j} \in \mathcal{K}_j$ that matches topic j (per article multiple phrases can match).

ii. Keep phrase p including ws terms before and after. Let $w_{t,p,i,j} = (w_{t,p,i,j,1}, w_{t,p,i,j,2}, ...w_{t,p,i,j,N_{t,p,i,j}})$ be the list of terms around phrase p. $N_{t,p,i,j}$, the total number of words is at most $2 * ws + 1$.

iii. Count the number of positive, negative, and the total number of words: $\sum_n \mathbb{1}(w_{t,p,i,j,n} \in \mathcal{P})$, $\sum_n \mathbb{1}(w_{t,p,i,j,n} \in \mathcal{N})$ and $N_{t,p,i,j}$

b. Calculate sentiment per matched phrase as

$$S_{t,p,i,j} = \frac{\sum_n \mathbb{1}(w_{t,p,i,j,n} \in \mathcal{P}) - \sum_n \mathbb{1}(w_{t,p,i,j,n} \in \mathcal{N})}{N_{t,p,i,j}}$$

c. Finally, daily news sentiment indicators,$S_{t,i,j}$, about the domestic and foreign economy for topic j are given by a simple average

$$S_{t,i,j} = \frac{1}{P_{t,i,j}} \sum_{p=1}^{P_{t,i,j}} S_{t,p,i,j}$$

where $P_{t,i,j}$ is the number of matched phrases.

Figure 1.A.3 — Daily news based recession indicators

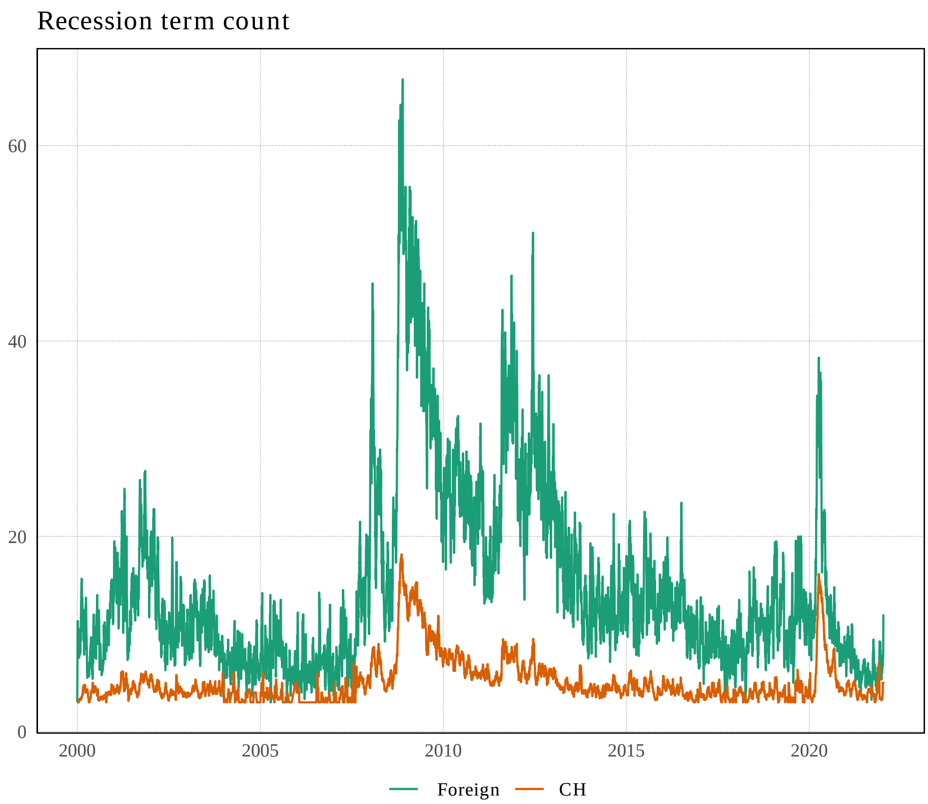


Figure 1.A.4 — Daily news based sentiment indicators

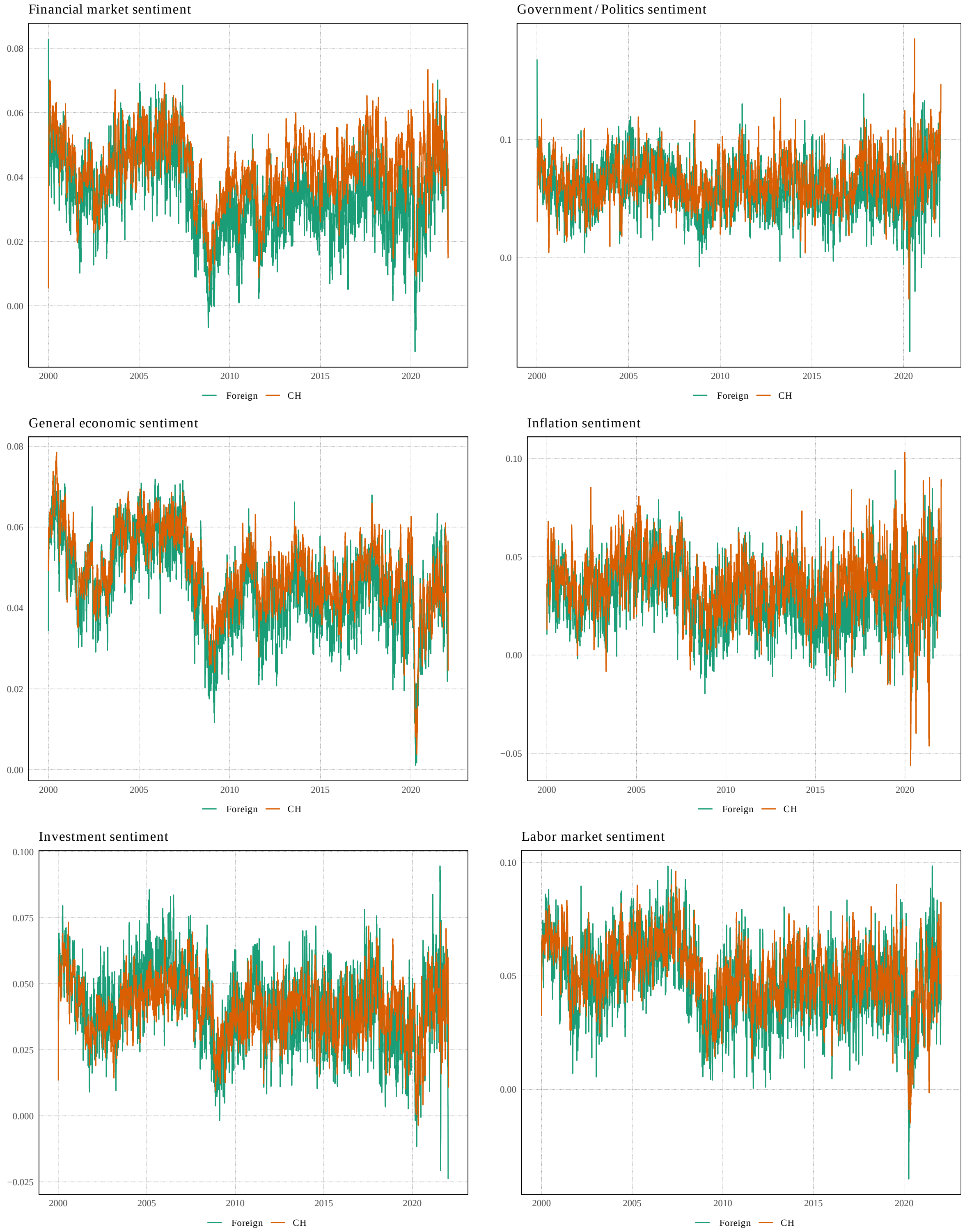
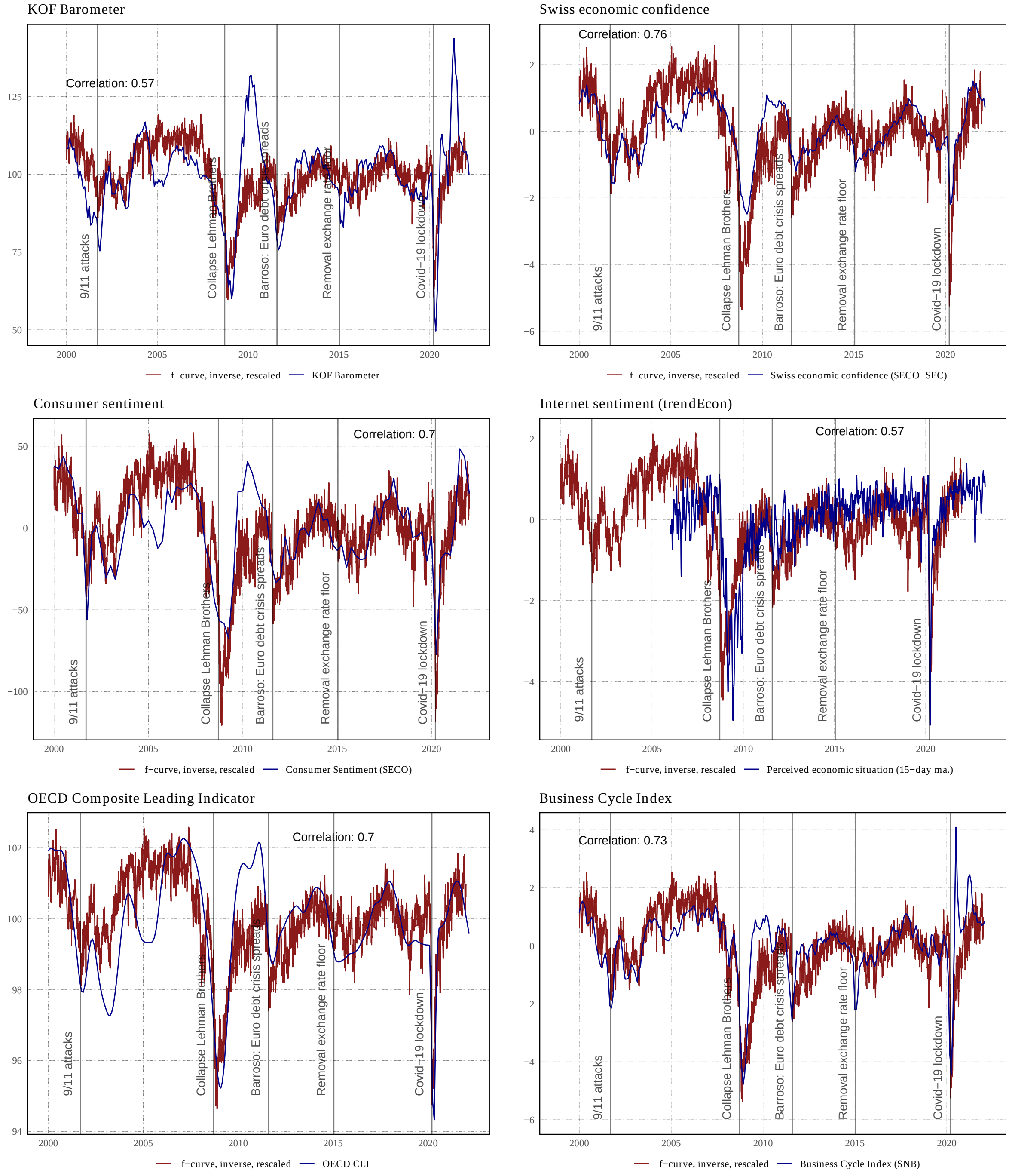
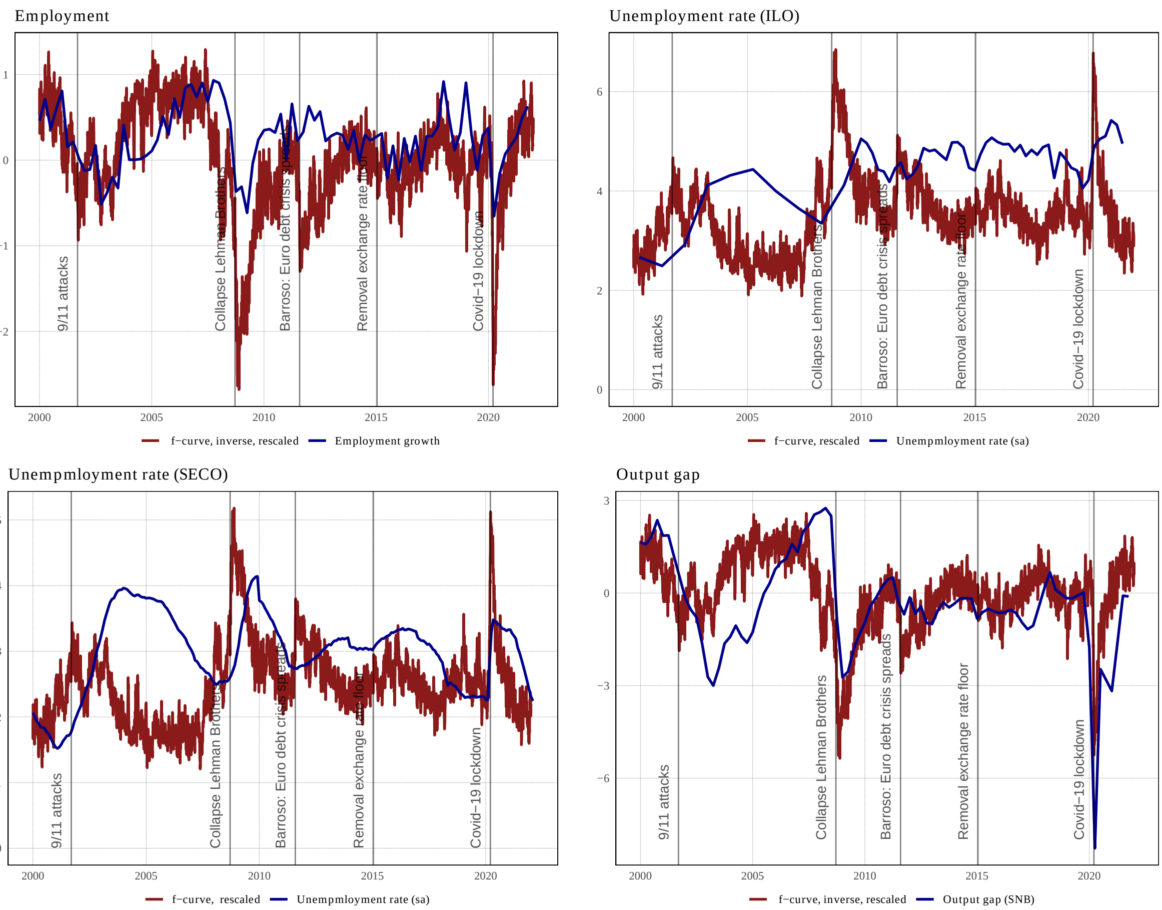


Figure 1.A.5 — Comparison with other indicators



Notes: f–curve rescaled such that it roughly matches the mean and volatility of the other data series.

Figure 1.A.6 — Comparison with other macroeconomic data



Notes: f–curve rescaled such that it roughly matches the mean and volatility of the other data series.

Table 1.A.4 — Relative RMSE and DMW tests using legendre polynomial

	Full sample				Excluding Covid crisis				Excluding all crisis periods			
Horizon	0	40	80	120	0	40	80	120	0	40	80	120
a) Hypothesis: Model > AR(1) model												
Bridge	0.64	0.84**	1.02	1.1	0.82*	0.78*	1.03	1.03	0.83***	0.76*	1	0.93
Midas	0.68	0.87*	1.06	1.09	0.83*	0.87	1.08	1.05	0.82**	0.83	1.04	0.96
Midas-IT	0.67	0.89*	1.02	1.11	0.82*	0.82*	1.09	1.05	0.82**	0.79*	1.09	0.97
b) Hypothesis: Model > Barometer bridge												
Bridge	1.04	0.93*	0.97	1.1	0.82**	0.86*	1.02	1.04	0.73**	0.81**	0.93	0.95
Midas	1.09	0.96	1	1.08	0.83*	0.96	1.06	1.06	0.73**	0.88	0.97	0.98
Midas-IT	1.09	0.98	0.97	1.1	0.82*	0.9	1.07	1.07	0.73**	0.84*	1.01	0.99
c) Hypothesis: Model < First Release												
Bridge	1.97*				1.13				1.17*			
Midas	2.06*				1.12				1.14*			
Midas-IT	2.05*				1.12				1.14*			

Notes: Relative Root-mean-squared errors (RMSE) for forecasts with selected forecast horizons. A lower RMSE implies higher predictive accuracy compared to the benchmark. Horizons from 0 to 59 (60 - 119) denote forecasts for the current (next) quarter. The full sample spans from January 1st 2002 to December 31st 2021. The sample exluding the Covid-19 crisis excludes quarters 2-3 of 2020. The sample excluding all crisis periods additionally excludes the GFC from 2008 -2009. I use three benchmarks. First, I use an AR(1) model (panel a). Second, I use a bridge model with the KOF Economic Barometer (panel b). Third, I use the first quarterly release of the corresponding quarter (panel c). Note, I restrict the sample in panel (c) to years where the GDP figures already include official annual figures by the SFSO (up to 2020). The Diebold-Mariano-West (DMW) test provides a p-value for the null hypothesis of equal predictive accuracy against the alternative (where $A > B$ means A has lower RMSE than B) written in the row header (Diebold & Mariano, 2002; West, 1996). I assume a quadratic loss function. Significance levels are given by: *** $p < 0.01$, ** $p < 0.05$, *$p < 0.1$

Table 1.A.5 — Relative RMSE and DMW tests for financial f–curve

	Full sample				Excluding Covid crisis				Excluding all crisis periods			
Horizon	0	40	80	120	0	40	80	120	0	40	80	120
a) Hypothesis: Model > AR(1) model												
Bridge	0.67	0.85**	1.01	1.07	0.77**	0.73*	1.02	0.98	0.83***	0.77*	0.96	0.84**
Midas	0.7	0.87**	0.99	1.06	0.71**	0.76*	1.07	1	0.76***	0.79*	1.01	0.91*
Midas-IT	0.72	0.87**	0.98	1.05	0.74**	0.74**	1	0.95	0.78**	0.77*	0.95	0.83*
b) Hypothesis: Model > Barometer bridge												
Bridge	1.09	0.95	0.96	1.07	0.77**	0.81*	1.01	1	0.73**	0.82**	0.89	0.86**
Midas	1.13	0.96	0.94	1.06	0.71***	0.84*	1.06	1.01	0.67***	0.84*	0.94	0.93*
Midas-IT	1.16	0.96	0.93	1.04	0.74**	0.81*	0.98	0.97	0.69**	0.82**	0.88	0.85**
c) Hypothesis: Model < First Release												
Bridge	2.07*				1.06				1.18**			
Midas	2.14*				0.96				1.06			
Midas-IT	2.19*				1				1.08			

Notes: Relative Root-mean-squared errors (RMSE) for forecasts with selected forecast horizons. A lower RMSE implies higher predictive accuracy compared to the benchmark. Horizons from 0 to 59 (60 - 119) denote forecasts for the current (next) quarter. The full sample spans from January 1st 2002 to December 31st 2021. The sample exluding the Covid-19 crisis excludes quarters 2-3 of 2020. The sample excluding all crisis periods additionally excludes the GFC from 2008 -2009. I use three benchmarks. First, I use an AR(1) model (panel a). Second, I use a bridge model with the KOF Economic Barometer (panel b). Third, I use the first quarterly release of the corresponding quarter (panel c). Note, I restrict the sample in panel (c) to years where the GDP figures already include official annual figures by the SFSO (up to 2020). The Diebold-Mariano-West (DMW) test provides a p-value for the null hypothesis of equal predictive accuracy against the alternative (where $A > B$ means A has lower RMSE than B) written in the row header (Diebold & Mariano, 2002; West, 1996). I assume a quadratic loss function. Significance levels are given by: *** $p < 0.01$, ** $p < 0.05$, *$p < 0.1$

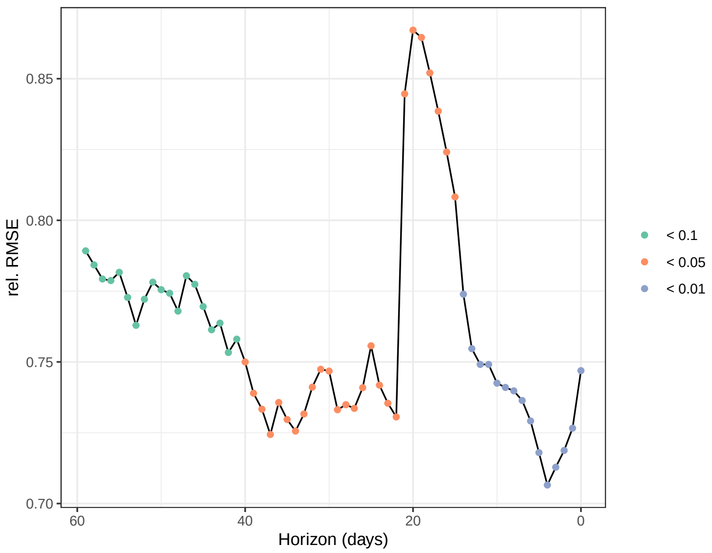
Table 1.A.6 — Relative RMSE and DMW tests for news-based f–curve

	Full sample				**Excluding Covid crisis**				**Excluding all crisis periods**			
Horizon	0	40	80	120	0	40	80	120	0	40	80	120
a) Hypothesis: Model > 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	1.08	0.78**	0.92	1.11	1.02	0.74***	0.9	1.11	0.98
Midas-IT	0.72	0.86*	0.98	1.07	0.8**	0.78**	0.99	0.96	0.77***	0.73**	0.98	0.88
b) Hypothesis: Model > 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	1
Midas-IT	1.17	0.96*	0.93	1.06	0.8**	0.86**	0.98	0.98	0.68***	0.77**	0.91	0.9
c) Hypothesis: Model < First Release												
Bridge	2.03*				1.2				1.22**			
Midas	2.13*				1.09				1.08			
Midas-IT	2.22*				1.08				1.05			

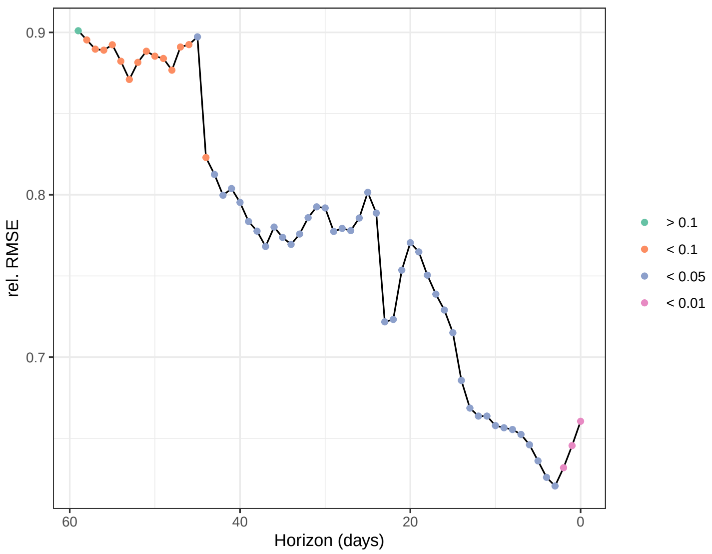
Notes: Relative Root-mean-squared errors (RMSE) for forecasts with selected forecast horizons. A lower RMSE implies higher predictive accuracy compared to the benchmark. Horizons from 0 to 59 (60 - 119) denote forecasts for the current (next) quarter. The full sample spans from January 1st 2002 to December 31st 2021. The sample exluding the Covid-19 crisis excludes quarters 2-3 of 2020. The sample excluding all crisis periods additionally excludes the GFC from 2008 -2009. I use three benchmarks. First, I use an AR(1) model (panel a). Second, I use a bridge model with the KOF Economic Barometer (panel b). Third, I use the first quarterly release of the corresponding quarter (panel c). Note, I restrict the sample in panel (c) to years where the GDP figures already include official annual figures by the SFSO (up to 2020). The Diebold-Mariano-West (DMW) test provides a p-value for the null hypothesis of equal predictive accuracy against the alternative (where $A > B$ means A has lower RMSE than B) written in the row header (Diebold & Mariano, 2002; West, 1996). I assume a quadratic loss function. Significance levels are given by: *** $p < 0.01$, ** $p < 0.05$, *$p < 0.1$

Figure 1.A.7 — Relative RMSE of daily current quarter nowcasts on sample excluding crisis periods

(a) MIDAS-IT vs. AR(1)



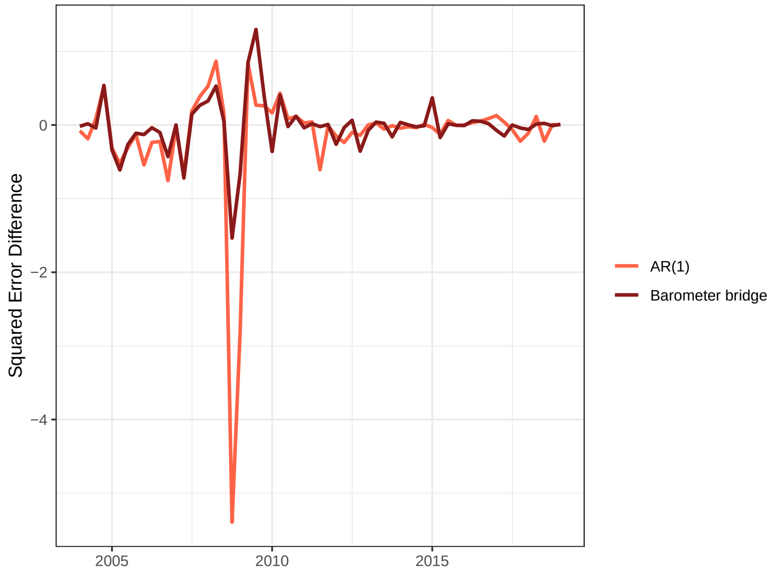
(b) MIDAS-IT vs. KOF Barometer bridge



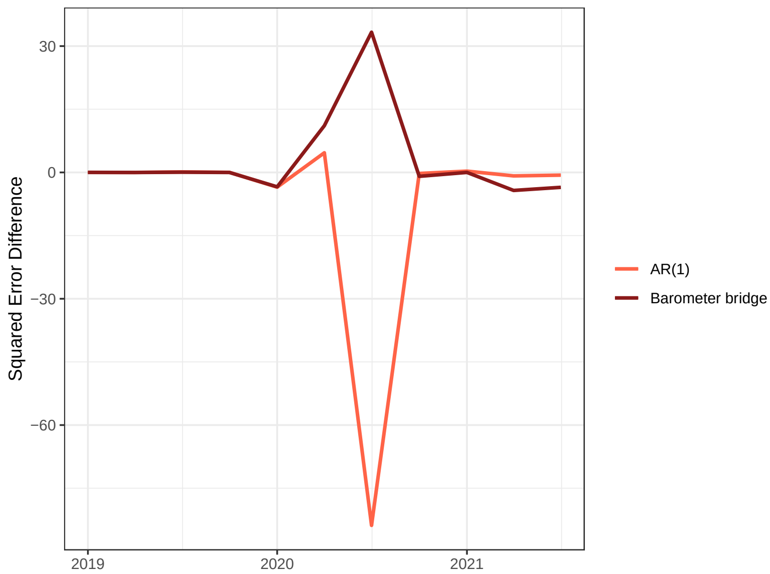
Notes: Relative Root-mean-squared errors (RMSE) for forecasts with forecast horizons from 119 to 0 days. Periods of the Covid–19 and the great financial crisis are excluded. A lower RMSE implies higher predictive accuracy compared to the benchmark. Horizons from 0 to 59 (60 - 119) denote forecasts for the current (next) quarter. We use two benchmarks. First, we use an AR(1) model (panel a). Second, we use a bridge model with the KOF Economic Barometer (panel b). The Diebold-Mariano-West (DMW) test provides a p-value for the null hypothesis of equal predictive accuracy against the alternative that the MIDAS-IT model is more accurate (Diebold & Mariano, 2002; West, 1996). We assume a quadratic loss function. Significance levels are given by: ● $p > 0.1$, ● $p < 0.1$, ● $p < 0.05$, ● $p < 0.01$

Figure 1.A.8 — Squared error differences: Midas.IT - Benchmark

(a) 2004 - 2019

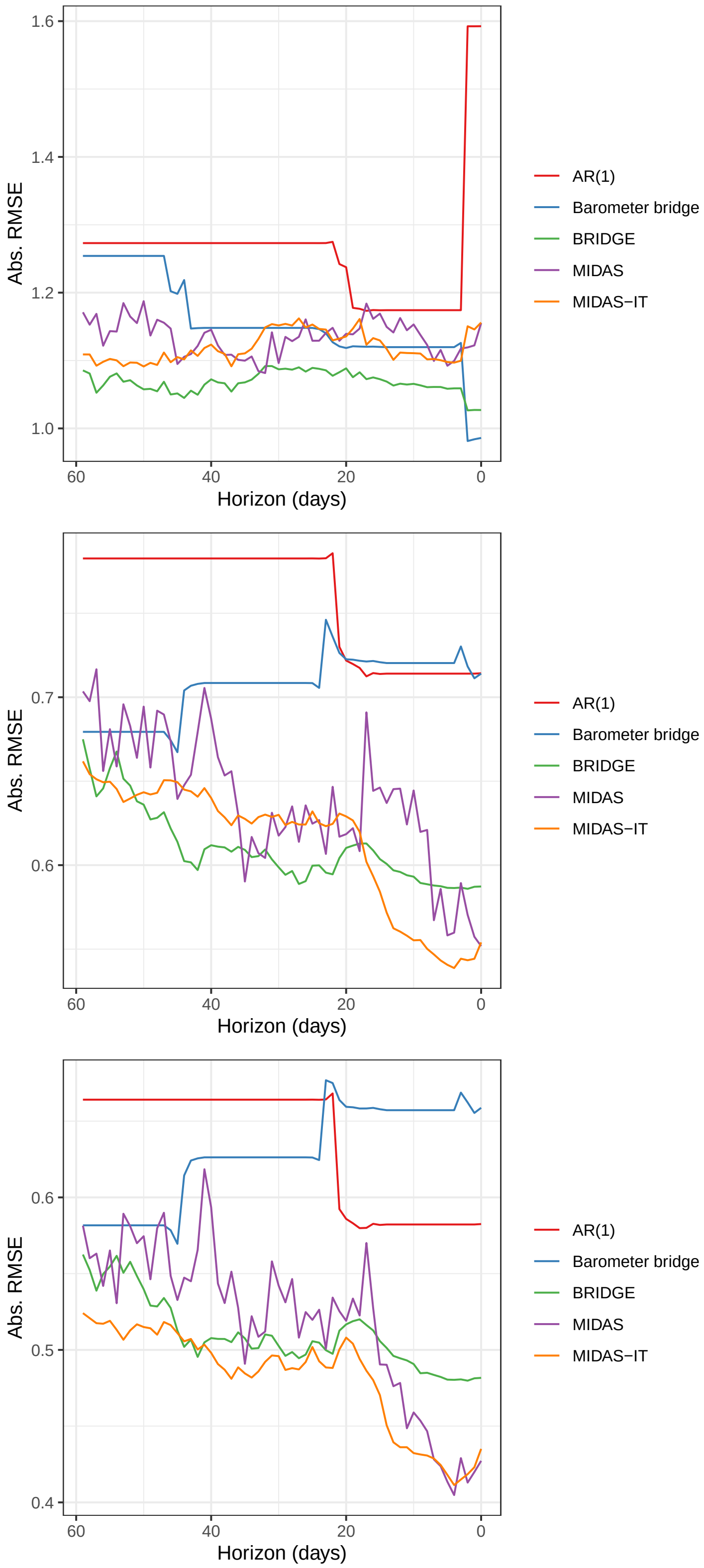


(b) 2019 - 2022



Notes: Squared error differences between Midas.IT model for horizon 0 and two benchmark models. The sample in panel (a) spans 2004 - 2019 and (b) 2019 - 2022.

Figure 1.A.9 — Absolute RMSE of daily current quarter nowcasts



Notes: RMSE for the current quarter nowcast. From top to bottom: Full sample, sample excluding Covid–19 crisis, sample excluding Covid–19 and GFC.

2 Exchange rate effects of US monetary policy – A multi-dimensional analysis using identification through heteroscedasticity†

2.1 Introduction

How do monetary policy decisions affect the exchange rate? To answer this question, economists resorted to theoretical and empirical approaches (Bjørnland, 2009; Dornbusch, 1976; Eichenbaum & Evans, 1995; Rogoff, 2002; Schmitt-Grohé & Uribe, 2022). However, there is no consensus about the size and persistence of the effects. Indeed, there is still a controversy about whether the exchange rate immediately overshoots, as in the theory by Dornbusch (1976). Theoretically, the response depends on rigidity in goods markets and the monetary policy regime (see, e.g., Benigno, 2004). Empirically, the response depends on the sample period (Kim et al., 2017) and the identification scheme (Bjørnland, 2009; Scholl & Uhlig, 2008). Some authors suggest that the exchange rate does not overshoot at all (Schmitt-Grohé & Uribe, 2022), or displays a delayed overshooting, for example, due to information rigidity (Müller et al., 2023). In addition, there is little evidence on whether different monetary policies, such as changes in the interest rate target, forward guidance, or large-scale asset purchases, affect the exchange rate differently.

This chapter aims to fill this gap. Our main contribution is twofold. First, we propose to combine a heteroscedasticity-based identification scheme with recursive zero restrictions and estimate the impulse responses with an instrumental variables (IV) approach. This

†This chapter is joint work with Daniel Kaufmann. We thank Martin Huber, Leif Anders Thorsrud, Eric Swanson, Mark W. Watson, Linyan Zhu, as well as participants at the IRENE PhD Meeting in Neuchâtel, the SSES Annual Congress in Lucerne, and the Workshop on Applied Macroeconomics and Monetary Policy in St. Gallen, for helpful comments.

allows us to identify multiple orthogonal monetary policy shocks exploiting the term structure of interest rates. Second, we apply our approach to investigate how various monetary policy shock dimensions in the United states affect the USD exchange rate at high (daily) and low (monthly) frequency.

Our identification strategy rests on two insights. First, we can identify a linear combination of a multi-dimensional monetary policy shock through heteroscedasticity, that is, using the difference in the variance of financial market variables during monetary policy event days and other days. Second, if we impose further restrictions, we can recover the underlying multiple dimensions of the monetary policy shock. Specifically, we impose recursive zero restrictions to identify a shock to the short-term interest rate (target shock), medium-term interest rate (path shock or forward guidance), and term spread (term premium shock or large-scale asset purchases). The zero restrictions impose that a path shock has no immediate impact on the short-term interest rate, while the term premium shock has no immediate impact on short- and medium-term interest rates. We show that dynamic causal effects can be estimated separately for every shock with a modification of the IV approach by Rigobon and Sack (2004). In addition, we can test for weak instruments and provide evidence on the existence of each shock by computing heteroscedasticity-robust F-statistics proposed by Lewis (2022b). Finally, we show how to estimate the monetary policy shock series from the term structure of interest rates, extending the procedure by Bu et al. (2021). This allows us to estimate the impulse responses of low-frequency macroeconomic variables and compare our shocks with existing high-frequency monetary policy surprises.

The main findings may be summarized as follows. The target, path, and term premium shocks we identify are correlated with the corresponding multi-dimensional high-frequency surprises for the United states by Swanson (2021). This suggests that our procedure is a valid alternative for countries and periods where high-frequency data is missing. In addition, we find evidence of a monetary policy target shock that affects financial markets via short-term interest rates. After a target tightening, the exchange rate immediately appreciates. Additionally, we find a path shock that affects financial markets through the 2Y interest rate. The path shock also leads to an (almost) immediate appreciation. Finally, we find evidence of a term premium shock affecting the 10Y - 2Y spread. Again, an increase in this spread appreciates the USD. For all three shocks, there is no evidence of a significantly delayed response. However, the estimation uncertainty at longer horizons is high. Therefore, we use the shocks to estimate monthly impulse responses in a structural vector-autoregression (SVAR) identified with external instruments. Although the exchange rate response is persistent, there is no evidence of a delayed overshooting puzzle.

Our first contribution relates to a large literature identifying causal effects of monetary policy using financial market data. The identification strategy is closely related to Canetg and Kaufmann (2022), who identify the effect of overnight rate and signaling shocks of central bank debt security auctions by the Swiss National Bank on the Swiss franc exchange rate. They assume the overnight rate shock immediately affects all financial market variables through a short-term interest rate. However, the signaling shock on impact affects only forward-looking variables, such as stock prices, but not the short-term interest rate. They estimate the impulse responses using a bootstrap algorithm, similar to Rigobon (2003). We show that impulse responses can be alternatively estimated with an IV approach, similar to Rigobon and Sack (2004).[1] Additionally, we show how to impose zero restrictions to disentangle multiple dimensions of the monetary policy shock.[2]

Following the seminal work by Kuttner (2001), most researchers use high-frequency identification schemes to identify multiple dimensions of monetary policy. The approach is based on the idea that, within a narrow window around monetary policy decisions, the only variation in financial market variables (mostly futures prices) stems from the policy decision. Gürkaynak et al. (2005) show that path surprises, associated with longer maturity interest rates, have a stronger effect on long-term interest rates than target surprises. Nakamura and Steinsson (2018) show that these high-frequency surprises comprise an 'information effect', that is, news the central bank communicates about the state of the economy. These information shocks may bias the results because they may lead to positive co-movement of interest rates and stock prices. In contrast, we would expect a negative co-movement in the case of a surprise tightening or loosening of monetary policy. Bauer and Swanson (2023b) argue that this information effect can be explained by the information available to the public before the monetary policy decision and, therefore, does not constitute inside information by the central bank. In addition, they highlight the relevance of speeches in addition to FOMC decisions. Because these monetary policy surprises are based on financial market instruments with varying

[1] The advantage of IV is that it saves computational time, is readily implemented in many software packages, and there is a large literature on testing for weak instruments.

[2] Lewis (2022a) estimates multiple dimensions of unconventional monetary policy announcements using intraday heteroscedasticity. His approach allows for varying importance of various shocks across announcements. He also finds that forward guidance has relevant effects. However, he also finds evidence of relevant information and large-scale asset purchase shocks.

maturities and because of recent non-conventional central bank policies, recent papers aim to disentangle various dimensions of monetary policy. Swanson (2021, 2024) and Altavilla et al. (2019) estimate multiple factors from a cross-section of high-frequency financial market data and rotate them so that they can be interpreted as a target, path, or large-scale asset purchase surprises.[3] Recently, Brennan et al. (2024) have shown that various measures of high-frequency surprises yield varying results.

Compared to the high-frequency literature, our approach has several advantages. High-frequency identification schemes remove background noise by computing changes in financial market variables in a very narrow window around a monetary policy announcement. This requires high-frequency data and exact knowledge of the intraday timing of the event. Our identification rests on comparing the variance-covariance of financial market variables on event and control days. The variance-covariance of financial market variables serves as a counterfactual of how markets respond in the absence of a monetary policy shock. Because our identification does not rest on the assumption that no other shocks occur during an event, we do not need to know the exact intraday timing of the event. In addition, we do not need high-frequency data and can use freely available daily data. Finally, the publication time of some events may be imprecise or completely unknown. The disadvantage is that the analysis may suffer from a weak instrument problem because daily data includes more background noise. However, standard tests exist to address this issue (Lewis, 2022b).

Our second contribution is related to a large literature estimating the effect of monetary policy on the exchange rate. Faust et al. (2003) use high-frequency surprises and estimate impulse responses with a VAR. They do not find evidence of a delayed overshooting of the exchange rate. Kearn and Manners (2006) estimate intraday dynamic responses for several countries. They find a statistically significant intraday effect of monetary policy on the exchange rate. Miranda-Agrippino and Rey (2022) use high-frequency monetary policy surprises in a VAR and estimate a persistent but not delayed response of the real exchange rate. Ciminelli et al. (2022) report similar findings to us, observing a peak in appreciation one week after the shock, then a gradual return to its initial level.

Delayed overshooting is mostly, but not always, found in VARs identified with zero or sign restrictions to estimate the causal effects of monetary policy on the exchange rate. Eichenbaum and Evans (1995) document a delayed overshooting puzzle using a VAR

[3]Recently, Schlaak et al. (2023) show how to combine high-frequency identification schemes with heteroscedasticity. As they exploit more exogenous variation, they can test the exogeneity of the instruments. They find evidence against the validity of high-frequency monetary policy surprises.

identified with zero restrictions. By contrast, Kim et al. (2017) suggest that this finding may be an artifact of including the volatile 1980s in the estimation sample. Bjørnland (2009) suggests that delayed overshooting disappears when long-term restrictions are used instead of short-term restrictions. Finally, Faust and Rogers (2003) and Scholl and Uhlig (2008) apply sign rather than zero restrictions and find little evidence in favor of delayed overshooting. However, similar to our results and Faust et al. (2003), the responses are imprecisely estimated. To the best of our knowledge, few papers examine the exchange rate response to different monetary policy shocks identified through heteroscedasticity. One notable exception is Wright (2012), who identifies monetary policy shocks at the effective lower bound through heteroscedasticity and provides the impact effect on bilateral exchange rates. However, he does not report dynamic causal effects, does not examine multiple dimensions of monetary policy, nor compares the resulting shocks to existing high-frequency surprises.

The remainder of the chapter is structured as follows. Section 2.2 presents the estimation and identification strategy. Section 2.3 presents the data and describes the baseline specification. Section 2.4 discusses the results before the last section concludes.

2.2 Estimation and identification

We identify dynamic causal effects of multiple dimensions of monetary policy in a heteroscedasticity-IV framework. We combine the IV-estimator suggested by Rigobon and Sack (2004) with recursive zero restrictions to identify orthogonal monetary policy shocks along the term structure of interest rates. In addition, we extend the approach by Bu et al. (2021) to estimate monetary policy shock series identified through heteroscedasticity via the term structure of interest rates to multi-dimensional shocks. Finally, we use these shocks as external instruments to identify a monthly SVAR.

2.2.1 Model and estimation

Suppose the data generating process reads:[4]

$$\begin{aligned} y_t = &\ \Psi\varepsilon_t + \Gamma v_t \quad \text{for } t \in P \\ y_t = &\ \Gamma v_t \quad \text{for } t \in C \end{aligned} \tag{2.1}$$

where y_t is a vector of N dependent variables, ε_t is a vector of E i.i.d. monetary policy shocks on policy event days (P), and v_t is a vector of N i.i.d. other shocks on policy

[4]We drop constant terms and lags of the dependent variable for ease of exposition.

event as well as control days (P and C). Furthermore, Γ and Ψ denote impact matrices of dimensions $N \times N$ and $N \times E$, respectively. Finally, we assume that Ψ is lower triangular.[5] We will justify this identifying assumption in more detail below.

Under these assumptions, we can sequentially estimate the causal impact of the E monetary policy shocks on y_t (Ψ) using a heteroscedasticity-IV estimator. The first equation of the model reads:

$$\begin{aligned} y_{1t} = &\quad \Psi_{11}\varepsilon_{1t} + \Gamma_1 v_t \quad \text{for } t \in P \\ y_{1t} = &\quad \Gamma_1 v_t \quad \text{for } t \in C \end{aligned} \tag{2.2}$$

where Ψ_{ij} denotes the ith row and jth column of Ψ and Γ_i denotes the ith row of Γ. Because Ψ is lower triangular, only the first monetary policy shock, ε_{1t}, affects y_{1t}.

As monetary policy shocks occur only on policy event days, the variance of y_{1t} differs between policy event and control days:

$$\begin{aligned} \mathbb{V}[y_{1t}] = &\quad \Psi_{11}^2\sigma_{1\varepsilon}^2 + \textstyle\sum_{n=1}^N \Gamma_{1n}^2\sigma_{nv}^2 \quad \text{for } t \in P \\ \mathbb{V}[y_{1t}] = &\quad \textstyle\sum_{n=1}^N \Gamma_{1n}^2\sigma_{nv}^2 \quad \text{for } t \in C \end{aligned} \tag{2.3}$$

where $\sigma_{e\varepsilon}^2$ and σ_{nv}^2 denote the variances of monetary policy shock e and other shock n, respectively.

Due to the recursive zero restrictions, the variance of y_{1t} changes only due to monetary policy shock 1. Thus, we can identify Ψ_{11} up to a scale from the difference in the variance on policy event and control days ($\Psi_{11}^2\sigma_{1\varepsilon}^2$). Intuitively, the variance of y_{1t} on control days serves as a counterfactual for the volatility of financial market variables in the absence of a monetary policy shock.

We can identify the impact on y_{2t} from changes in the covariance:

$$\begin{aligned} \mathbb{COV}[y_{1t}, y_{2t}] = &\quad \Psi_{11}\Psi_{21}\sigma_{1\varepsilon}^2 + \textstyle\sum_{n=1}^N \Gamma_{1n}\Gamma_{2n}\sigma_{nv}^2 \quad \text{for } t \in P \\ \mathbb{COV}[y_{1t}, y_{2t}] = &\quad \textstyle\sum_{n=1}^N \Gamma_{1n}\Gamma_{2n}\sigma_{nv}^2 \quad \text{for } t \in C \end{aligned} \tag{2.4}$$

where $\Psi_{11}\Psi_{21}\sigma_{1\varepsilon}^2$ corresponds to the difference in the covariances between policy event and control days. Having identified Ψ_{11}, this difference allows us to identify Ψ_{21}.

[5] Canetg and Kaufmann (2022) used this assumption to identify distinct overnight and signaling effects of central bank debt security auctions.

Because only the first shock affects the first variable, we can identify Ψ_{i1} for $i = 1, \ldots, N$ using a standard heteroscedasticity-based identification scheme (Rigobon, 2003). Therefore, we can estimate the impulse response using an IV-estimator (Lewis, 2022b; Rigobon & Sack, 2004). The instrument, the first, as well as the second stage read:

$$\begin{aligned} Z_{1t} &= \left[\mathbf{1}(t \in P)\frac{T}{T_P} - \mathbf{1}(t \in C)\frac{T}{T_C}\right] y_{1t} \\ y_{1t} &= \alpha_1 + \beta_1 Z_{1t} + u_{1t} \\ y_{it} &= \alpha_i + \tilde{\Psi}_{i1}\hat{y}_{1t} + e_{it} \end{aligned} \tag{2.5}$$

where $\mathbf{1}(t \in X)$ denotes an indicator function that equals one if the condition in parentheses is true and zero otherwise, and T, T_P, and T_C are the number of total, policy event and control days, respectively. In addition, Z_{1t} and $\hat{y}_{1t} = \hat{\beta}_1 Z_{1t}$ denote the instrument, as well as the first-stage prediction based on OLS estimates. Finally, α_i, β_i, and $\tilde{\Psi}_{ij}$ are regression parameters and u_{1t} and e_{it} are regression residuals.

Four comments are in order. First, the instrument is uncorrelated with v_t, and therefore e_{it}, because ε_{1t} occurs only during policy event periods (see e.g. Lewis, 2022b, for a detailed discussion). Second, the instrument is also uncorrelated with other monetary policy shocks because the variance of y_{1t} changes on event days only due to ε_{1t} (recursive zero restriction). Third, as we construct the instrument with y_{1t}, we assume that the first shock changes the variance of this variable on event days. As we will see below, we can construct an F-statistic to verify this. Finally, we identify the impulse responses only up to a scale because the impact on y_{1t} is normalized to unity, that is, $\tilde{\Psi}_{i1} = \Psi_{i1}/\Psi_{11}$.

We can then identify the causal effect of monetary policy shock 2 using the second equation of the model:

$$\begin{aligned} y_{2t} &= \Psi_{21}\varepsilon_{1t} + \Psi_{22}\varepsilon_{2t} + \Gamma_2 v_t \quad \text{for } t \in P \\ y_{2t} &= \Gamma_2 v_t \quad \text{for } t \in C. \end{aligned} \tag{2.6}$$

Because Ψ is lower triangular, the variance of y_{2t} changes on policy event days due to the first and second monetary policy shock. If we fail to control for shock 1, the instrument will be correlated with ε_{1t}, and the error term. Therefore, the exclusion restriction would be violated. Controlling for monetary policy shock 1 is straightforward by separately controlling for y_{1t} on event and control days. This is because, on control days, y_{1t} depends on ε_{1t} but not on ε_{2t}. We can construct an additional instrument using the second variable and then use both instruments to estimate Ψ_{21}. The instrument, the first, as well as the second stage, then read:

$$\begin{aligned} Z_{2t} &= \left[\mathbf{1}(t \in P)\frac{T}{T_P} - \mathbf{1}(t \in C)\frac{T}{T_C} \right] y_{2t} \\ y_{1t} &= \alpha_1 + \beta_{11} Z_{1t} + \beta_{12} Z_{2t} + u_{1t} \\ y_{2t} &= \alpha_2 + \beta_{21} Z_{1t} + \beta_{22} Z_{2t} + u_{2t} \\ y_{it} &= \alpha_i + \tilde{\Psi}_{i2}\hat{y}_{2t} + \tilde{\Psi}_{i1}\hat{y}_{1t} + e_{it}. \end{aligned} \tag{2.7}$$

On event days, y_{2t} and Z_{2t} are correlated with shocks ε_{1t} and ε_{2t}. However, by including $\hat{y}_{1t}$ as a control, the instrument will not be correlated with the error term. Again, we identify Ψ_{i2} only up to a scale because the initial response on y_{2t} is normalized to unity, that is, $\tilde{\Psi}_{i2} = \Psi_{i2}/\Psi_{22}$.

More generally, we can recursively identify the impact matrix of E-dimensional monetary policy shocks using the following instruments, first and second stages:

$$\begin{aligned} Z_{et} &= \left[\mathbf{1}(t \in P)\frac{T}{T_P} - \mathbf{1}(t \in C)\frac{T}{T_C} \right] y_{et}, \quad e = 1, \ldots, E \\ y_{et} &= \alpha_e + \sum_{j=1}^{E} \beta_{ej} Z_{jt} + u_{et}, \quad e = 1, \ldots, E \\ y_{it} &= \alpha_i + \sum_{j=1}^{E} \tilde{\Psi}_{ij}\hat{y}_{jt} + e_{it}, \quad i = 1, \ldots, N. \end{aligned} \tag{2.8}$$

We can extend this framework to include additional control variables and estimate cumulative daily impulse responses. Details are given in Appendix 2.A.

2.2.2 Identifying assumptions

The main identifying assumption is that policy events occur on pre-determined days (see Canetg & Kaufmann, 2022). For example, suppose the Federal Reserve is more likely to make decisions during economic distress. In that case, the variance of other shocks is different between policy event days and control days. Focusing on FOMC decisions, as well as planned speeches, fulfills this requirement. We will distinguish between scheduled and unscheduled FOMC decisions in a robustness test.

Furthermore, we assume that other shocks occur randomly across policy and control days. This assumption is violated if the FOMC schedules its policy meeting as a function of major economic data releases. For example, if FOMC meetings are usually scheduled after the release of quarterly GDP figures, there is a 'news' shock that will not affect the economy on policy event days. In addition, it may be that other central banks schedule their meetings briefly after the Federal Reserve to take into account policy surprises in their own decisions. Therefore, the assumption that the variance of other shocks is constant between policy event days and control days is violated. We account for this issue by excluding various other events from the control days across multiple robustness checks.

Finally, we impose recursive zero restrictions. These restrictions are unnecessary to identify the overall monetary policy shock, that is, the weighted average of all orthogonal dimensions. But, they serve to disentangle various orthogonal dimensions if they exist.[6] In our application, we order interest rates along the term structure, ordering a short-term interest rate first. The recursive zero restrictions imply that the first shock potentially affects all variables and has to increase the variance of the short-term interest rate. Therefore, it resembles the target surprise by Altavilla et al. (2019), which is a factor estimated on a high-frequency data set that loads only on short-term interest rates. Because of the zero restriction, our second shock has no immediate impact on the short-term interest rate but affects the medium-term interest rate. Therefore, it resembles the path surprise Altavilla et al. (2019) identified as being orthogonal to the target surprise. Finally, our term premium shock does not immediately impact short- or medium-term interest rates. But it has to affect the term spread. This resembles the QE shock by Altavilla et al. (2019), which is identified using high-frequency surprises that are orthogonal to the target and path surprises.

2.2.3 Weak instruments and number of monetary policy shocks

To test for weak instruments, we follow Lewis (2022b) and compute a heteroscadasticity-robust F-statistic for every instrument ($e = 1, \ldots, E$):

$$F_e = \frac{\hat{\beta}_e^2 \left(\sum_{t=1}^{T} Z_{et}^2 \right)^2}{\sum_{t=1}^{T} Z_{et}^2 \hat{u}_{et}^2} \tag{2.9}$$

where $\hat{\beta}_e$ and $\hat{u}_{et}$ are the OLS estimates of the first-stage coefficient and residuals,

[6]Swanson (2021) uses similar assumptions to identify three-dimensional monetary policy shocks. First, he imposes that changes in forward guidance and LSAP do not affect the current federal funds rate. Second, he imposes the restriction that the LSAP shock is as small as possible in the pre-ELB period. Therefore, our assumptions are weaker since we do not assume his second restriction.

respectively. Intuitively, the F-statistic increases in the absolute size of the first-stage regression coefficient. The instrument is stronger if the instrument is more highly correlated with the outcome variable. The correlation will be higher the more the variance changes during policy event days. Then, the F-statistic falls with a higher covariance between the instrument and the first-stage residuals. Intuitively, the instrument is weaker if it is more highly correlated with unobserved factors ('background noise'). As Z_{et} is generally correlated with u_{et}, an increase in the variance of u_{et} leads to an increase in the covariance between Z_{et} and u_{et}. Intuitively, if more of the variation in Z_{et} stems from background noise rather than the changes in the variance between policy event days and control days, the instrument will be weaker.

Interestingly, the F-statistic also indicates whether there is a relevant additional monetary policy shock dimension in the first place. Recall that we construct an additional instrument for equation e with variable y_{et}. Therefore, we assume that shock e changes the variance of this variable, conditional on shocks $1, \ldots, e-1$. Suppose there are E monetary policy shocks, but we estimate the first-stage regression for a nonexistent shock $e = E + 1$.[7] Then, the variance of $y_{E+1,t}$ increases during policy events only due to the shocks $1, \ldots, E$, for which we control by including the other instruments in the first-stage $Z_{jt}, j = 1, \ldots, E$. That is, conditional on the other shocks, the variance of $y_{E+1,t}$ does not increase on policy event days. Therefore, the OLS estimate on instrument $Z_{t,E+1}$ is zero, and $F_{E+1} = 0$ asymptotically.[8] This suggests that a positive F-statistic is a sufficient condition for the existence of a monetary policy shock.

2.2.4 Estimation of the monetary policy shock series

We adapt the method developed by Bu et al. (2021) to estimate three-dimensional monetary policy shock series from the term structure of interest rates. Bu et al. (2021) propose a two-step regression in the spirit of Fama and MacBeth (1973) to estimate a one-dimensional unobserved monetary policy shock, ε_t. First, they estimate the impact of a monetary policy shock on interest rate changes of maturities from 1 to 30 years via heteroscedasticity. Second, for every period with an FOMC decision, they perform a cross-sectional regression of the interest rate changes on the impact matrix. The OLS coefficients of these regressions are then proportional to the underlying unobserved monetary policy shocks.

[7]See also Appendix 2.B.

[8]Note, however, that a low F-statistic can be the result of a lot of background noise relative to the variance of an existing structural shock or of a sign of the absence of an additional shock.

Their approach can be readily extended to multiple monetary policy shocks. Suppose there are three orthogonal dimensions of monetary policy:

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}. \tag{2.10}$$

We construct an interest rate data set with the following ordering:

$$y_t = [i_t^{3m},\ i_t^{2y},\ i_t^{10y-2y},\ i_t^{20y}]' \tag{2.11}$$

where the first three variables are required to impose the recursive zero restrictions, we can add further interest rate data to estimate the shocks. For ease of exposition, we only add one additional interest rate.[9]

The model, therefore, reads:

$$y_t = \begin{bmatrix} \Psi_{11} & 0 & 0 \\ \Psi_{21} & \Psi_{22} & 0 \\ \Psi_{31} & \Psi_{32} & \Psi_{33} \\ \Psi_{41} & \Psi_{42} & \Psi_{43} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} + \Gamma v_t \quad \text{for } t \in P. \tag{2.12}$$

However, we can estimate each column of Ψ only up to scale, where we assume:

$$\Psi = \tilde{\Psi} \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \tag{2.13}$$

where a, b, and c are constants.

Bu et al. (2021) suggest regressing the vector of dependent variables on the impact

[9] In our application, we use the three and six-month interest rates and the one to thirty-year Treasury yields from the Gürkaynak et al. (2007) dataset to estimate the shocks.

matrix for every event day. In our multi-dimensional setting, we have one impact matrix for every shock and, therefore, a multi-variate regression: The OLS estimator of y_t on $\tilde{\Psi}$ on day t can be written as:

$$
\begin{aligned}
\hat{\varepsilon}_t &= [\tilde{\Psi}'\tilde{\Psi}]^{-1}\tilde{\Psi}'y_t \\
&= [\tilde{\Psi}'\tilde{\Psi}]^{-1}\tilde{\Psi}'\Psi\varepsilon_t + [\tilde{\Psi}'\tilde{\Psi}]^{-1}\tilde{\Psi}'\Gamma v_t \\
&= [\tilde{\Psi}'\tilde{\Psi}]^{-1}\tilde{\Psi}'\Psi\varepsilon_t + [\tilde{\Psi}'\tilde{\Psi}]^{-1}\tilde{\Psi}'\Gamma v_t \\
&= [\tilde{\Psi}'\tilde{\Psi}]^{-1}\tilde{\Psi}'\tilde{\Psi}\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}\varepsilon_t + [\tilde{\Psi}'\tilde{\Psi}]^{-1}\tilde{\Psi}'\Gamma v_t.
\end{aligned} \tag{2.14}
$$

OLS is an unbiased estimator if the term $[\tilde{\Psi}'\tilde{\Psi}]^{-1}\tilde{\Psi}'\Gamma v_t$ is zero in expectation. We can show that this is formally the case if $\mathbb{E}[\Gamma v_t|\tilde{\Psi}] = 0$, that is, we need an orthogonality assumption between $\tilde{\Psi}$ and Γv_t (see Appendix 2.C). This implies that the cross-sectional variation in the responses of the dependent variables to the monetary policy shocks is unrelated to the variation in the responses to other shocks. Under this assumption, we have:

$$
\mathbb{E}\begin{bmatrix} \hat{\varepsilon}_{1t} \\ \hat{\varepsilon}_{2t} \\ \hat{\varepsilon}_{3t} \end{bmatrix} = \mathbb{E}\begin{bmatrix} a\varepsilon_{1t} \\ b\varepsilon_{2t} \\ c\varepsilon_{3t} \end{bmatrix} \propto \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}. \tag{2.15}
$$

That is, we can recover the underlying shocks up to a scale. If this assumption is violated, the shocks will suffer from an unobserved variables bias because we fail to control for variation in Γv_t. This introduces time-varying 'background noise' via the other shocks v_t (see Appendix 2.C).

2.2.5 The structural VAR model

Having an estimate for the three dimensions of monetary policy shocks, it is of interest to see how these shocks affect macroeconomic variables on a lower frequency. To calculate the responses at a monthly frequency, we use the estimated shocks in an SVAR identified with external instruments (also known as proxy VARs) (Mertens & Ravn, 2013; Stock & Watson, 2018).

In a first step, we estimate a reduced-form monthly VAR with $n = 7$ macroeconomic variables:

$$Y_t = \alpha + B(L)Y_{t-1} + u_t \tag{2.16}$$

where Y_t is a vector of endogenous variables, $B(L)$ denotes the lag operator, and u_t is a vector of reduced-form residuals with covariance matrix $\mathbb{V}[u_t] = \Omega$. As is common in the literature, we use a specification with 12 monthly lags (Bauer & Swanson, 2023b; Gertler & Karadi, 2015; Ramey, 2016). We assume the relationship between the reduced-form residuals, u_t, and the structural shocks, ε_t, is linear

$$u_t = S\varepsilon_t \tag{2.17}$$

where S is the $n \times n$ structural impact matrix. Following common practice, we assume that the economy is governed by a sequence of uncorrelated structural shocks, ε_t, with $\mathbb{V}(\varepsilon_t) = I$. Without loss of generality, we assume that the first shock is a target shock, ε_{1t}, the second shock is a path shock, ε_{2t}, and the third shock is a term premium shock, ε_{3t}. Therefore, columns 1 through 3 of S specify the impact effects of different dimensions of monetary policy on u_t and Y_t. It follows from the variances of u_t and ε_t that

$$SS' = \Omega. \tag{2.18}$$

The goal is to identify the first three columns of S. However, infinitely many potential matrices S satisfy the equation (2.18). To estimate S, we need additional information or assumptions.

2.2.6 VAR identification using external instruments

We use the external instruments approach to estimate the effects of different dimensions of monetary policy. We use the monetary policy shocks as our instruments z_t, which are converted to a monthly series by summing the shocks on event days within each month.

The main assumption behind the external instruments approach is that the instruments are correlated with the structural shock of interest but uncorrelated with all other shocks. For example, this condition might be violated when the instrument for the target shock is not only correlated with the target shock but also with the term premium shock. Table 2.H.3 in the Appendix provides evidence that this might be the case. This correlation might result from a violated orthogonality condition when estimating the shocks (see Section 2.2.4). To account for this, we jointly identify the three dimensions of the monetary policy shock. The identifying restrictions are given by

$$\mathbb{E}\left[z_t \varepsilon_{1,t}'\right] = \Phi \tag{2.19}$$

$$\mathbb{E}\left[z_t \varepsilon_{2,t}'\right] = 0_{k\times(n-k)} \tag{2.20}$$

where Φ is a $k \times k$ matrix of full rank. $\varepsilon_{1,t}$ is the vector of k structural shocks to be identified, and $\varepsilon_{2,t}$ denotes all the other structural shocks. The relevance (2.19) and exogeneity (2.20) conditions together with the variances for u_t and ε_t imply

$$\mathbb{E}\left[z_t u_t'\right] = \mathbb{E}\left[z_t \varepsilon_t'\right] S' = \mathbb{E}\left[z_t \begin{pmatrix} \varepsilon_{1,t}' & \varepsilon_{2,t}' \end{pmatrix}\right] \begin{pmatrix} S_1' \\ S_2' \end{pmatrix} = (\Phi, 0) \begin{pmatrix} S_1' \\ S_2' \end{pmatrix} = \Phi S_1' \tag{2.21}$$

$$\mathbb{E}\left[z_t u_t'\right] \mathbb{E}\left[u_t u_t'\right]^{-1} \mathbb{E}\left[u_t z_t'\right] = \Phi S_1' \left(S S'\right)^{-1} S_1 \Phi' = \Phi S_1' (S') S^{-1} S_1 \Phi' = \Phi\Phi' \tag{2.22}$$

where $S^{-1}S_1 = \begin{pmatrix} I_k & 0_{(n-k)\times k} \end{pmatrix}'$ and S_1 is the first k columns of S. If $k = 1$, Φ is a scalar, and the identification is unique up to sign and scale. If $k > 1$, Φ has k^2 unique elements, while $\Phi\Phi'$ is symmetric with only $\frac{k(k+1)}{2}$ unique elements. Hence, the instrument moment restrictions are not sufficient. Therefore, we additionally impose the restriction that the path and the term premium shocks do not affect the three-month interest rate on impact. Moreover, the term premium shock does not affect the two-year Treasury yield on impact. These additional assumptions identify the three structural shocks. We provide more details on the joint identification of k structural shocks with k instruments in Appendix 2.D.

2.3 Data

In what follows, we present the dependent and control variables used to estimate daily and monthly exchange rate responses, policy event days, and control days. We use daily data from 1988–2022.[10] The monthly variables span from January 1973 to February 2020. Note that in an SVAR identified with external instruments, the estimation and identification sample do not need to be congruent. Following Bauer and Swanson (2023b), we use the sample spanning January 1988 to December 2019 for identification. The exact data sources are listed in Appendix 2.E.

[10]There are some missing values due to weekends and public holidays. We remove all these values and interpolate a few additional missings before transforming the data.

2.3.1 Dependent variables

We use exchange rate data recorded at noon EST from the Federal Reserve Board. Besides a nominal effective exchange rate, we also examine the USD exchange rate vis-à-vis the CHF, GBP, JPY, CAD, and EUR. Before the euro-changeover, we use the USD/DEM exchange rate.[11] All exchange rates are defined as one USD in terms of foreign currency. A decrease in the exchange rate is an appreciation of the USD. All exchange rates are included as log-changes, multiplied by 100, such that the cumulative impulse responses are measured in percent.

We use interest rates along the term structure to identify multiple dimensions of monetary policy shocks on daily frequency. Specifically, we use the three-month interest rate to represent the target shock and the two-year Treasury bill and a term spread (ten-year minus the two-year Treasury bill) to represent the path and term premium shocks, respectively. They are recorded at market close, typically 4 pm EST. We include them in first-differences so that the cumulative impulse responses are measured in percentage points.

Recall that the model reads

$$y_t = \sum_{l=1}^{L} \Phi_l y_{t-l} + \Psi \varepsilon_t + \Gamma v_t \quad \text{for } t \in P \tag{2.23}$$

where we include $L = 4$ lags of the dependent variables.

In the baseline, we set

$$y_t = [i_t^{3m},\ i_t^{2y},\ i_t^{10y-2y},\ \text{neer}_t]' \tag{2.24}$$

with

$$\Psi \varepsilon_t = \begin{bmatrix} \Psi_{11} & 0 & 0 \\ \Psi_{21} & \Psi_{22} & 0 \\ \Psi_{31} & \Psi_{32} & \Psi_{33} \\ \Psi_{41} & \Psi_{42} & \Psi_{43} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}. \tag{2.25}$$

The ordering implies that the target shock (ε_{1t}), which changes the variance of the

[11] The USD/DEM is transformed to a hypothetical USD/EUR using the official euro-changeover exchange rate.

three-month interest rate, affects on impact all variables.[12] The path shock (ε_{2t}), which changes the variance of the two-year rate (conditional on the target shock), does not affect the three-month rate on impact. The term premium shock (ε_{3t}), which changes the variance of the term spread (ten-year minus two-year rate), conditional on the target and path shocks, does not affect the three-month and two-year rate on impact. In addition, we include the nominal effective exchange rate as the outcome variable of main interest.

We estimate a reduced-form VAR with $n = 7$ macroeconomic variables to calculate the responses at monthly frequency. As has become standard in monetary policy VARs following Gertler and Karadi (2015), we use the log of industrial production, the log of the consumer price index (CPI), and the Gilchrist and Zakrajšek (2012) excess bond premium. Additionally, we add the log of the nominal exchange rate index. We add the three-month interest rate to represent the target dimension of monetary policy. For the path component, we follow Swanson (2024) and add the two-year Treasury yield. Finally, we add a term spread (ten-year minus two-year Treasury yield) to represent the term premium dimension. Gertler and Karadi (2015) use the one-year instead of the two-year Treasury yield. However, the two-year rate was unconstrained during the ZLB, making it a better stance of the path component of monetary policy.[13] Following Bauer and Swanson (2023b) and Brennan et al. (2024) we use the end-of-month values for all interest rates. Our sample spans January 1973 to February 2020. We end the sample in February 2020 because we do not want to estimate a model with the large swings of industrial production during the Covid–19 pandemic. All variables except interest rates and the bond premium are expressed in logarithms multiplied by 100. Therefore, their responses are measured in percent, while the responses of the interest rates and the bond premium are in percentage points.

2.3.2 Events

As monetary policy events, we use the 323 FOMC announcement dates (announcements only) for the period 1988-2019 by Swanson and Jayawickrema (2023), extending them to include 2020–2022. We end up with 344 FOMC announcement dates, whereby 284 are regularly scheduled FOMC meetings.[14] The remaining 60 correspond to unscheduled

[12] In the baseline, we use the three-month rate instead of the Federal Funds Rate to identify the target shock because FOMC meetings do not occur every day. Therefore, target surprises are usually expected to change the interest rate for more than one day.

[13] The results are very similar whether we use the one- or two-year Treasury yield in our analysis.

[14] Note that the FOMC only started in 1994 to announce its decisions for the federal funds rate target after each FOMC meeting.

FOMC (intermeeting) conference calls.[15] In addition, we use dates of relevant speeches and testimony before Congress by the Chair and the Vice-Chair of the Federal Reserve. Recent work by Swanson (2023) and Swanson and Jayawickrema (2023) shows that these speeches are an important source of variation in U.S. monetary policy. In contrast to Swanson and Jayawickrema (2023) who read the newspaper the following day to judge whether the speech had implications for monetary policy, we use a more data-driven approach to identify relevant speeches.[16] We use a topic modeling approach, identifying transcripts that mainly concern monetary policy decisions.[17] We aim to exclude transcripts that refer mainly to the state of the economy or regulatory changes. Thereby, we aim to exclude potential information effects (see Nakamura & Steinsson, 2018). In total, we add 81 speeches to our event dataset.[18] As a robustness test, we also estimate monetary policy shock series over a longer sample starting in 1982. Between 1982 and 1988, we identified 17 more important speeches, 48 scheduled FOMC meetings, and 19 discount rate changes that we use as event days. Our baseline specification uses all other days as control. However, we test the robustness of the results by excluding days that may systematically increase the variance of control days relative to policy event days.

2.4 Empirical results

This section discusses the results. We start by showing how our three-dimensional monetary policy shock is related to the high-frequency surprises by Swanson (2021). Then, we discuss the exchange rate responses on a daily and monthly basis. Moreover, we show how sensitive daily and monthly results are to different modeling and data specifications.

2.4.1 The monetary policy shock series

How do lower-frequency non-financial variables respond to our novel target and path shocks? To answer these questions, we estimate the monetary policy shock series, extending the approach by Bu et al. (2021) to multiple dimensions and estimate impulse responses in an IV-SVAR framework at monthly frequency.

[15] Note that these numbers differ from Swanson and Jayawickrema (2023) because we use daily data and define events on a daily basis. In contrast, they sometimes have two events on one day. We refer to their paper for more details about the selection of event dates.

[16] Unfortunately, we can not compare the speeches we identified with the speeches by Swanson and Jayawickrema (2023) because their dataset is not available to the public at the time of writing this chapter.

[17] A detailed description of how relevant articles are identified can be found in Appendix 2.F

[18] Two of them take place on the same date as a FOMC announcement.

Table 2.1 — Comparison to existing shocks

	Swanson FFR	Swanson Path	Swanson LSAP	Bu et al. (2021)
Target	1.534***	0.073	−0.382***	0.118***
	(0.285)	(0.203)	(0.127)	(0.003)
Path	0.631**	2.388***	0.342	0.178***
	(0.281)	(0.525)	(0.281)	(0.003)
Term premium	−0.982***	0.240	1.593***	-0.003
	(0.348)	(0.378)	(0.584)	(0.002)
Observations	241	241	241	215
R^2	0.352	0.331	0.287	0.986
Adjusted R^2	0.344	0.323	0.278	0.986

Notes: The table shows regressions of our estimated monetary policy shocks on the shocks provided by Swanson (2021) and Bu et al. (2021). Significance levels are given by *p<0.1; **p<0.05; ***p<0.01. HAC robust standard errors are in parentheses.

To check whether the procedure produces reasonable shock series, we compare the estimated shocks to the three-dimensional high-frequency shocks by Swanson (2021). Figure 2.H.1 in the Appendix shows that the estimated target and path shocks are substantially correlated with Swanson's high-frequency counterparts. The correlation amounts to around 0.5.[19] In Table 2.1, we regress our shocks on Swanson's high-frequency shocks as well as the Bu et al. (2021) shocks. Our shock explains about one-third of the variation in Swanson's FFR, path, and LSAP shocks. Moreover, the corresponding coefficients are highly statistically significant. Moreover, our shocks explain 99.8 % of the variation in the Bu et al. (2021) shock, with all coefficients being statistically significant. Bu et al. (2021) show that their shock is not predictable by any other information available before the release of the FOMC announcement. This suggests that our shocks are not predictable as well. These results show that even if the methodology and the data compared to Swanson (2021) are different, the shocks are substantially correlated. This suggests that our shocks are reasonable and can be

[19]For robustness, we used alternative interest rate data published by the Federal Reserve Board instead of the estimates by Gürkaynak et al. (2007). The advantage is that we know that the time stamp is 4 pm. The disadvantage is that we have fewer maturities (3M, 6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y, 20Y, 30Y). The correlation with the shocks by Swanson (2021) is even higher (see Figure 2.H.3 and Table 2.H.2 in the Appendix), while the correlation with the shocks by Bu et al. (2021), who use the same data as in our baseline, is lower. However, our results are robust to using either data source.

used to estimate the impact of monetary policy on macroeconomic variables at a lower frequency.

2.4.2 Daily effects on the exchange rate

Figure 2.1 shows the impulse responses to a target, path, and a term premium shock, respectively. The solid lines show the local projection estimates. In addition, the figure provides the F-statistics for each shock. The F-statistics for all three shocks are relatively large. Lewis (2022b) suggests, as a rule of thumb, that for $F > 23$ the bias due to weak instruments is sufficiently small. This condition is satisfied for all three dimensions. This suggests that monetary policy events comprise at least three orthogonal dimensions affecting short-, medium-, and long-term interest rates.

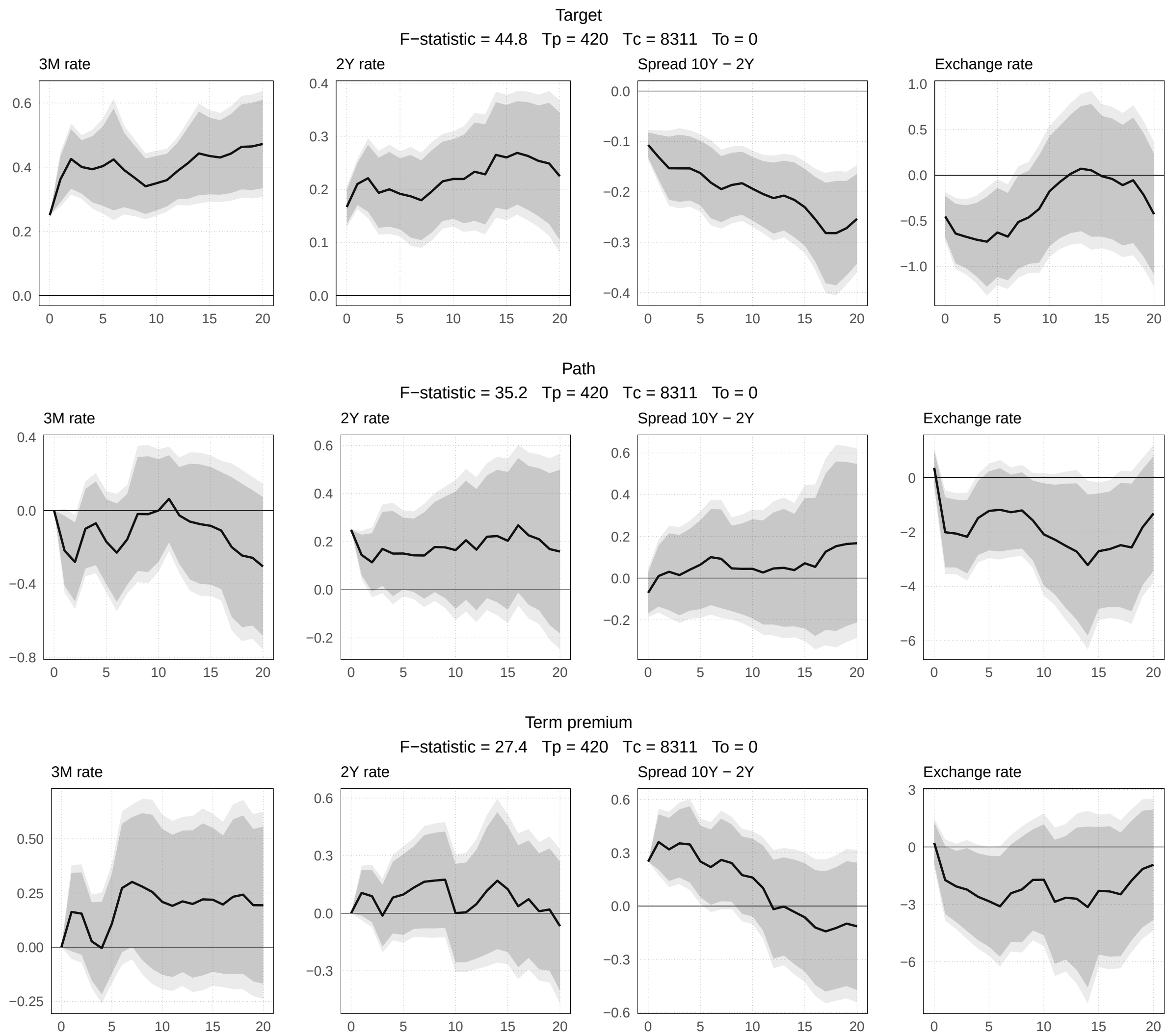
Turning to the impulse responses, a target shock tightening leads to a temporary appreciation of the USD, which vanishes after ten working days. The response is slightly delayed. However, the delay is not statistically significant. In addition, this may be partly related to the fact that exchange rates are recorded at noon, so it takes one working day until the full effect of the monetary policy shock is recorded in the data. Overall, there is no indication of a delayed overshooting puzzle.

A path shock tightening does not affect the 3M interest rate on impact due to the recursive zero restriction. As we would expect, the short-term interest rate is not significantly affected even after a few working days. However, we observe a rapid appreciation of the exchange rate. The exchange rate remains persistently stronger for the entire horizon we examine. However, the response is relatively imprecisely estimated. Finally, the term premium shock raises the term spread for up to ten working days while not significantly affecting the 3M or 2Y interest rates. The exchange rate also appreciates. Although the response is slightly delayed, it reaches the trough already after about five working days.

The delayed overshooting puzzle observed in monthly or quarterly VARs occurs at longer lags. We, therefore, estimate the exchange rate response to a target and path shock for up to 100 working days. Figure 2.2 shows that the exchange rate response is not statistically significantly different from zero at any horizon between 10 and 100 working days for the target shock.[20] Therefore, we do not find evidence in favor of delayed overshooting. However, we find a more persistent response for the path and term premium shocks. Given the large estimation uncertainty, other patterns are also

[20]The F-statistics are slightly different from the baseline because we adapt the sample to estimate the long-run responses with the same number of observations as the short-run responses.

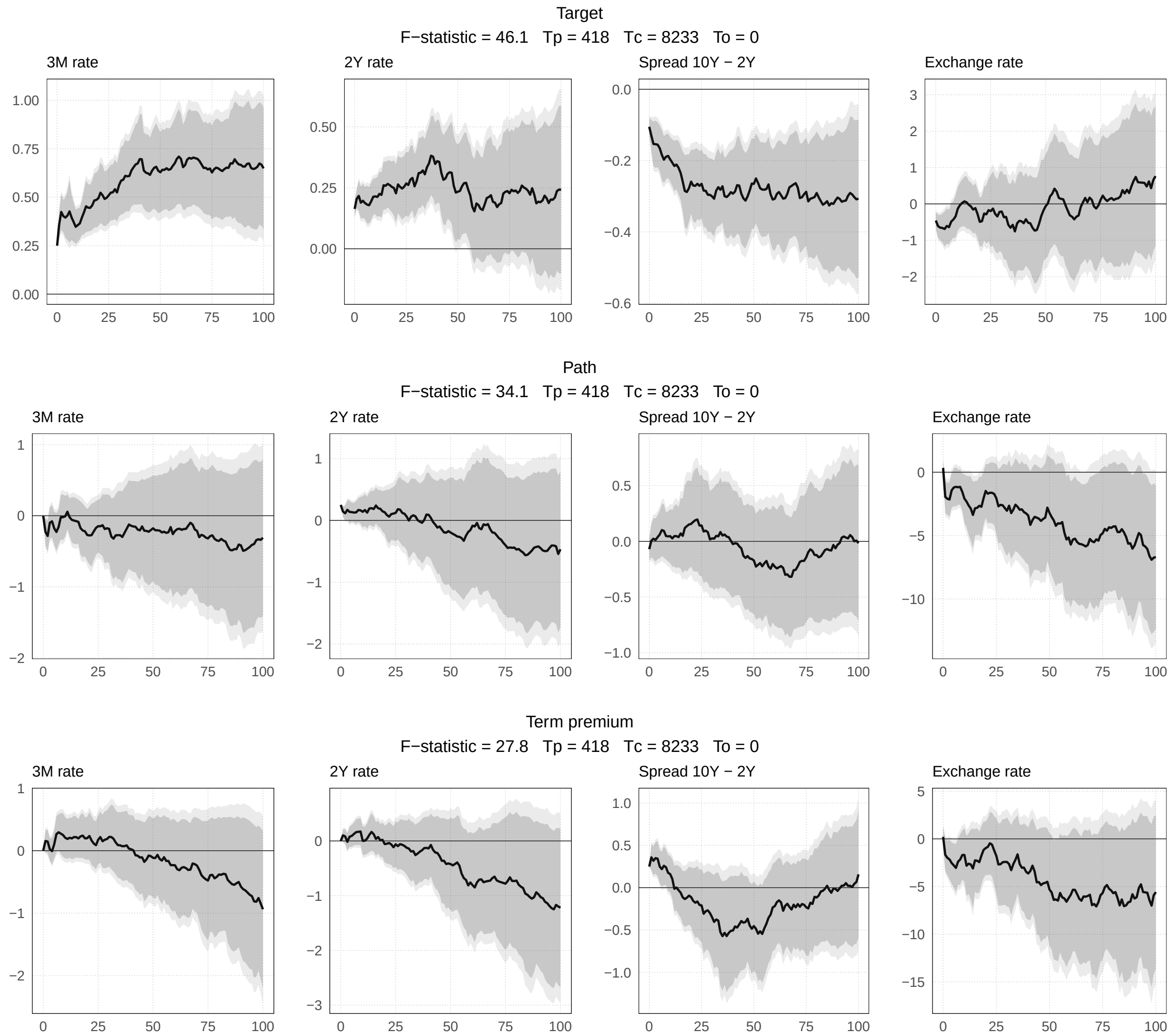
Figure 2.1 — Impulse responses to orthogonal monetary policy shocks



Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

possible, in line with Faust et al. (2003). The daily responses are not accurate enough to provide evidence in favor or against delayed overshooting for the path and term premium shocks. In the next section, we will address the question in more detail when estimating monthly impulse responses in a SVAR.

Figure 2.2 — Long-run impulse responses to orthogonal monetary policy shocks



Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

2.4.3 Monthly macroeconomic effects

Is the lack of delayed overshooting due to the daily frequency of our data? To answer this question, we estimate the responses of the nominal exchange rate and other macroeconomic variables on monthly frequency using a SVAR identified with external instruments.

Table 2.2 — First-stage regressions

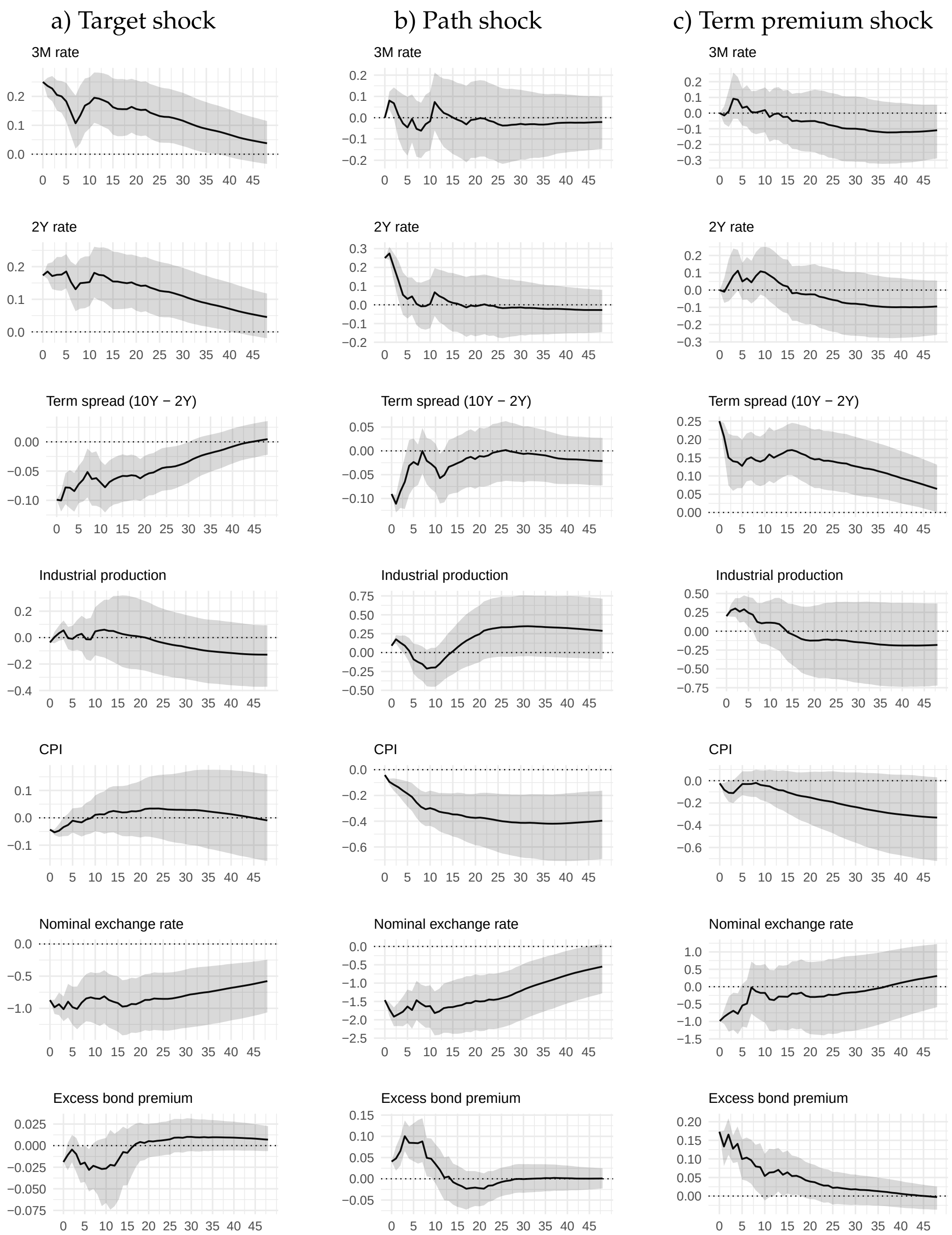
	Burri and Kaufmann			Swanson (2021)			Bu et al. (2021)
	Target	Path	Term premium	FFR	Path	LSAP	
F-stat	12.05	4.35	15.43	7.29	2.32	7.30	0.21
R^2	0.09	0.03	0.11	0.06	0.02	0.06	0
Obs.	384	384	384	336	336	336	312

Notes: The table shows the results of the first-stage regressions of the corresponding interest rate residual (3M, 2Y, 10Y - 2Y) on the three external instruments in the column header. The one-dimensional Bu et al. (2021) shock is regressed on the two-year Treasury bill residual. F-statistics above 10 indicate strong instruments.

The core assumption behind the method of external instruments is that the instruments are correlated with the structural shocks to be identified yet uncorrelated with any other structural shocks. Because the true value of the monetary policy shocks is unobserved, these conditions must be justified economically. Our shocks capture news about monetary policy transmitted via FOMC announcements and speeches by FOMC officials. Therefore, our series is expected to meet the relevance condition economically, though there is a concern that the instrument may be only weakly relevant. In this case, standard inference may fail to produce reliable results. An important statistic to check for the strength of the instrument is the F-statistic in the first-stage regression of the corresponding interest rate residual from the VAR on the instruments (Montiel Olea et al., 2021). To be confident that a weak instrument problem is not present, Montiel Olea et al. (2021) use a rule of thumb of $F > 10$. Table 2.2 shows the results of the first-stage regressions of the corresponding interest rate residual on the three external instruments. The F-statistics are above 10 for the target and term premium shocks, suggesting that the instruments are strong. However, the F-statistic is below 10 for the path shock. This suggests that the path shock is not precisely estimated. Therefore, we should interpret the results of the path shock with caution. For comparison, we also show the F-statistics for the Swanson (2021) and Bu et al. (2021) shocks. The F-statistics are well below 10, suggesting that the instruments are potentially weak.

The exogeneity condition is more difficult to justify. Recent research findings put into question that this condition holds for commonly used high-frequency monetary policy shocks by showing that they are predictable by information available to the public before the FOMC announcements (See, e.g. Bauer & Swanson, 2023a, 2023b;

Figure 2.3 — Macroeconomic effects of monetary policy



Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The identification period spans from 1988 – 2019. The horizontal axis is measured in months. All interest rate and bond premium responses are measured in percentage points. All other responses are measured in percent. 90% confidence intervals are based on a moving block bootstrap with 10,000 replications.

Miranda-Agrippino & Ricco, 2021). However, our shock series, derived from the methodology outlined in Bu et al. (2021), explains 98.6% of the variation in the shock identified by Bu et al. (2021). This shock has been demonstrated to be unpredictable by any information available prior to the FOMC announcement. Consequently, this suggests that the exogeneity condition may indeed be satisfied.

We now turn to the findings from the SVAR model, which was identified with external instruments. Figure 2.3 illustrates the impulse responses to identified three-dimensional monetary policy shocks, standardized to cause a 0.25 basis points increase in the corresponding interest rate. The solid black lines represent the point estimates, while the shaded regions indicate 90 percent confidence intervals, which are derived from 10,000 bootstrap replications.[21]

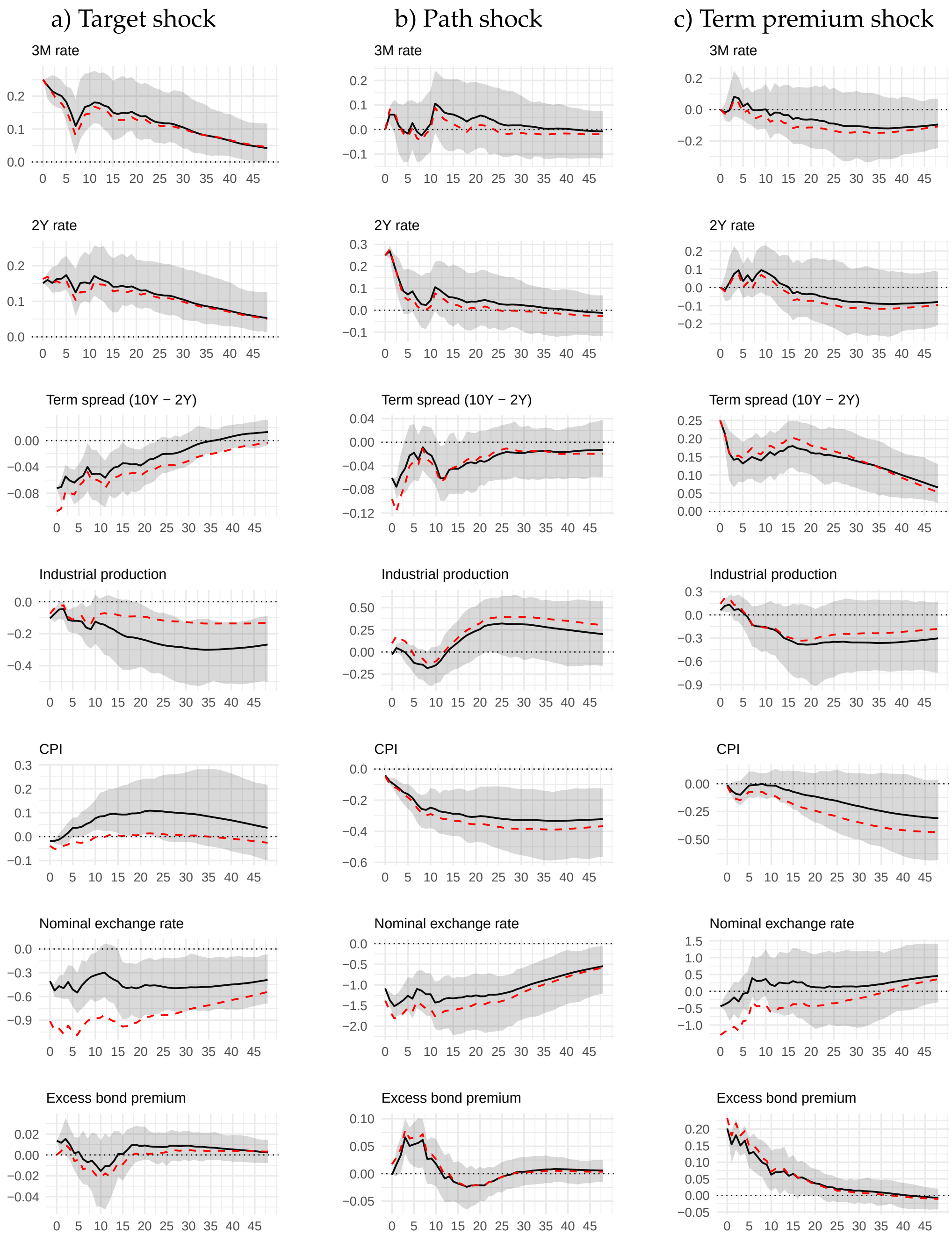
The responses to the target shock are depicted in panel a) of Figure 2.3. By construction, the three-month interest rate increases by 0.25 percentage points on impact and then gradually declines. The same holds for the two-year Treasury yield with a slightly lower increase on impact. The term spread decreases by 0.1 percentage points on impact and then gradually increases. The exchange rate significantly appreciates by about one percent on impact and then slowly depreciates. Industrial production is hardly affected. The CPI drops slightly on impact, by about 0.05 percent, then increases gradually. The response turns insignificant after about three months. The excess bond premium decreases on impact, becoming statistically insignificant as it increases again.

Panel b) of Figure 2.3 shows the responses to the path shock. By construction, the three-month interest rate is not affected on impact, and the two-year Treasury yield increases by 0.25 percentage points on impact. It then gradually decreases. The responses of the term spread, industrial production, and the exchange rate are similar to the target shock. The CPI drops slightly on impact and then declines around 0.4 percent over the following years. The excess bond premium increases around five basis points on impact, increases a further five basis points over a few months, and then declines back.

Finally, panel c) of Figure 2.3 shows the responses to the term premium shock. The three-month interest rate and the the two-year Treasury yield are not affected on impact.

[21] For calculating the confidence intervals, a moving block bootstrap technique is employed as suggested by Jentsch and Lunsford (2019, 2021). This approach yields confidence intervals that are asymptotically accurate under relatively mild α-mixing conditions. The length of each block is fixed at 24, and to address the difference between estimation and identification samples, any missing values in the instruments are set to zero (see, e.g. Känzig, 2021).

Figure 2.4 — Macroeconomic effects of monetary policy using long identification sample



Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The red dashed lines indicate our baseline responses. The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The identification period spans from 1982 – 2019. The horizontal axis is measured in months. All interest rate and bond premium responses are measured in percentage points. All other responses are measured in percent. 90% confidence intervals are based on a moving block bootstrap with 10,000 replications.

By construction, the term spread increases by 0.25 percentage points on impact and then gradually declines. Industrial production increases by 0.25 percent on impact and progressively reverts to steady-state, rendering the response insignificant after about six months. The exchange rate appreciates by about one percent on impact, then depreciates towards steady-state within a year. The CPI drops slightly on impact and then decreases further gradually. However, the response is not significantly different from zero. The excess bond premium increases on impact and then gradually decreases.

The impulse responses to these three monetary policy shocks are generally consistent with the predictions of standard macroeconomic models. We do not find evidence for a delayed overshooting of the exchange rate. Nevertheless, some estimates are not statistically significant. Moreover, the excess bond premium response to the target shock and the industrial production response to the path shock are puzzling. In a recent paper, Swanson (2024) also finds similar puzzling responses, particularly for industrial production, the CPI, and the bond spread, to what he calls FFR, forward guidance, and LSAP shocks. He argues this is due to the "Fed response to news bias" channel. This means that the Fed often tightens monetary policy when there is positive economic and financial news. This action moves economic variables in the opposite direction of what is expected from real monetary policy shocks, potentially leading to attenuation bias of the impulse responses. This channel may also explain the puzzling responses we find.

Because we are not using high-frequency data to construct our shocks, we can estimate them going back even further than 1988. We identify another 92 event days between 1982 and 1988 and use them to estimate the monetary policy shock series. Figure 2.4 shows the impulse responses using this longer identification sample. The responses are qualitatively similar to our baseline, with F-statistics of comparable magnitude. Therefore, the results are not driven by the sample period. This contrasts the finding of Kim et al. (2017), who finds that the exchange rate overshooting is an artifact of the 80s. Figure 2.H.2 in the Appendix shows that the shocks over the longer and shorter samples are highly correlated.

2.4.4 Robustness

We conducted a range of robustness tests reported in Appendix 2.G.

Robustness of daily responses We examined the response of bilateral exchange rates. For all three shocks, the USD appreciates bilaterally against a variety of other currencies (CHF, JPY, GBP, EUR, CAD). As bilateral exchange rates are more volatile, estimation uncertainty is larger. Therefore, although the point estimates sometimes deviate from the response of the trade-weighted exchange rate, the differences are not statistically different from zero.

The main results carry over when excluding major data releases. The number of observations falls for two reasons. We exclude $T_o = 1,240$ data release dates from the sample. As a consequence, the F-statistics fall slightly. Still, for the path and term premium shocks, it fulfills the rule of thumb by Lewis (2022b). In addition, the responses of the exchange rate remain qualitatively similar.

We may expect that the volatility of changes in financial market variables differs according to the day of the week. For example, the exchange rate change on Monday (compared to Friday) may be different because news is released over the weekend. We therefore restrict the control days to Tuesday and Wednesday, the days of the week where 78% of all FOMC decisions take place. The F-statistics of the path shock indeed increases. However, the responses of the exchange rate remain qualitatively unchanged.

Bauer and Swanson (2023b) suggest that speeches comprise relevant information about monetary policy. However, our results are not driven by the relevant speeches we included as policy events. The impulse responses are estimated on a sample excluding speeches ($T_o = 79$). The impulse responses and the F-statistics remain qualitatively unchanged.

We also exclude a series of periods associated with increased financial market volatility ($T_o = 938$). The number of observations falls substantially, leading to higher estimation uncertainty. However, the F-statistics are still reasonably high, and the point estimate of the exchange rate response is qualitatively similar to the baseline.

Recent research has shown that minutes released by the FOMC affect financial markets (Swanson & Jayawickrema, 2023). We exclude $T_o = 261$ such days from the sample. The results remain virtually unchanged.

We also exclude unscheduled FOMC decisions from the sample. Note that this removes $T_o = 59$ events, which occur predominantly before 1994. Before then, policy meetings were not announced in advance (Swanson & Jayawickrema, 2023). The F-statistics fall for all three shocks. In particular, the responses to the target shock are not well identified anymore, and the exchange rate response is statistically insignificant. This suggests that with better central bank communication, target surprises lost importance, but path and term premium surprises gained importance.

Other central banks may respond to US monetary policy surprises, which also affect the exchange rate. We, therefore, exclude monetary policy announcements by the ECB (starting in 1999) and the Bank of England (starting in 1997). In total, this removes $T_o = 440$ observations. The results remain virtually unchanged.

We hypothesize that the importance of various shocks varied over time. The path and term premium shocks were probably particularly relevant at the effective lower bound (ELB).[22] Indeed, when estimating the model on an effective lower bound sample, the F-statistic for the target shock drops to 0, suggesting that no target monetary policy shock exists during this episode. Meanwhile, the exchange rate responses remain qualitatively similar for the path and term premium shocks. However, as the sample is much smaller, the F-statistics are lower, and the estimation uncertainty is higher. We also estimated a specification with data starting in 1982. Note that the event periods before 1988 are defined as days with policy rate changes. However, the F-statistics fall for all three shocks, suggesting that the background noise with unannounced policy changes during this early sample period was much higher and, therefore, the instruments very weak.

Finally, we estimated specifications by removing all controls, adding more lags of the dependent variables, and adding additional controls (stock price index, commodity price index, news sentiment index, corporate bond spread). The results are hardly affected.

Robustness of monthly responses To check whether our puzzling responses on monthly frequency are driven by the "Fed response to news bias" channel, following Jarociński and Karadi (2020) and Miranda-Agrippino and Nenova (2022), we censor our monetary policy shocks to zero whenever they move in the same direction as stock prices.[23] These periods may be affected by either a "Fed response to news bias" or an information effect (Nakamura & Steinsson, 2018). For example, suppose the Federal Reserve provides a more optimistic view of the state of the economy. In that case, stock prices may rise while markets are surprised by the monetary policy tightening. Figure 2.G.14 in the Appendix shows that the puzzles are eliminated when censoring the shocks. The response of the excess bond premium to the target shock is now positive, and the response of industrial production to the path shock is now negative and statistically significant. The other responses remain qualitatively similar, although somewhat less

[22] We define the ELB period from December 16, 2008, to December 16, 2015, as well as from March 16, 2020, to March 17, 2022.

[23] This procedure is also known as the poor man's sign restrictions.

attenuated. This suggests that the "Fed response to news bias" might be present, even if Bu et al. (2021) find that their method estimates shocks that are not predictable.

Another robustness check addresses the concern that the target shock is quite volatile during the ELB period when, by definition, no such shocks should exist. We, therefore, censored the target shock to zero whenever the federal funds rate was at the effective lower bound. The results in Figure 2.G.15 in the Appendix are almost identical to the baseline.

Is there evidence for the delayed overshooting puzzle using existing monetary policy shocks? To address this question, we estimate the impulse responses to the Swanson (2021) and Bu et al. (2021) shocks. The results are shown in Figure 2.G.16 and 2.G.17 in the Appendix. The path shock of Swanson (2021) generates a puzzling increase in industrial production. The LSAP shock generates puzzling responses regarding the sign of the CPI, the exchange rate, and the excess bond premium. However, these responses should be taken with a grain of salt, as the first-stage F-Statistics are well below 10. The Bu et al. (2021) shock generates responses in line with macroeconomic models. However, the first-stage F-statistic is very low, suggesting that the instrument is weak. Thus, even with existing monetary policy shocks, there is no evidence for a well-identified delayed overshooting puzzle.

Finally, we identify the three shocks using our shocks and the Swanson (2021) shocks as instruments. Thus the model is overidentified. These different types of shocks might complement each other and provide more precise estimates. However, the first-stage F-statistics are below ten and lie somewhere in between our baseline and the Swanson (2021) F-statistics (See Table 2.2). The results are shown in Figure 2.G.18 in the Appendix. The responses constitute a mix of the responses to the individual shocks, being closer to our baseline responses.

2.5 Concluding remarks

In this paper, we propose to combine a heteroscedasticity-based identification scheme with recursive zero restrictions to identify multiple orthogonal monetary policy shocks along the term structure of interest rates. We then show how to estimate daily dynamic causal effects by modifying the IV approach by Rigobon and Sack (2004).

Applying this identification scheme, we contribute to the ongoing debate in the literature about the effects and the timing of monetary policy shocks on the exchange rate.

So far, there is little evidence in the literature on whether different monetary policy actions, such as changes in the interest rate target, forward guidance, or large-scale asset purchases, affect the exchange rate differently.

The results show that all three dimensions have the expected effect on the exchange rate. A monetary policy tightening appreciates the exchange rate immediately. There is no evidence to support the presence of delayed exchange rate overshooting. This finding holds true for all three identified monetary policy dimensions. Moreover, we do not find evidence that different policy actions affect the exchange rate in different directions. Differences appear mainly in the persistence of the exchange rate effects. However, the responses on horizons that are longer than two weeks suffer from large estimation uncertainty.

To estimate the effects on lower frequency and on other macroeconomic variables, we extend the methodology developed by Bu et al. (2021) to estimate multiple dimensions of monetary policy shocks. Using the shocks in an SVAR identified via external instruments, we find persistent responses of the exchange rate. However, we do not find evidence for an exchange rate overshooting or that different policy actions affect the exchange rate in different directions. This holds when we estimate the model with data including the 1980s and extend our shock series back to 1982. Therefore, this suggests that the delayed overshooting puzzle is an artifact of the identification scheme rather than the estimation period (Kim et al., 2017). This also has implications for the calibration of theoretical models of exchange rate dynamics based on portfolio-adjustment costs (see e.g. Bacchetta & Van Wincoop, 2021).

Further, we find that our shocks are substantially correlated with the multi-dimensional high-frequency surprises by Swanson (2021), even though we use a different method based on daily data. This suggests that our procedure is a valid alternative for countries and periods where high-frequency data is missing or the exact time-stamp of monetary policy announcements is unknown.

Because this approach is not reliant on high-frequency data and the exact timing of events, a promising avenue for future research could be to apply it to other fields where identifying causal relationships is challenging. For example, similar to Bianchi et al. (2024), it could be used to estimate the effects of news shocks on financial markets. Moreover, it could be used to estimate the effects of monetary policy in a historical context, provided that daily data exist.

References

Altavilla, C., Brugnolini, L., Gürkaynak, R. S., Motto, R., & Ragusa, G. (2019). Measuring euro area monetary policy. *Journal of Monetary Economics*, *108*, 162–179. https://doi.org/10.1016/j.jmoneco.2019.08.016

Bacchetta, P., & Van Wincoop, E. (2021). Puzzling exchange rate dynamics and delayed portfolio adjustment. *Journal of International Economics*, *131*, 103460. https://doi.org/10.1016/j.jinteco.2021.103460

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Bauer, M. D., & Swanson, E. T. (2023a). An alternative explanation for the "fed information effect". *American Economic Review*, *113*(3), 664–700. https://doi.org/10.1257/aer.20201220

Bauer, M. D., & Swanson, E. T. (2023b). A reassessment of monetary policy surprises and high-frequency identification. *NBER Macroeconomics Annual*, *37*, 87–155. https://doi.org/10.1086/723574

Benigno, G. (2004). Real exchange rate persistence and monetary policy rules. *Journal of Monetary Economics*, *51*(3), 473–502. https://doi.org/10.1016/j.jmoneco.2003.06.001

Bianchi, F., Ludvigson, S., & Ma, S. (2024). What hundreds of economic news events say about belief overreaction in the stock market. *NBER Working Paper No. 32301*. https://doi.org/10.3386/w32301

Bjørnland, H. C. (2009). Monetary policy and exchange rate overshooting: Dornbusch was right after all. *Journal of International Economics*, *79*(1), 64–77. https://doi.org/10.1016/j.jinteco.2009.06.003

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, *1*(1), 17–35. https://doi.org/10.1214/07-AOAS114

Braun, R., Miranda-Agrippino, S., & Saha, S. (2023). Measuring monetary policy in the UK: The UK monetary policy event-study database. *Bank of England Working Paper No. 1050*. https://doi.org/10.2139/ssrn.4698305

Brennan, C. M., Jacobson, M. M., Matthes, C., & Walker, T. B. (2024). Monetary policy shocks: Data or methods? *Finance and Economics Discussion Series 2024-011*, (2024). https://doi.org/https://doi.org/10.17016/FEDS.2024.011

Bu, C., Rogers, J., & Wu, W. (2021). A unified measure of fed monetary policy shocks. *Journal of Monetary Economics*, *118*, 331–349. https://doi.org/10.1016/j.jmoneco.2020.11.002

Canetg, F., & Kaufmann, D. (2022). Overnight rate and signalling effects of central bank bills. *European Economic Review*, *143*, 104060. https://doi.org/10.1016/j.euroecorev.2022.104060

Ciminelli, G., Rogers, J., & Wu, W. (2022). The effects of u.s. monetary policy on international mutual fund investment. *Journal of International Money and Finance*, *127*, 102676. https://doi.org/10.1016/j.jimonfin.2022.102676

Dornbusch, R. (1976). Expectations and exchange rate dynamics. *Journal of Political Economy*, *84*(6), 1161–1176. https://doi.org/10.1086/260506

Eichenbaum, M., & Evans, C. L. (1995). Some empirical evidence on the effects of shocks to monetary policy on exchange rates. *The Quarterly Journal of Economics*, *110*(4), 975–1009. https://doi.org/10.2307/2946646

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636. https://doi.org/10.1086/260061

Faust, J., & Rogers, J. H. (2003). Monetary policy's role in exchange rate behavior. *Journal of Monetary Economics*, *50*(7), 1403–1424. https://doi.org/10.1016/j.jmoneco.2003.08.003

Faust, J., Rogers, J. H., Swanson, E., & Wright, J. H. (2003). Identifying the effects of monetary policy shocks on exchange rates using high frequency data. *Journal of the European Economic Association*, *Vol. 1*, pp. 1031–1057.

Gertler, M., & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, *7*(1), 44–76. https://doi.org/10.1257/mac.20130329

Gilchrist, S., & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, *102*(4), 1692–1720. https://doi.org/10.1257/aer.102.4.1692

Gürkaynak, R. S., Sack, B., & Swanson, E. T. (2005). Do actions speak louder than words? the response of asset prices to monetary policy actions and statements. *International Journal of Central Banking*, *1*(1), 39.

Gürkaynak, R. S., Sack, B., & Wright, J. H. (2007). The u.s. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, *54*(8), 2291–2304. https://doi.org/10.1016/j.jmoneco.2007.06.029

Jarociński, M., & Karadi, P. (2020). Deconstructing monetary policy surprises— the role of information shocks. *American Economic Journal: Macroeconomics*, *12*(2), 1–43. https://doi.org/10.1257/mac.20180090

Jentsch, C., & Lunsford, K. G. (2019). The dynamic effects of personal and corporate income tax changes in the united states: Comment. *American Economic Review*, *109*(7), 2655–2678. https://doi.org/10.1257/aer.20162011

Jentsch, C., & Lunsford, K. G. (2021). Asymptotically valid bootstrap inference for proxy SVARs. *Journal of Business & Economic Statistics*, 1–16. https://doi.org/10.1080/07350015.2021.1990770

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, *95*(1), 161–182. https://doi.org/10.1257/0002828053828518

Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review*, *111*(4), 1092–1125. https://doi.org/10.1257/aer.20190964

Kearn, J., & Manners, P. (2006). The impact of monetary policy on the exchange rate: A study using intraday data. *International Journal of Central Banking*, *2*(4).

Kim, S.-H., Moon, S., & Velasco, C. (2017). Delayed overshooting: Is it an '80s puzzle? *Journal of Political Economy*, *125*(5), 1570–1598. https://doi.org/10.1086/693372

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of Monetary Economics*, *47*(3), 523–544. https://doi.org/10.1016/S0304-3932(01)00055-1

Lakdawala, A. (2019). Decomposing the effects of monetary policy using an external instruments SVAR. *Journal of Applied Econometrics*, *34*(6), 934–950. https://doi.org/10.1002/jae.2721

Lewis, D. J. (2022a). Announcement-specific decompositions of unconventional monetary policy shocks & their macroeconomic effects. *Staff Report, No. 891, Federal Reserve Bank of New York*.

Lewis, D. J. (2022b). Robust inference in models identified via heteroskedasticity. *The Review of Economics and Statistics*, *104*(3), 510–524. https://doi.org/10.1162/rest_a_00963

Mertens, K., & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review*, *103*(4), 1212–1247. https://doi.org/10.1257/aer.103.4.1212

Miranda-Agrippino, S., & Nenova, T. (2022). A tale of two global monetary policies. *Journal of International Economics*, *136*, 103606. https://doi.org/10.1016/j.jinteco.2022.103606

Miranda-Agrippino, S., & Rey, H. (2022). The global financial cycle. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), *Handbook of international economics: International macroeconomics, volume 6* (pp. 1–43). Elsevier. https://doi.org/https://doi.org/10.1016/bs.hesint.2022.02.008

Miranda-Agrippino, S., & Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, *13*(3), 74–107. https://doi.org/10.1257/mac.20180124

Montiel Olea, J. L., Stock, J. H., & Watson, M. W. (2021). Inference in structural vector autoregressions identified with an external instrument. *Journal of Econometrics*, *225*(1), 74–87. https://doi.org/10.1016/j.jeconom.2020.05.014

Müller, G. J., Wolf, M., & Hettig, T. (2023). Delayed overshooting: The case for information rigidities. *American Economic Journal: Macroeconomics*, *forthcoming*.

Nakamura, E., & Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: The information effect*. *The Quarterly Journal of Economics*, *133*(3), 1283–1330. https://doi.org/10.1093/qje/qjy004

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703. https://doi.org/10.2307/1913610

Ramey, V. (2016). Macroeconomic shocks and their propagation. *Handbook of macroeconomics* (pp. 71–162). Elsevier. https://doi.org/10.1016/bs.hesmac.2016.03.003

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, *85*(4), 777–792. https://doi.org/10.1162/003465303772815727

Rigobon, R., & Sack, B. (2004). The impact of monetary policy on asset prices. *Journal of Monetary Economics*, *51*(8), 1553–1575. https://doi.org/10.1016/j.jmoneco.2004.02.004

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An r package for structural topic models. *Journal of Statistical Software*, *91*(2). https://doi.org/10.18637/jss.v091.i02

Rogoff, K. (2002). Dornbusch's overshooting model after twenty-five years: International monetary fund's second annual research conference mundell-fleming lecture. *IMF Staff Papers*, *49*, 1–34. https://doi.org/10.2307/3872469

Schlaak, T., Rieth, M., & Podstawski, M. (2023). Monetary policy, external instruments, and heteroskedasticity. *Quantitative Economics*, *14*(1), 161–200. https://doi.org/10.3982/QE1511

Schmitt-Grohé, S., & Uribe, M. (2022). The effects of permanent monetary shocks on exchange rates and uncovered interest rate differentials. *Journal of International Economics*, *135*, 103560. https://doi.org/10.1016/j.jinteco.2021.103560

Scholl, A., & Uhlig, H. (2008). New evidence on the puzzles: Results from agnostic identification on monetary policy and exchange rates. *Journal of International Economics*, *76*(1), 1–13. https://doi.org/10.1016/j.jinteco.2008.02.005

Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, *128*(610), 917–948. https://doi.org/10.1111/ecoj.12593

Swanson, E. T. (2021). Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. *Journal of Monetary Economics*, *118*, 32–53. https://doi.org/10.1016/j.jmoneco.2020.09.003

Swanson, E. T. (2023). The importance of fed chair speeches as a monetary policy tool. *mimeo*.

Swanson, E. T. (2024). The macroeconomic effects of the federal reserve's conventional and unconventional monetary policies. *mimeo*.

Swanson, E. T., & Jayawickrema, V. (2023). Speeches by the fed chair are more important than FOMC announcements: An improved high-frequency measure of u.s. monetary policy shocks. *mimeo*.

Wright, J. H. (2012). What does monetary policy do to long-term interest rates at the zero lower bound? *The Economic Journal*, *122*(564), F447–F466. https://doi.org/10.1111/j.1468-0297.2012.02556.x

Appendix 2.A Extension daily model

We may extend this framework to include observed control variables, such as lags of the dependent variables, in the information set. Let the model comprise M control variables:

$$\begin{aligned} y_t &= \sum_{l=1}^{L} \Phi_l x_{t-l} + \Psi \varepsilon_t + \Gamma v_t \quad \text{for } t \in P \\ y_t &= \sum_{l=1}^{L} \Phi_l x_{t-l} + \Gamma v_t \quad \text{for } t \in C \end{aligned} \tag{2.26}$$

where Φ_l are $N \times M$ matrices of coefficients for every lag $l = 1, \ldots, L$. Then, we add x_{t-l}, for $l = 1, \ldots, L$ as additional regressors in the IV estimation. The construction of the instrument remains unchanged.[24]

Following Jordà (2005), we can also estimate dynamic effects by iterating the dependent variable forward:

$$\begin{aligned} y_{t+h} &= \sum_{l=1}^{L} \Phi_l^{(h)} x_{t-l} + \sum_{n=0}^{h} \Psi^{(h-n)} \varepsilon_{t+n} + \Gamma^{(h-n)} v_{t+n} \quad \text{for } t \in P \\ y_{t+h} &= \sum_{l=1}^{L} \Phi_l^{(h)} x_{t-l} + \sum_{n=0}^{h} \Gamma^{(h-n)} v_{t+n} \quad \text{for } t \in C \end{aligned} \tag{2.27}$$

where $\Psi^{(h)}$ and $\Gamma^{(h)}$ are the impulse response functions after h periods and $\Phi_l^{(h)}$ are coefficients on the control variables, which differ for every horizon h.[25]

The error term in the IV estimation includes future monetary policy and other shocks. As the instruments are only affected by current shocks, not by future shocks, the exclusion restriction is still valid.[26] We can therefore use the same first-stage and then replace the dependent variable in the second stage with y_{t+h} to estimate the impulse response after h periods. For the same reason, we can estimate cumulative responses by using $\sum_{n=0}^{h} y_{t+h}$ as the dependent variable with the same instrument.

Appendix 2.B Simulation study

We conducted a simulation study to show that we can recover the impulse responses of a multi-dimensional monetary policy shock with our IV estimator. We simulate 5'000 observations for $N = 3$ variables using VAR with $P = 2$ lags. There are $R = 3$ i.i.d.

[24] If $x_{t-l} = y_{t-l}$ and $L = 1$ we have the structure of a VAR of order 1.

[25] For example, if the data generating process is a VAR(1) we have that $y_{t+h} = \Phi^h y_{t-1} + \sum_{n=0}^{h} \Phi^{h-n} \Psi \varepsilon_{t+n} + \Phi^{h-n} \Gamma v_{t+n}$, with $\Psi^{(h)} \equiv \Phi^h \Psi$ and $\Gamma^{(h)} \equiv \Phi^h \Gamma$.

[26] However, the error term is autocorrelated, such that it is important to use a HAC-consistent variance estimator (see Newey & West, 1987)

structural shocks that occur on all periods and $E = 2$ i.i.d. structural shocks that occur every 3rd period only. Both impact matrices are lower triangular. The specific values of the impact matrices and VAR coefficients are drawn randomly subject to the constraint that the VAR is stationary.

Figure 2.B.1 shows that we can estimate the impulse responses of the three variables to the two shocks that occur only every 3rd period using the recursive heteroskedasticity-based IV estimator. The impact effect is accurately estimated, while there are larger deviations and wider confidence intervals at longer horizons. Looking at the estimated responses for the third dimension, which does actually not exist, we see that the confidence intervals are wide and include zero. In addition, the F-statistic is lower than 23, the rule-of-thumb by Lewis (2022b).

Appendix 2.C Estimation of shocks via OLS

According to Bu et al. (2021) we can estimate the underlying shocks via an OLS regression of the cross-sectional variation of the interest rates across the term structure and the impact matrix. Let our estimated $\tilde{\Psi}$ be proportional to the true impact matrix: $\tilde{\Psi}A = \Psi$, where A is a diagonal matrix. The OLS estimator for a given time period t reads:

$$\hat{\varepsilon}_t = (\tilde{\Psi}'\tilde{\Psi})^{-1}\tilde{\Psi}'y_t. \tag{2.28}$$

On policy event days, the interest rates are affected by monetary policy shocks and other shocks: $y_t = \Psi\varepsilon_t + \Gamma v_t$. Therefore

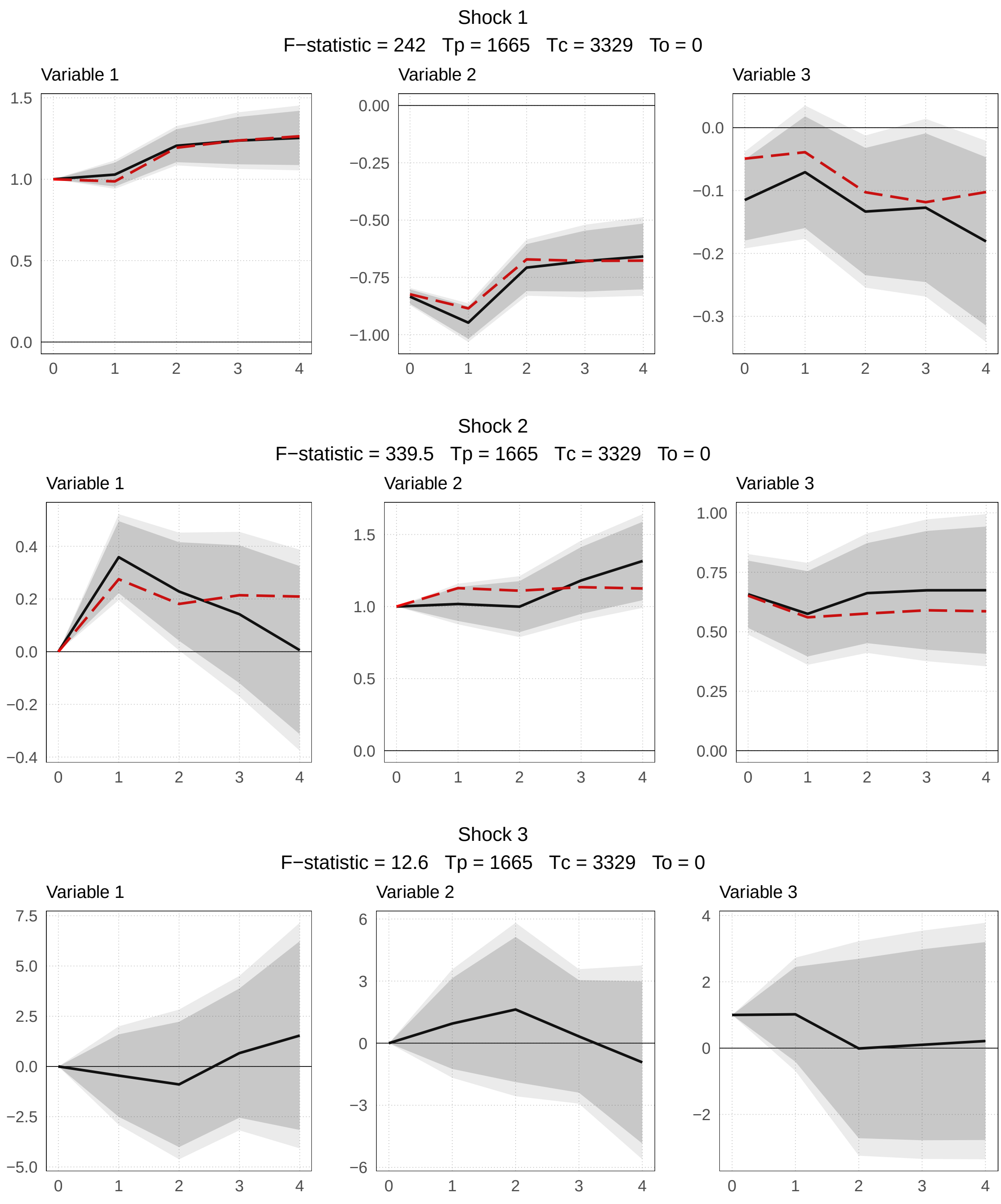
$$\hat{\varepsilon}_t = (\tilde{\Psi}'\tilde{\Psi})^{-1}\tilde{\Psi}'\Psi\varepsilon_t + (\tilde{\Psi}'\tilde{\Psi})^{-1}\tilde{\Psi}'\Gamma v_t. \tag{2.29}$$

Using the fact that $\tilde{\Psi}$ is proportional to Ψ yields

$$\hat{\varepsilon}_t = A\varepsilon_t + (\tilde{\Psi}'\tilde{\Psi})^{-1}\tilde{\Psi}'\Gamma v_t. \tag{2.30}$$

Note that if $(\tilde{\Psi}'\tilde{\Psi})^{-1}\tilde{\Psi}'\Gamma$ is different from zero, we will introduce background noise

Figure 2.B.1 — Impulse responses simulated data $E = 2$



Notes: Cumulative impulse responses two-dimensional shock occurring every third period estimated using local projections. The red dashed lines give the true impulse responses. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

because the other structural shocks v_t will lead to time-variation in the estimated monetary policy shocks. Therefore, we need an orthogonality assumption. Let us assume that:

$$\mathbb{E}[\Gamma v_t|\tilde{\Psi}] = 0 \tag{2.31}$$

which corresponds to the classical OLS assumption that the error term is uncorrelated with the regressors. Using the law of total expectation we then obtain:

$$\mathbb{E}[\hat{\varepsilon}_t] = A\varepsilon_t + \mathbb{E}\left[\left(\tilde{\Psi}'\tilde{\Psi}\right)^{-1}\tilde{\Psi}'\mathbb{E}[\Gamma v_t|\tilde{\Psi}]\right] = A\varepsilon_t. \tag{2.32}$$

Therefore, OLS yields an unbiased estimate of the underlying shocks (up to a scale) if the sensitivity of the interest rates across the term structure to other structural shocks is unrelated to the sensitivity across the term structure to the monetary policy shocks ($\mathbb{E}[\Gamma v_t|\tilde{\Psi}] = 0$).

Appendix 2.D External instrument identification with k shocks and k instruments

In this Appendix, we provide more details on the joint identification strategy of the $k = 3$ monetary policy shocks in a SVAR with k instruments (See also Känzig, 2021; Lakdawala, 2019; Mertens & Ravn, 2013, for other applications). Specifically, we allow the instruments to be correlated not only with one structural shock but impose additional restrictions on their impact effects.

We start by organizing the structural shocks as $\varepsilon_t = \begin{pmatrix}\varepsilon'_{1,t} & \varepsilon'_{2,t}\end{pmatrix}'$, where $\varepsilon_{1,t}$ represents a $k \times 1$ vector of the structural shocks we aim to identify, and $\varepsilon_{2,t}$ denotes a $(n-k)\times 1$ vector encompassing the remaining structural shocks. Similarly, we express u_t as $\begin{pmatrix}u'_{1,t} & u'_{2,t}\end{pmatrix}'$. The identification of these shocks relies on moment restrictions for the instrument as follows:

$$\mathbb{E}\left[z_t\varepsilon'_{1,t}\right] = \Phi_{k\times k} \tag{2.33}$$
$$\mathbb{E}\left[z_t\varepsilon'_{2,t}\right] = 0_{k\times(n-k)} \tag{2.34}$$

where Φ is of full rank. The covariance restrictions are expressed by the equation:

$$SS' = \Omega. \tag{2.35}$$

In the next step, we partition S as

$$S^{n\times n} = \left(S_1^{n\times k}, S_2^{n\times(n-k)}\right) = \begin{pmatrix} S_{11}^{k\times k} & S_{12}^{k\times(n-k)} \\ S_{21}^{(n-k)\times k} & S_{22}^{(n-k)\times(n-k)} \end{pmatrix}. \tag{2.36}$$

The instrument moment conditions together with $u_t = S\varepsilon_t$ imply

$$\Sigma_{zu'} = \mathbb{E}\left[z_t u_t'\right] = \mathbb{E}\left[z_t \varepsilon_t'\right] S' = \mathbb{E}\left[z_t \begin{pmatrix}\varepsilon_{1,t}' & \varepsilon_{2,t}'\end{pmatrix}\right] \begin{pmatrix} S_1' \\ S_2' \end{pmatrix} = (\Phi, 0) \begin{pmatrix} S_1' \\ S_2' \end{pmatrix} = \Phi S_1'. \tag{2.37}$$

Together with the variances of u_t and ε_t, this yields

$$\mathbb{E}\left[z_t u_t'\right] \mathbb{E}\left[u_t u_t'\right]^{-1} \mathbb{E}\left[u_t z_t'\right] = \Phi S_1' \left(SS'\right)^{-1} S_1 \Phi' = \Phi S_1'(S')S^{-1}S_1\Phi' = \Phi\Phi' \tag{2.38}$$

where $S^{-1}S_1 = \begin{pmatrix} I_k & 0_{(n-k)\times k}\end{pmatrix}'$. If $k = 1$, Φ is a scalar and the identification is unique up to sign and scale. If $k > 1$, Φ has k^2 unique elements, while $\Phi\Phi'$ is symmetric with only $\frac{k(k+1)}{2}$ unique elements. Hence, S_1 is only identified up to a rotation.

Another way to show this is by partitioning $\Sigma_{zu'} = \begin{pmatrix}\Sigma_{zu_1'} & \Sigma_{zu_2'}\end{pmatrix}$ or equivalently

$$\Phi S_{11}' = \Sigma_{zu_1'} \tag{2.39}$$

$$\Phi S_{21}' = \Sigma_{zu_2'}. \tag{2.40}$$

Combining the two yields

$$S_{21}S_{11}^{-1} = \left(\Sigma_{zu_1'}^{-1}\Sigma_{zu_2'}\right)'. \tag{2.41}$$

This corresponds to the 2SLS estimator in a regression of $u_{2,t}$ on $u_{1,t}$ using z_t as an instrument for $u_{1,t}$.

The covariance restrictions yield

$$SS' = \Omega \tag{2.42}$$
$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} S'_{11} & S'_{21} \\ S'_{12} & S'_{22} \end{pmatrix} = \begin{pmatrix} S_{11}S'_{11} + S_{12}S'_{12} & S_{11}S'_{21} + S_{12}S'_{22} \\ S_{21}S'_{11} + S_{22}S'_{12} & S_{21}S'_{21} + S_{22}S'_{22} \end{pmatrix} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}.$$

Since Ω is a covariance matrix, it is symmetric, i.e., $\Omega'_{12} = \Omega_{21}$. Thus, this system yields three equations:

$$S_{11}S'_{11} + S_{12}S'_{12} = \Omega_{11} \tag{2.43}$$
$$S_{11}S'_{21} + S_{12}S'_{22} = \Omega_{12} \tag{2.44}$$
$$S_{21}S'_{21} + S_{22}S'_{22} = \Omega_{22}. \tag{2.45}$$

To identify S up to a rotation, it is sufficient to find $S_{11}S'_{11}$, $S_{22}S'_{22}$, $S_{21}S_{11}^{-1}$ and $S_{12}S_{22}^{-1}$. This is because one can write

$$S = \begin{pmatrix} L_1 & S_{12}S_{22}^{-1}L_2 \\ S_{21}S_{11}^{-1}L_1 & L_2 \end{pmatrix} \tag{2.46}$$

where $L_1 = \text{chol}(S_{11}S'_{11})$ and $L_2 = \text{chol}(S_{22}S'_{22})$. This still satisfies $SS' = \Omega$. Thus, it proves useful to rewrite these equations in terms of $S_{11}S'_{11}$, $S_{22}S'_{22}$, $S_{21}S_{11}^{-1}$ and $S_{12}S_{22}^{-1}$:

$$S_{11}S'_{11} + S_{12}S_{22}^{-1}S_{22}S'_{22}(S'_{22})^{-1}S'_{12} = \Omega_{11} \tag{2.47}$$
$$S_{11}S'_{11}S_{11}^{-1}S'_{21} + S_{12}S_{22}^{-1}S_{12}S'_{22} = \Omega_{12} \tag{2.48}$$
$$S_{21}S_{11}^{-1}S_{21}S'_{21}S_{11}^{-1}S'_{21} + S_{22}S'_{22} = \Omega_{22}. \tag{2.49}$$

Note that $S_{21}S_{11}^{-1}$ is identified by the instrument conditions. Thus, this is a system of 3 matrix equations in 3 unknown matrices. The solutions are given by

$$S_{12}S_{12}' = \left(\Omega_{21} - S_{21}S_{11}^{-1}\Omega_{11}\right)' \zeta^{-1} \left(\Omega_{21} - S_{21}S_{11}^{-1}\Omega_{11}\right) \tag{2.50}$$

$$\zeta = \left(\Omega_{22} + S_{21}S_{11}^{-1}\Omega_{11}(S_{11}')^{-1}S_{21}' - S_{21}S_{11}^{-1}\Omega_{12} - \Omega_{21}(S_{11}')^{-1}S_{21}'\right) \tag{2.51}$$

$$S_{11}S_{11}' = \Omega_{11} - S_{12}S_{12}' \tag{2.52}$$

$$S_{22}S_{22}' = \Omega_{22} - S_{21}S_{11}^{-1}S_{11}S_{11}'(S_{11}')^{-1}S_{21}' \tag{2.53}$$

$$S_{12}S_{22}^{-1} = \left(\Omega_{12} - S_{11}S_{11}'(S_{11}')^{-1}S_{21}'\right)\left(S_{22}S_{22}'\right)^{-1} \tag{2.54}$$

which is sufficient to evaluate S.

However, this only identifies S up to a rotation. The parameter space can be characterized by

$$SR = \begin{pmatrix} L_1 & S_{12}S_{22}^{-1}L_2 \\ S_{21}S_{11}^{-1}L_1 & L_2 \end{pmatrix} \begin{pmatrix} R_k & 0 \\ 0 & R_{n-k} \end{pmatrix} = \begin{pmatrix} L_1R_k & S_{12}S_{22}^{-1}L_2R_{n-k} \\ S_{21}S_{11}^{-1}L_1R_k & L_2R_{n-k} \end{pmatrix} \tag{2.55}$$

where R is an orthonormal rotation matrix. Our focus is on pinpointing the first k shocks, which entails selecting an appropriate R_k rotation submatrix for S_1's identification. Setting $R_k = I_k$ is deemed a suitable choice for the scenario under consideration. This presupposes that the VAR orders the three-month interest rate first, followed by the two-year Treasury yield, and then the ten-year Treasury yield. Given that L_1 is a lower triangular matrix, it implies the assumption that initially, the path and term premium shocks have no immediate effect on the three-month interest rate. Furthermore, it assumes that the term premium shock does not immediately impact the two-year Treasury yield. These additional assumptions identify the three structural shocks.

Appendix 2.E Data

Table 2.E.1 — Time series

Category	**Source**	**Variants**	**Time stamp**	**Comments**
Treasury bill yields	Board of Governors	3M, 6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y, 20Y, 30Y	4pm EST	www.federalreserve.gov/releases/h15/
Treasury bill yields	Gürkaynak et al. (2007)	1Y to 30Y		https://www.federalreserve.gov/data/nominal-yield-curve.htm
Federal Funds Rate	Board of Governors		Close	www.federalreserve.gov/releases/h15/
Exchange rates	Board of Governors	NEER, USD/CHF, USD/JPY, USD/GBP, USD/EUR	Noon EST	www.federalreserve.gov/releases/h10/. For the nominal effective exchange rate, we linked the discountinued series with FRED identifier DTWEXM with DTWEXAFEGS. For the USD/EUR exchange rate, we linked the USD/DEM with the USD/EUR exchange rate using the official changeover exchange rate from www.eu-info.de/euro-waehrungsunion/5007/5222/5170/.
Stock prices	TradingView	S&P 500	4pm EST	de.tradingview.com/symbols/SPX/
Bond spreads	Moody's	AAA, BAA		fred.stlouisfed.org/series/DAAA fred.stlouisfed.org/series/DBAA. We computed the spreads as the difference to the 10Y government bond yield.
Economic Policy Uncertainty	Baker et al. (2016)			fred.stlouisfed.org/series/USEPUINDXD
Commodity price index	Dow Jones, Bloomberg			
Industrial production	FRED		Monthly	FRED variable key: INDPRO
CPI	FRED		Monthly	FRED variable key: CPIAUCSL
Excess bond premium	Gilchrist and Zakrajšek (2012)		Monthly	

Table 2.E.2 — Events

Category	Source	Comments
FOMC announcements	1982 to 1987: https://www.federalreserve.gov/monetarypolicy/fomc_historical_year.htm, 1988 to 2019: Swanson and Jayawickrema (2023), from 2020: https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm	
Discount rate changes	Monetary Policy and Open Market Operations 1982 - 1989, FRBNY Quarterly Review, 1983 - 1990, https://www.newyorkfed.org/research/quarterly_review/75th.html	Collected for 1982 - 1989
Speeches and Congressional Testimony	To 1996: https://alfred.stlouisfed.org/, from 1997: https://www.federalreserve.gov/newsevents/speeches.htm	
FOMC minutes	https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm	Available from 1988
ECB decisions	Altavilla et al. (2019)	Available from 1999
BoE decisions	Braun et al. (2023)	Available from 1997
CPI releases	https://www.bls.gov/bls/news-release/cpi.htm, https://alfred.stlouisfed.org/	
PPI releases	https://www.bls.gov/bls/news-release/ppi.htm	Available from 1994
Employment situation releases	https://www.bls.gov/bls/news-release/empsit.htm	Available from 1994
Employment cost releases	https://www.bls.gov/bls/news-release/eci.htm	
GDP releases	https://www.bea.gov/index.php/news/archive?field_related_product_target_id=All&created_1=All&title=gross%20domestic%20product&page=0	Includes first, second and third estimates. Available from 1996
Industrial production releases	https://www.federalreserve.gov/releases/g17/release_dates.htm	

Notes: Prior to 1994, the FOMC did not explicitly announce its target for the federal funds rate, but implemented changes in its target via open market operations. These open market operations were conducted at 11:30am the next morning (see Swanson & Jayawickrema, 2023). Therefore, we use the next day after a regularly scheduled FOMC meeting as event day from 1982 - 1987.

Appendix 2.F Identification of relevant speeches

Swanson (2023) and Swanson and Jayawickrema (2023) show that speeches and Congressional testimony (henceforward just speeches) of the Chair and the Vice-Chair of the Federal Reserve Board are an important source of variation in U.S. monetary policy. Therefore, we augment the event dataset with carefully selected speeches. In contrast to Swanson and Jayawickrema (2023) who read the newspaper the next morning to judge whether the speech had implications for monetary policy, we use a more data-driven approach to identify relevant speeches.

To identify speeches of the Chair and the Vice-Chair of the Federal Reserve, we apply a Correlated Topic Model (CTM).[27] Specifically, we downloaded all speeches by officials of the Federal Reserve from 1996 onwards from the website of the Federal Reserve.[28] Speeches before 1996 are collected from ALFRED. [29] We then identify and link together words that often appear together (e.g., interest rate or Federal Reserve) by calculating bigrams (contiguous sequences of two words). Finally, after cleaning the corpus from stopwords, numbers, and punctuation, we apply the CTM.

The CTM was initially developed by Blei and Lafferty (2007). Here, we use the algorithm described in Roberts et al. (2016), Roberts et al. (2019). Specifically, we estimate a Structural Topic Model (STM), which reduces to a fast implementation of the CTM if estimated without covariates.[30] The CTM is a statistical model used to analyze large sets of documents. It assumes that each document in the collection is made up of a mixture of different topics, and each topic is a probability distribution over the words in the vocabulary. It is superior in this context to other topic modeling approaches, such as Latent Dirichlet Allocation (LDA), because it explicitly models the correlations between the topics, which may be important for understanding the underlying structure of the data. For classifying speeches of Federal Reserve officials, there may be certain topics that are frequently discussed together (such as inflation and monetary policy) or that have a strong influence on each other.

In the following, we use the notation as in Roberts et al. (2016). We denote the documents

[27] We follow Swanson and Jayawickrema (2023) and focus on the most influential members of the FOMC: the Federal Reserve Board Chair and the Federal Reserve Board Vice Chair. However, to estimate the topic model we use all speeches given by Federal Reserve Board Governors. Thus, the resulting topics are rendered more interpretable by providing the algorithm with additional data.

[28] https://www.federalreserve.gov/newsevents/speeches.htm

[29] https://alfred.stlouisfed.org/

[30] See Roberts et al. (2016), Roberts et al. (2019) for more details on the STM.

using the index $d \in \{1, \ldots, D\}$ and the words (or positions within the documents) using the index $n \in \{1, \ldots, N\}$. Each word in a document, represented as $w_{d,n}$, is an instance of distinct words drawn from a vocabulary that is indexed by $v \in \{1, \ldots, V\}$. Additionally, the model assumes the selection of a certain number of topics, K, which are indexed by $k \in \{1, \ldots, K\}$.

The CTM is a generative model assuming that each document d, given the number of topics K and observed words $w_{d,n}$, is generated in the following way:

$$\eta_d \sim \mathcal{N}_{K-1}(\mu, \Sigma) \tag{2.56}$$

$$\theta_{d,k} = \frac{\exp(\eta_{d,k})}{\sum_{i=1}^{K} \exp(\eta_{d,i})} \tag{2.57}$$

where η_d is the latent topic proportion vector for document d, transformed to the simplex via a logistic function to get θ_d. $\eta_{d,K}$ is fixed to zero to identify the model. μ is the mean vector, and Σ is the covariance matrix (capturing topic correlations) of the topic proportion. Given the topic proportion vector, θ_d, for each word, indexed by n, within document d, a topic indicator is sampled from

$$z_{d,n} \sim \text{Multinomial}_K(\theta_d) \tag{2.58}$$

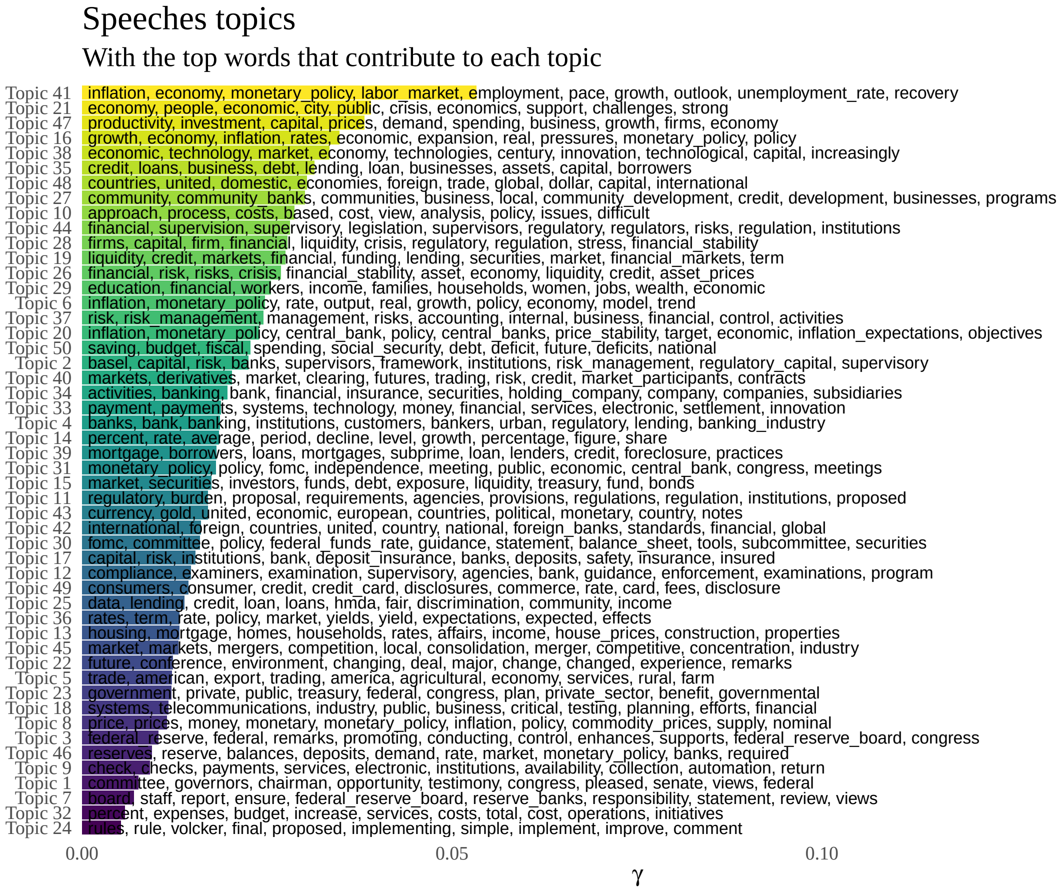
whose positive component indicates the topic associated with that particular position. Conditional on such a topic indicator, a word is sampled from

$$w_{d,n} \sim \text{Multinomial}_V(\beta_{z_{d,n}}) \tag{2.59}$$

where V is the size of the vocabulary and β is the $K \times V$ matrix representing the distributions of terms in the vocabulary corresponding to the K topics.

The objects of interest in a CTM include the distributions of topics within documents (θ_d), the distributions of words across topics (β), the topic assignments for each word ($z_{d,n}$), and the parameters (μ, Σ) of the logistic normal distribution. Estimating these components allows for a comprehensive understanding of the thematic structure present in a text corpus. However, inference in a CTM is challenging due to the non-conjugate nature of the logistic normal and multinomial distributions. Here, an approximate variational EM algorithm using a Laplace approximation developed by Roberts et al. (2016) is used. We refer to their paper for more details on the estimation procedure.

Figure 2.F.1 — Topic prevalence with the top words that contribute to topics

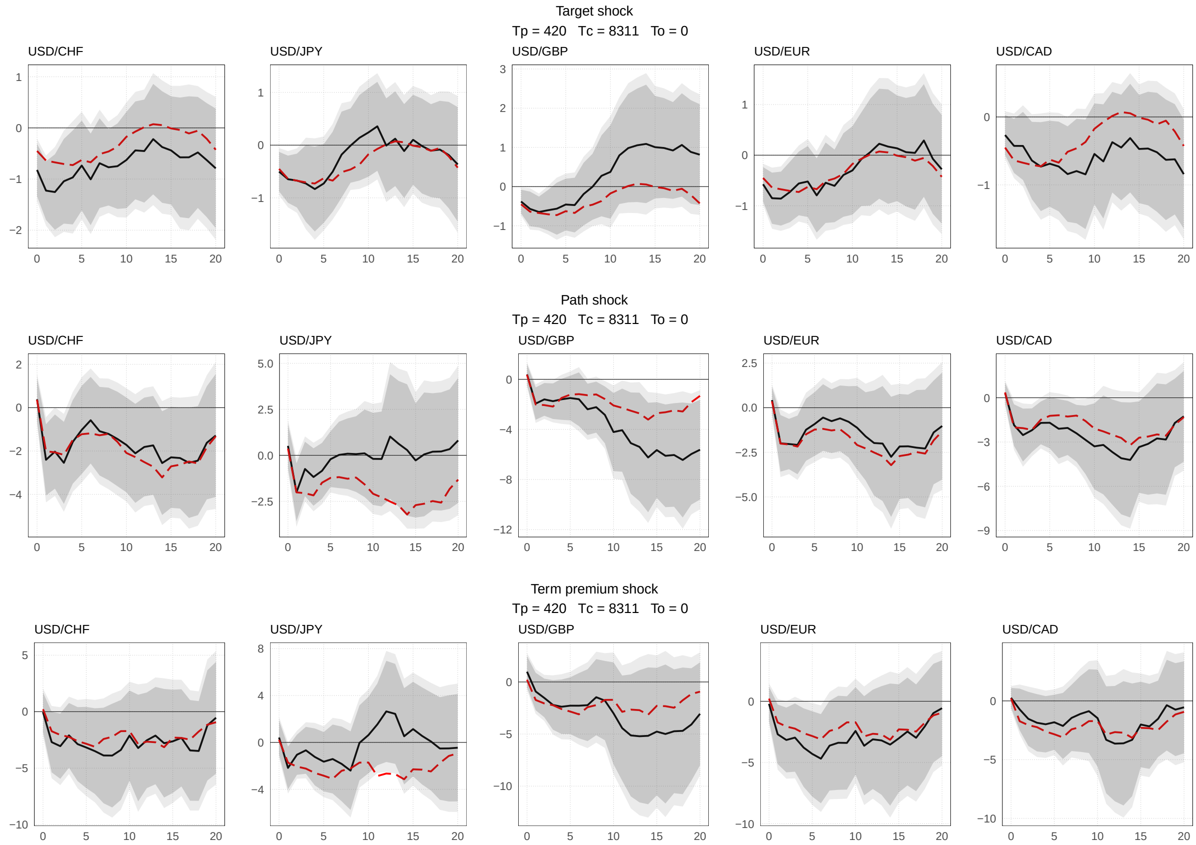


We use the default values of the algorithm developed by Roberts et al. (2019) and set the number of topics to $K = 50$. Figure 2.F.1 illustrates the identified topics in terms of their frequency of occurrence within the text corpus and the words that most accurately describe them. From a human standpoint, the top words are perceived as coherent and meaningful, resulting in the interpretability of the topics. Therefore, we use the top words to identify the topics associated with monetary policy. We explicitly choose only those topics that are directly related to monetary policy. In doing so, we can avoid possible concerns regarding information effects (see Nakamura & Steinsson, 2018). According to our judgment, the following six topics hold the greatest relevance for monetary policy: 6, 20, 30, 31, 36, 46. Accordingly, we identify and include 81 speeches in the event dataset.

Appendix 2.G Robustness tests

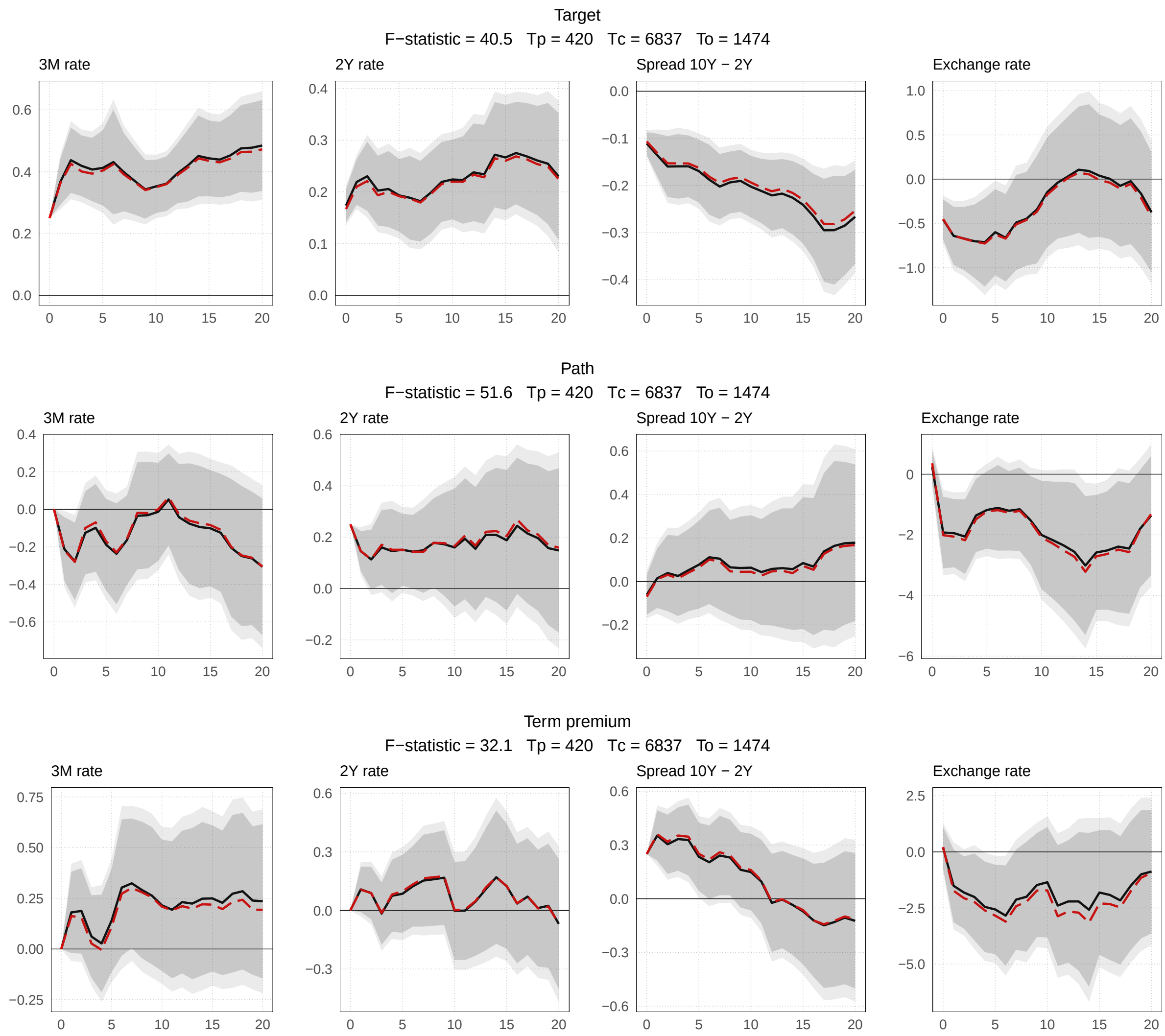
2.G.1 Daily responses

Figure 2.G.1 — Bilateral exchange rates



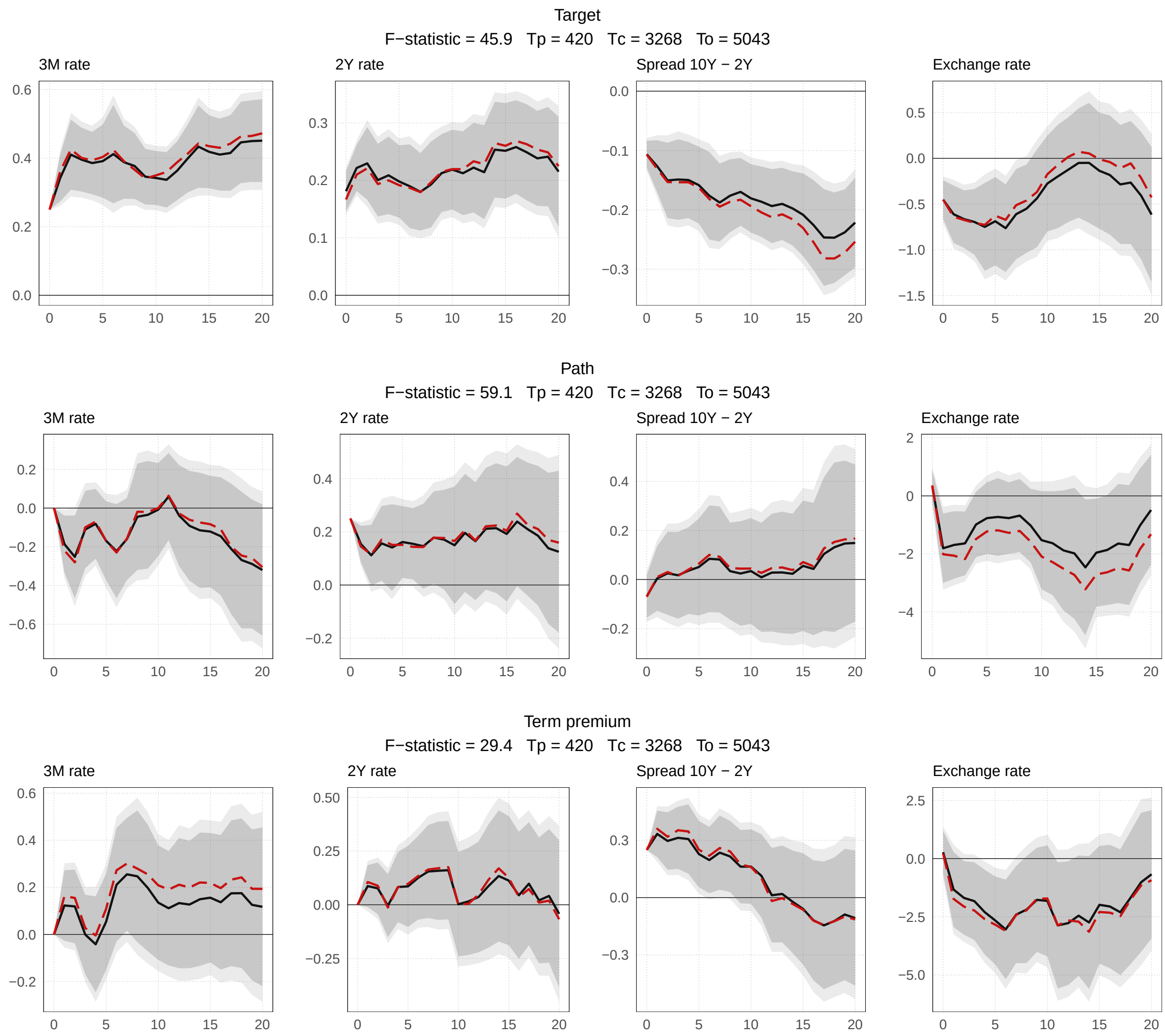
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated for every bilateral exchange rate. The horizontal axis is measured in working days (excluding weekends and holidays). The exchange rate response is measured in percent. Red dashed lines give the baseline response of the trade-weighted exchange rate. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.2 — Excluding data releases



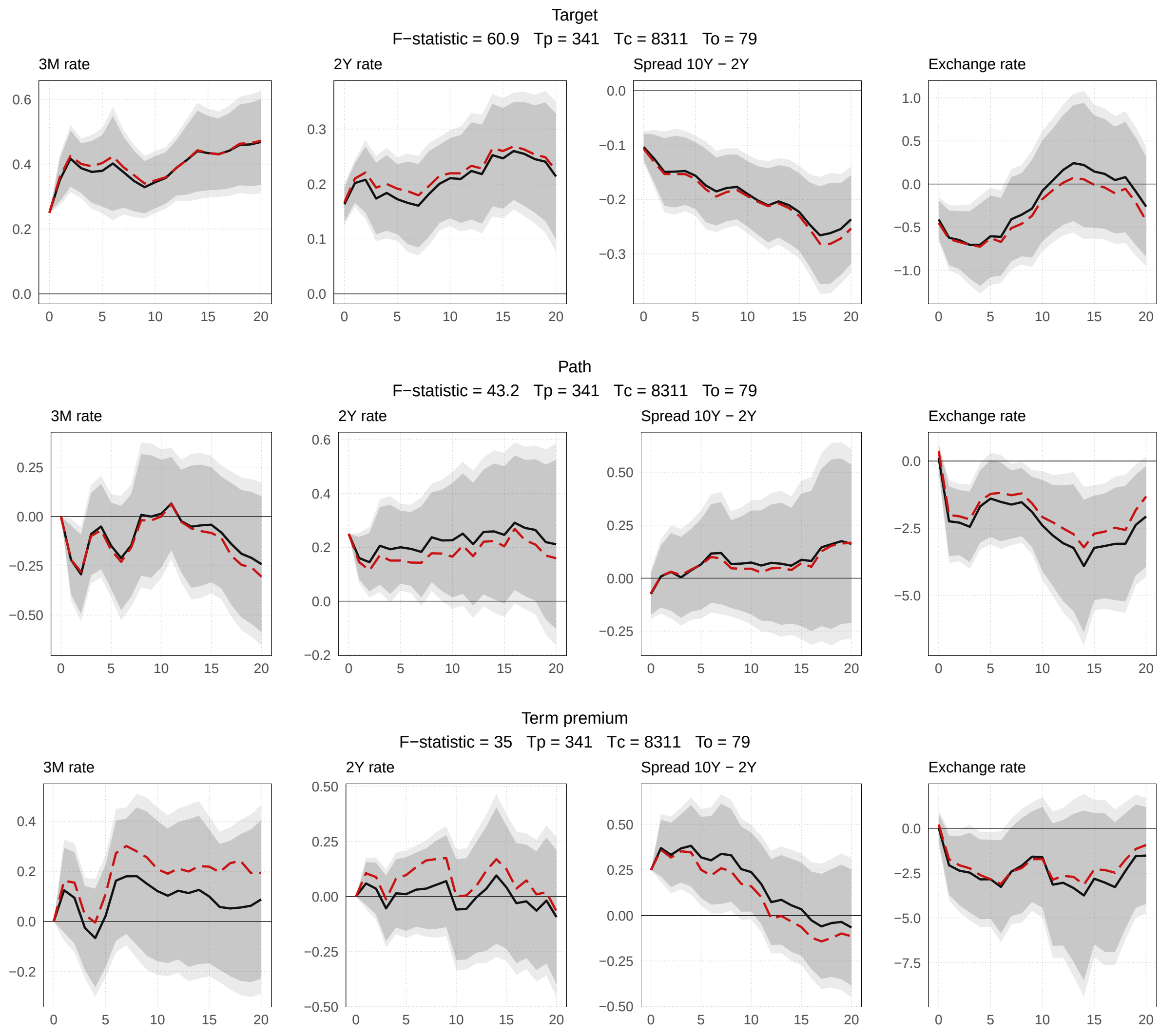
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated on a sample excluding important data releases. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.3 — Tuesday and Wednesday as control days



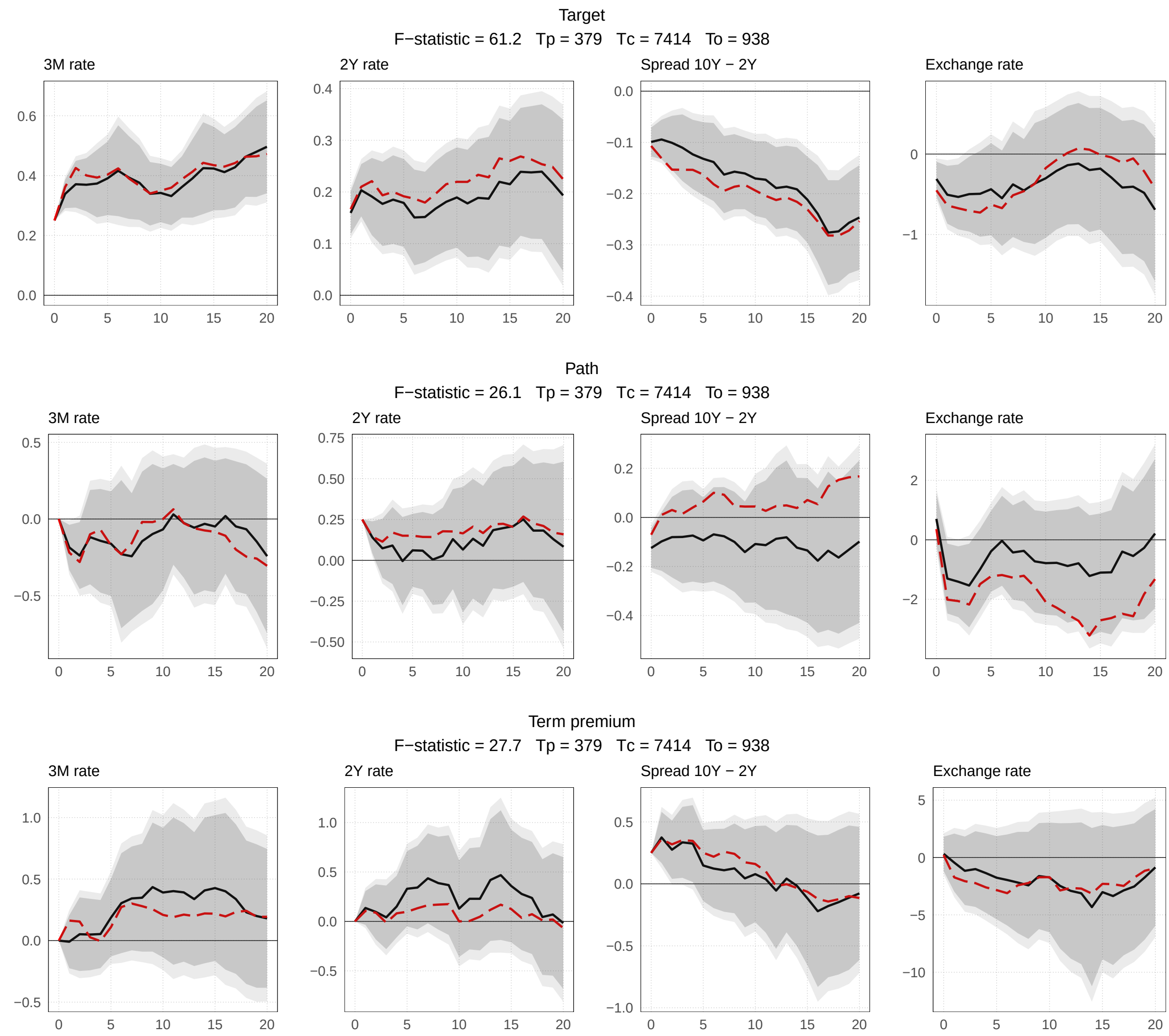
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The horizontal axis is measured in working days (excluding weekends and holidays). The model is estimated on a sample using only Tuesdays and Wednesdays as control days. All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.4 — Excluding speeches



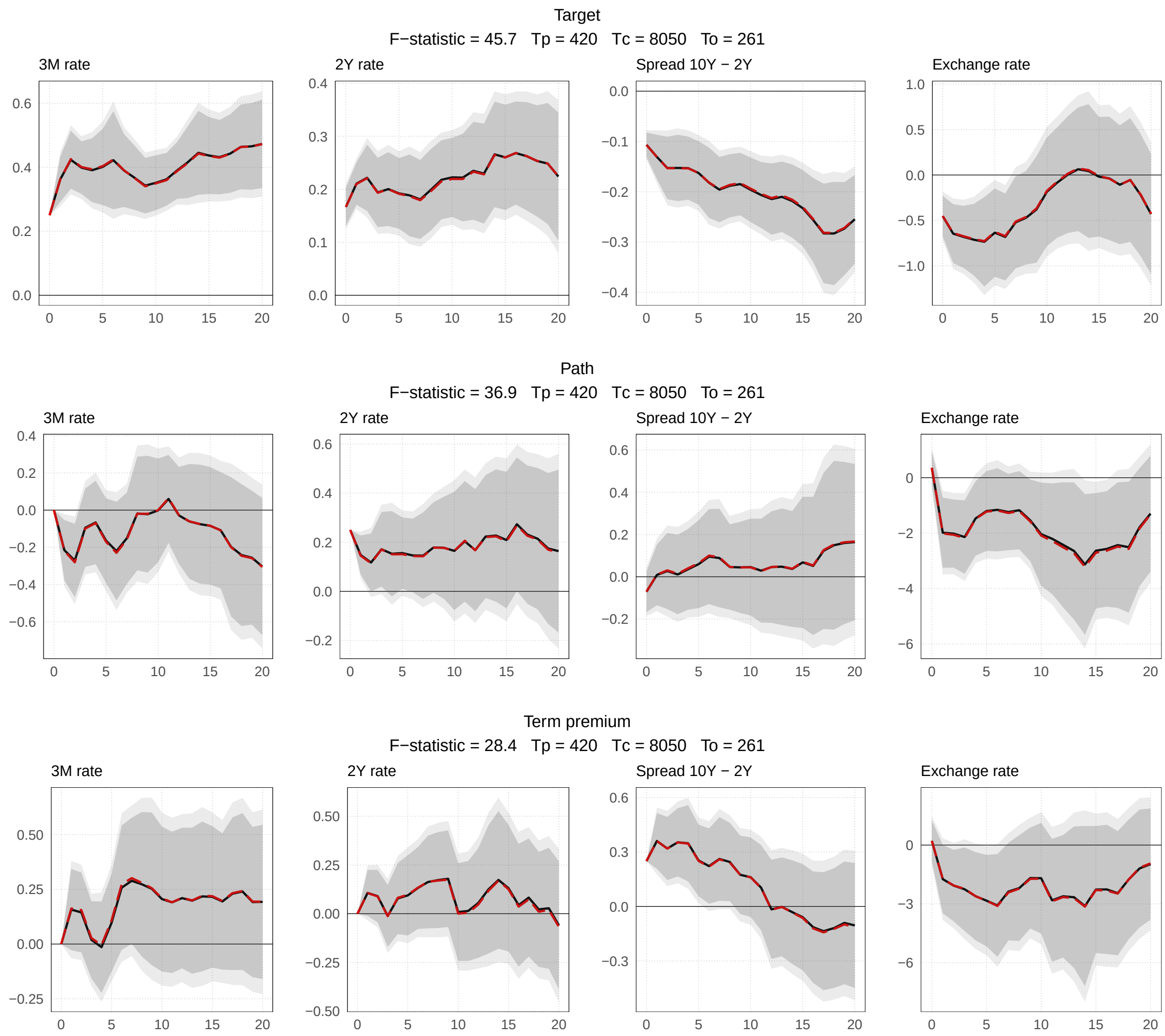
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated on a sample excluding speeches. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy yevent da, control, and other days, respectively.

Figure 2.G.5 — Excluding volatile crisis periods



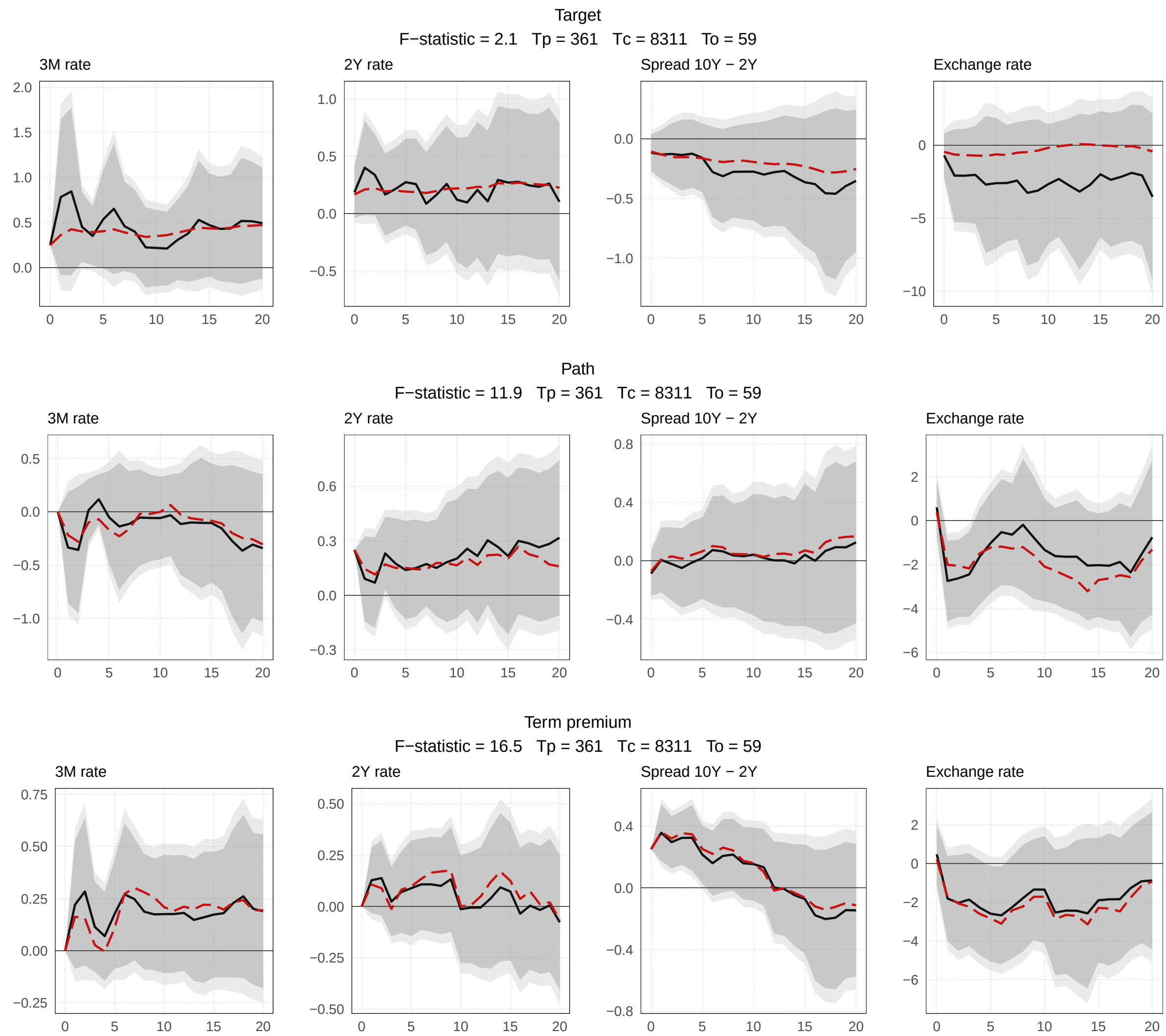
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated on a sample excluding volatile crisis periods. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.6 — Excluding FOMC minutes releases



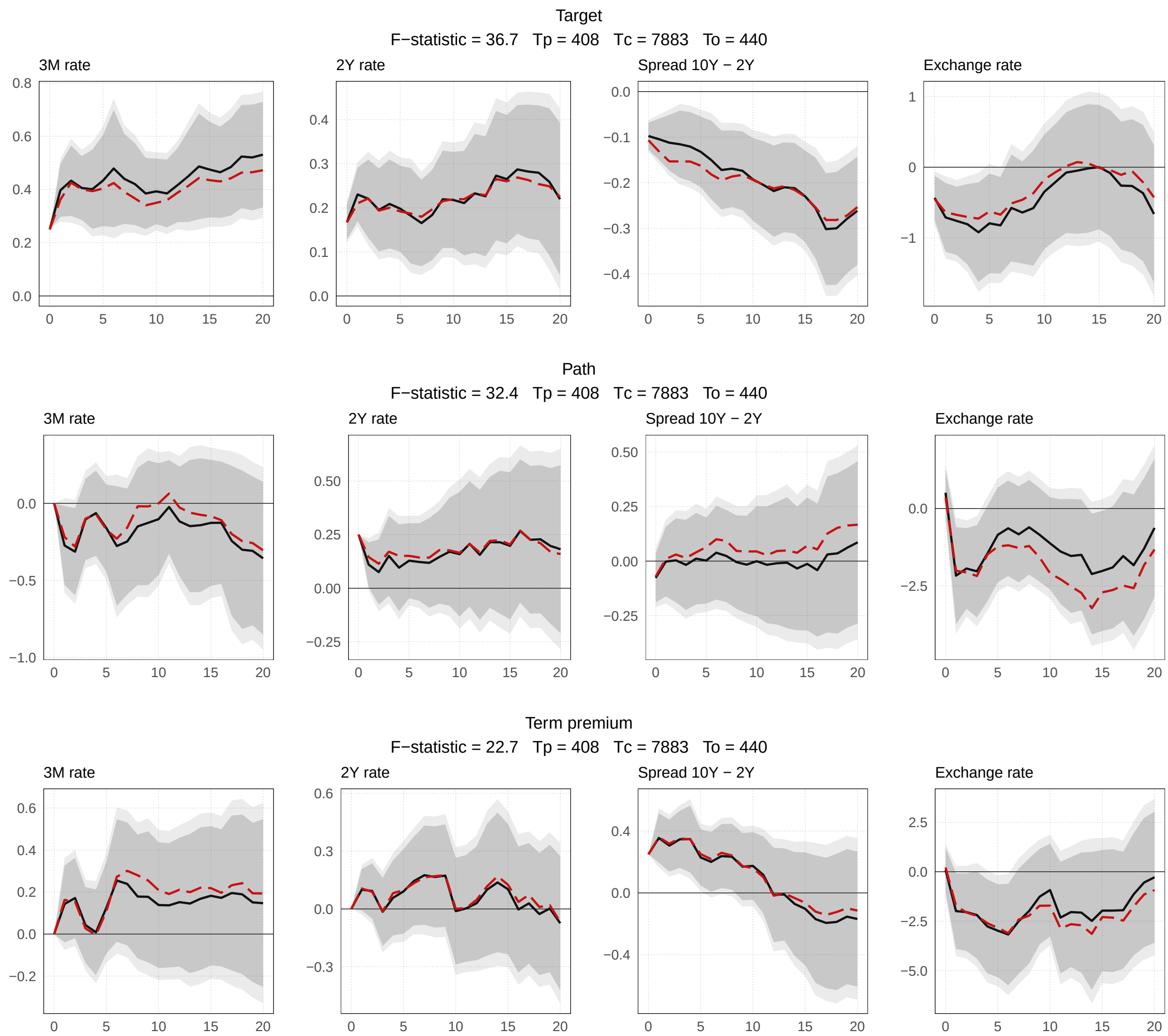
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated on a sample excluding days with releases of FOMC minutes. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.7 — Excluding unscheduled policy events



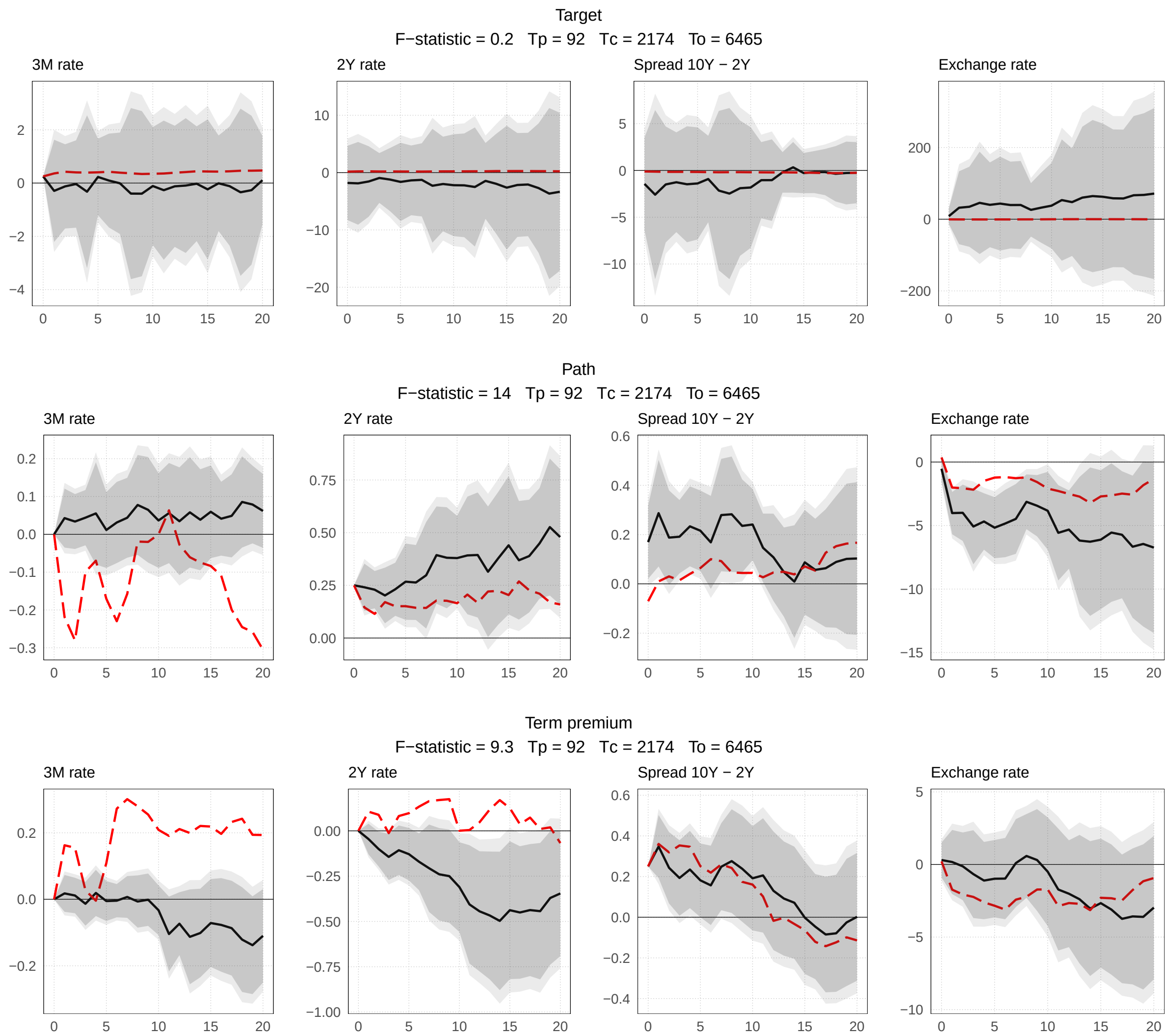
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated on a sample excluding unscheduled FOMC decisions. The horizontal axis is measured in working days (excluding unscheduled policy events). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.8 — Excluding ECB and BoE decisions



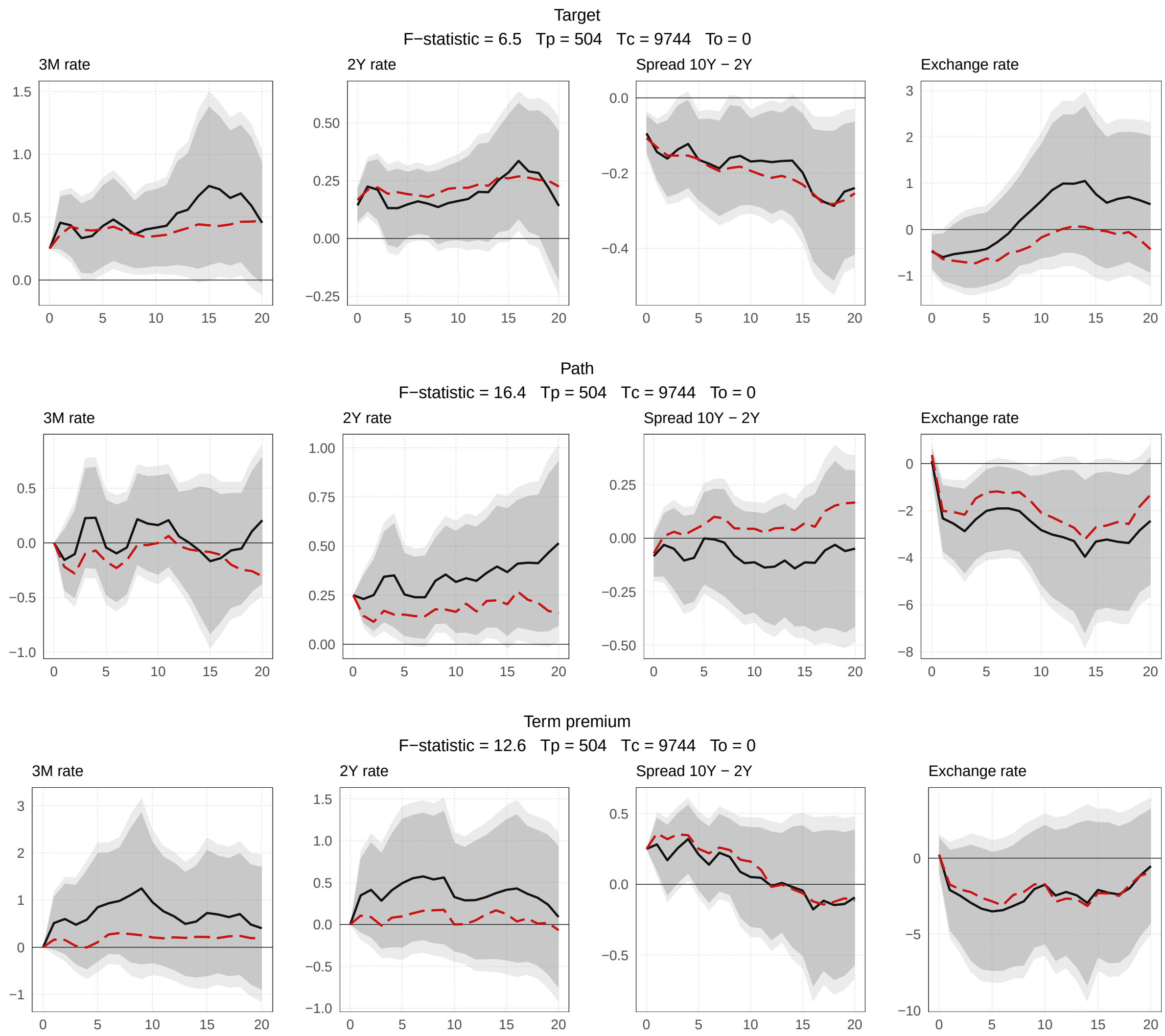
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated on a sample excluding days with decisions by the ECB and the Bank of England. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.9 — At the effective lower bound



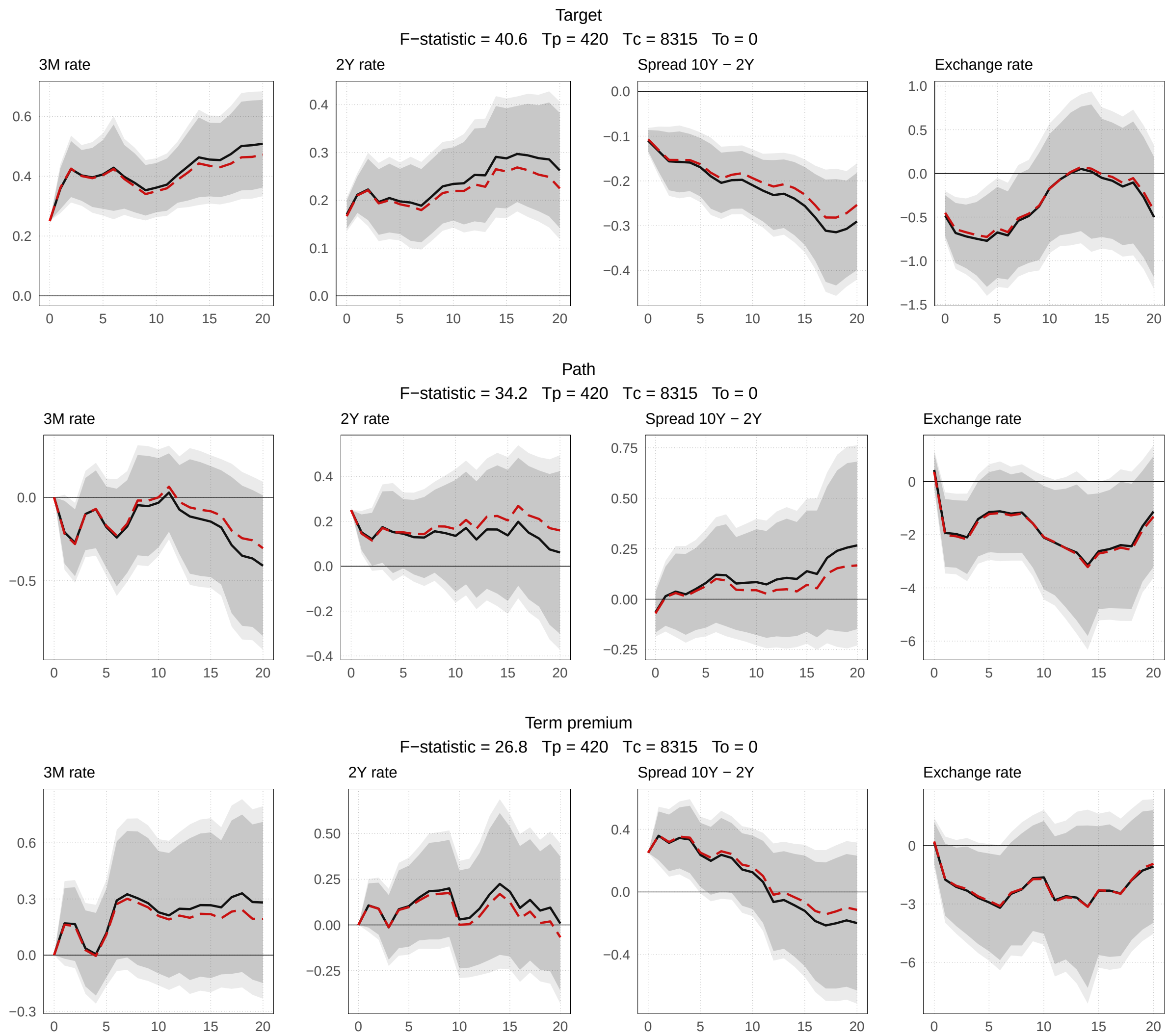
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated on a sample restricted to the effective lower bound period. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.10 — Long sample (1982–2022)



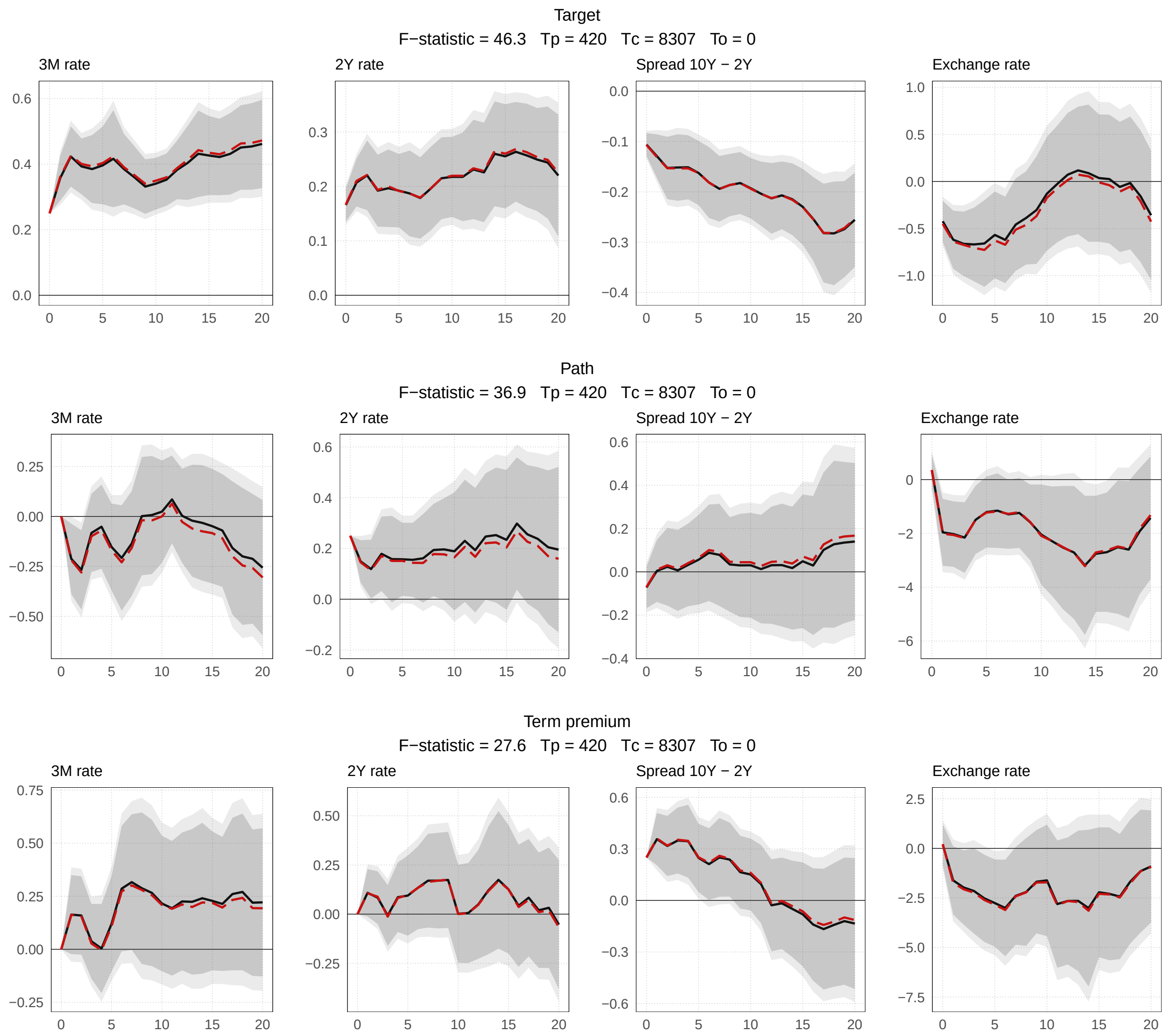
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated on a long sample from 1982–2022. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.11 — No controls



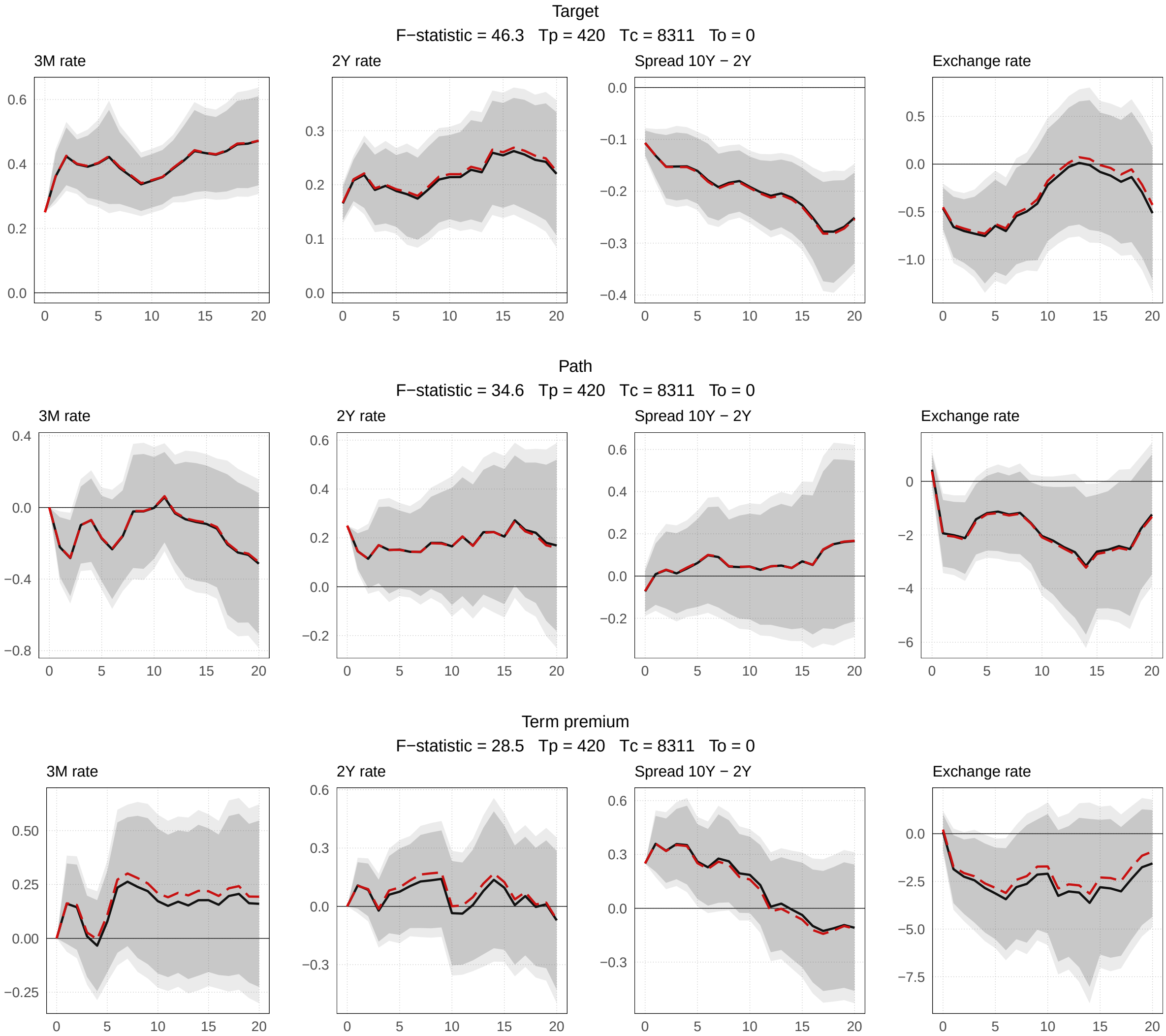
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated without controls. The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

Figure 2.G.12 — Additional lags ($P = 8$)



Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model includes additional lags of the dependent variables ($P = 8$). The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

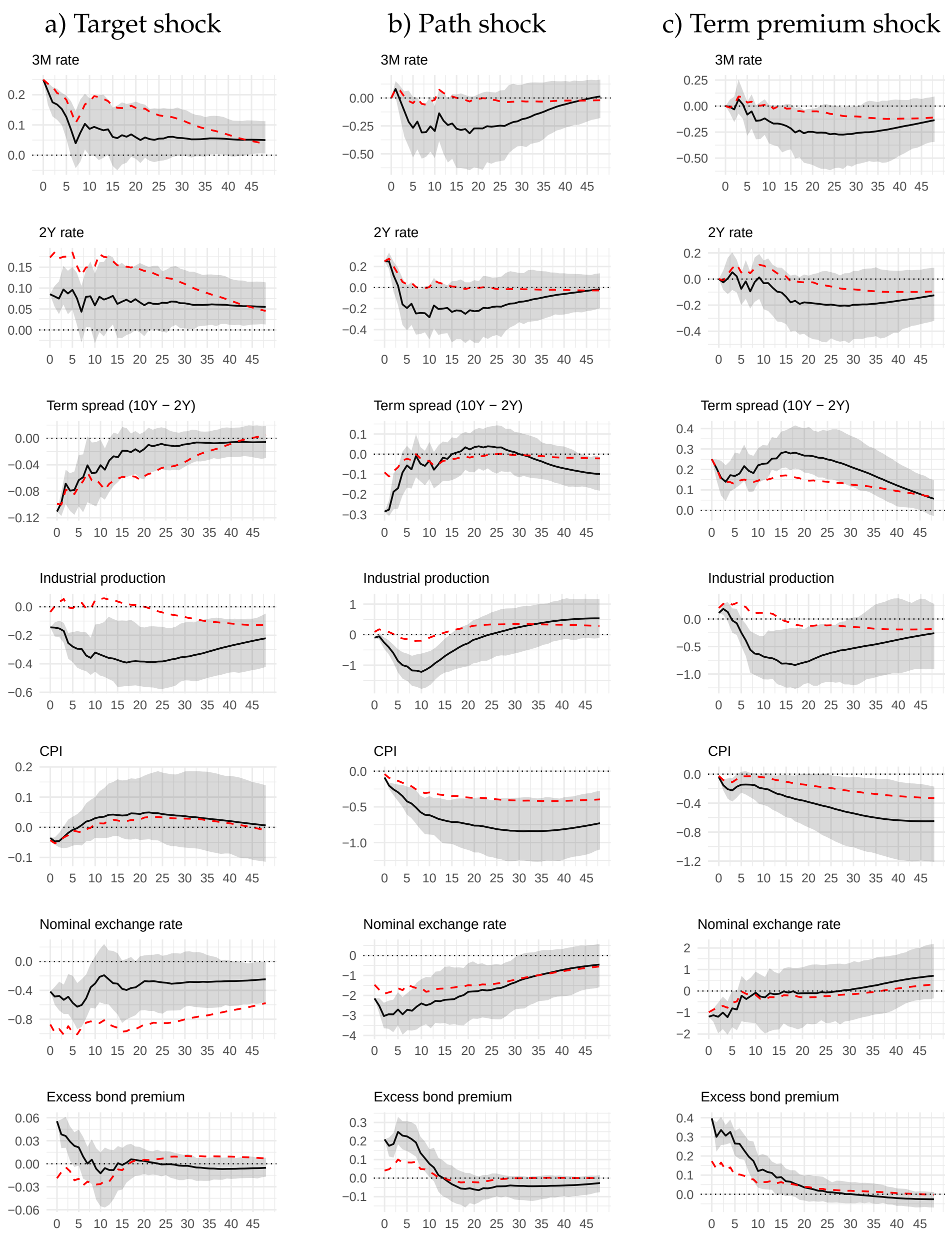
Figure 2.G.13 — Additional controls



Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model is estimated with additional controls (stock price index, commodity price index, corporate bond spread, news sentiment index). The horizontal axis is measured in working days (excluding weekends and holidays). All interest rate responses are measured in percentage points. The exchange rate response is measured in percent. The red dashed lines give the responses in the baseline model. 90% and 95% confidence intervals are based on HAC-robust standard errors. T_p, T_c, T_o denote the number of policy event days, control days, and other days, respectively.

2.G.2 Monthly responses

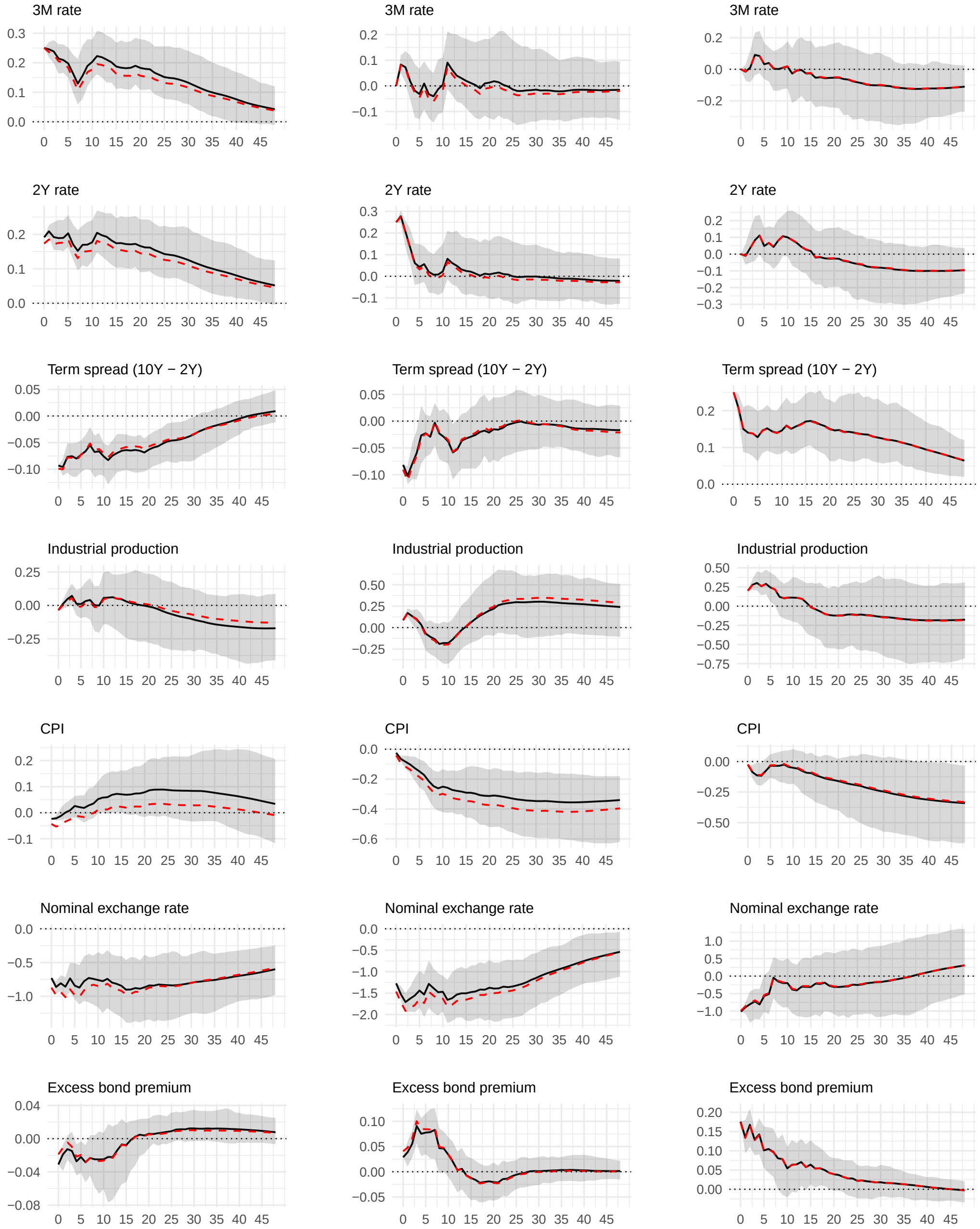
Figure 2.G.14 — With poor man's sign restrictions



Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The red dashed lines indicate our baseline responses. The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. Shocks that move in the same direction as stock prices are set to zero (see Jarociński & Karadi, 2020). The horizontal axis is measured in months. All interest rate and bond premium responses are measured in percentage points. All other responses are measured in percent. 90% confidence intervals are based on a moving block bootstrap with 10,000 replications.

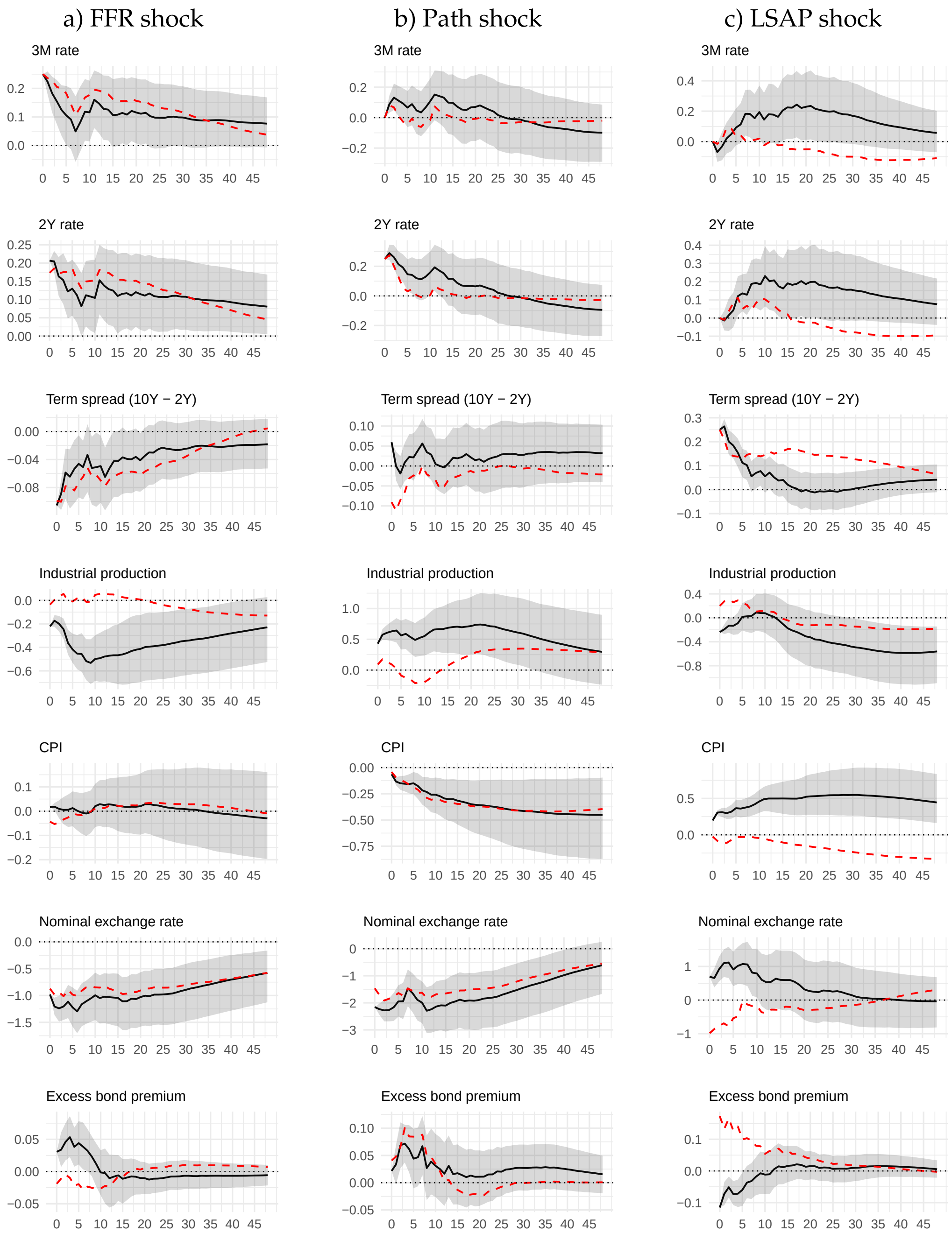
Figure 2.G.15 — Setting the target to zero at the effective lower bound

a) Target shock b) Path shock c) Term premium shock



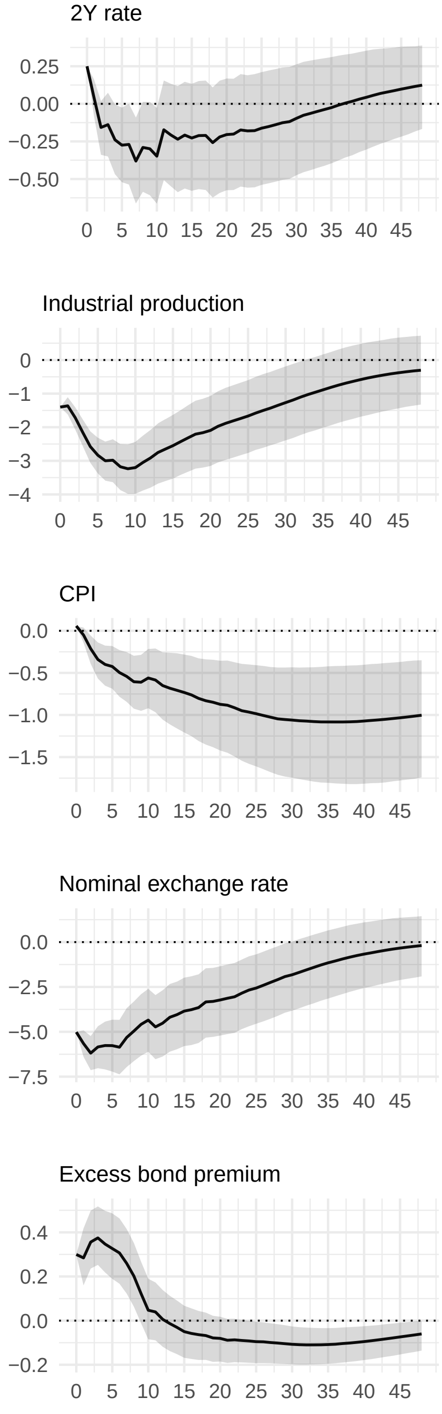
Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The red dashed lines indicate our baseline responses. The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The target shock is set to zero at the effective lower bound. The horizontal axis is measured in months. All interest rate and bond premium responses are measured in percentage points. All other responses are measured in percent. 90% confidence intervals are based on a moving block bootstrap with 10,000 replications.

Figure 2.G.16 — Using the shocks of Swanson (2021) as external instruments



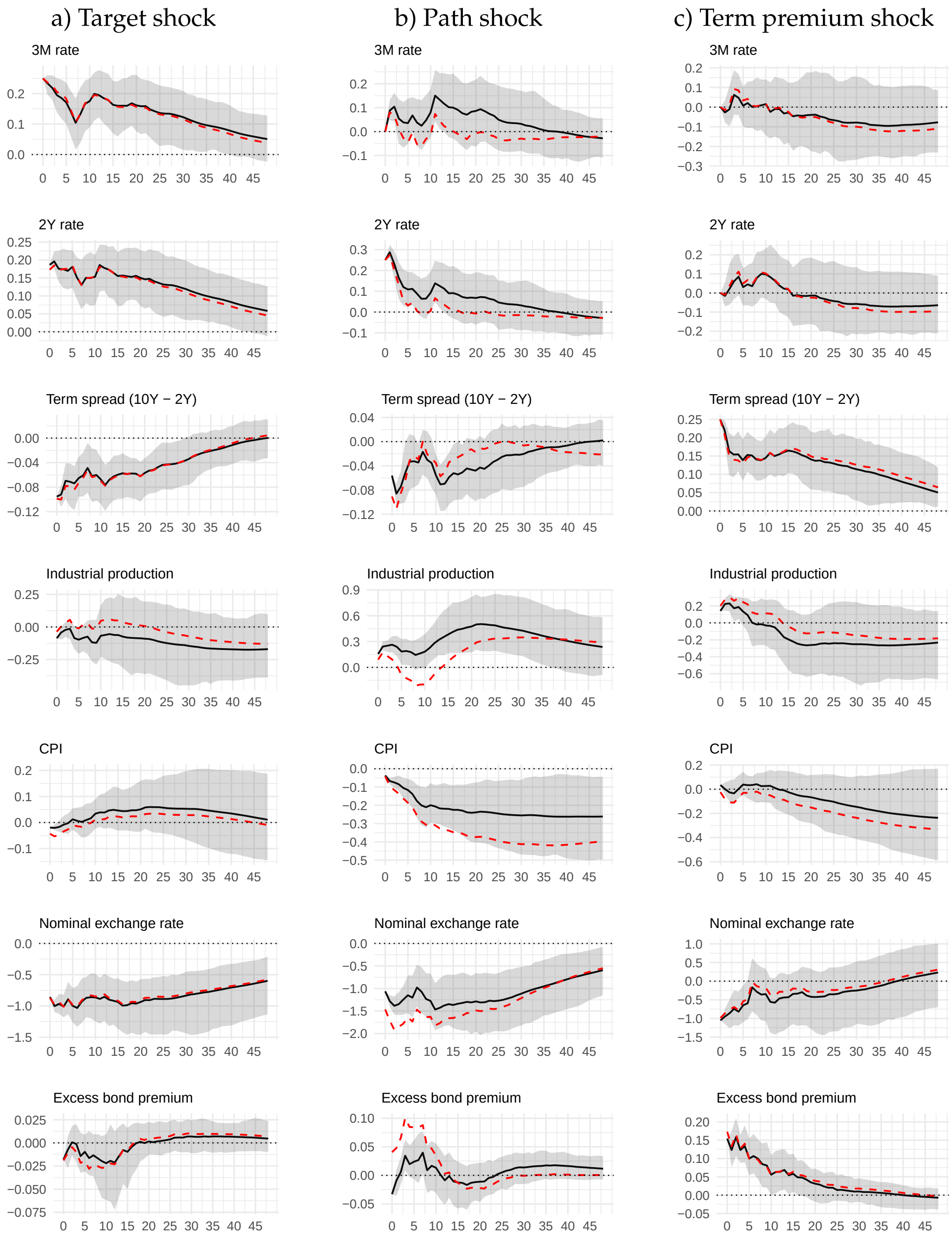
Notes: Impulse responses to monetary policy shocks identified by Swanson (2021). The red dashed lines indicate our baseline responses. The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The horizontal axis is measured in months. All interest rate and bond premium responses are measured in percentage points. All other responses are measured in percent. 90% confidence intervals are based on a moving block bootstrap with 10,000 replications.

Figure 2.G.17 — Using the shocks of Bu et al. (2021) as external instruments



Notes: Impulse responses to monetary policy shocks identified by Bu et al. (2021). The responses are normalized to a 25 bp increase in the 2Y rate. The horizontal axis is measured in months. All interest rate and bond premium responses are measured in percentage points. All other responses are measured in percent. 90% confidence intervals are based on a moving block bootstrap with 10,000 replications.

Figure 2.G.18 — Using our shocks together with the Swanson (2021) shocks as external instruments



Notes: Impulse responses to monetary policy shocks (target, path, and term premium). The red dashed lines indicate our baseline responses. The responses are normalized to a 25 bp increase in the 3M rate, 2Y rate, and 10Y - 2Y spread, respectively. The model uses our baseline shocks and the shocks by Swanson (2021) as external instruments. It is, therefore, overidentified. The horizontal axis is measured in months. All interest rate and bond premium responses are measured in percentage points. All other responses are measured in percent. 90% confidence intervals are based on a moving block bootstrap with 10,000 replications.

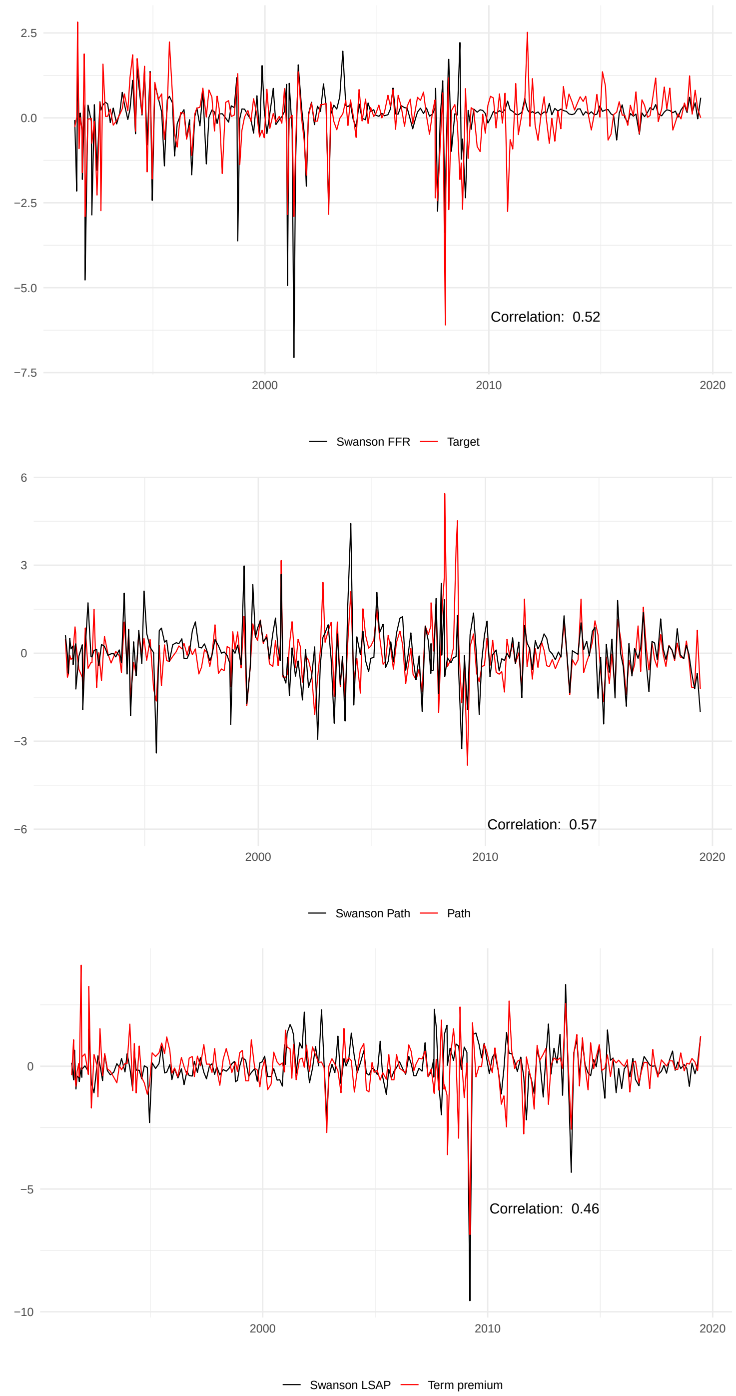
Appendix 2.H Supplementary material

Table 2.H.1 — Comparison to existing shocks aggregated to monthly frequency

	Swanson FFR	Swanson Path	Swanson LSAP	Bu et al. (2021)
Target	0.955^{***}	0.042	-0.293^{***}	0.052^{***}
	(0.312)	(0.184)	(0.098)	(0.019)
Path	0.448	1.739^{***}	0.478^{*}	0.104^{***}
	(0.304)	(0.358)	(0.260)	(0.017)
Term premium	-1.049^{**}	0.372	1.259^{**}	0.005
	(0.413)	(0.327)	(0.583)	(0.020)
Observations	336	336	336	312
R^2	0.252	0.250	0.250	0.551
Adjusted R^2	0.245	0.243	0.244	0.546

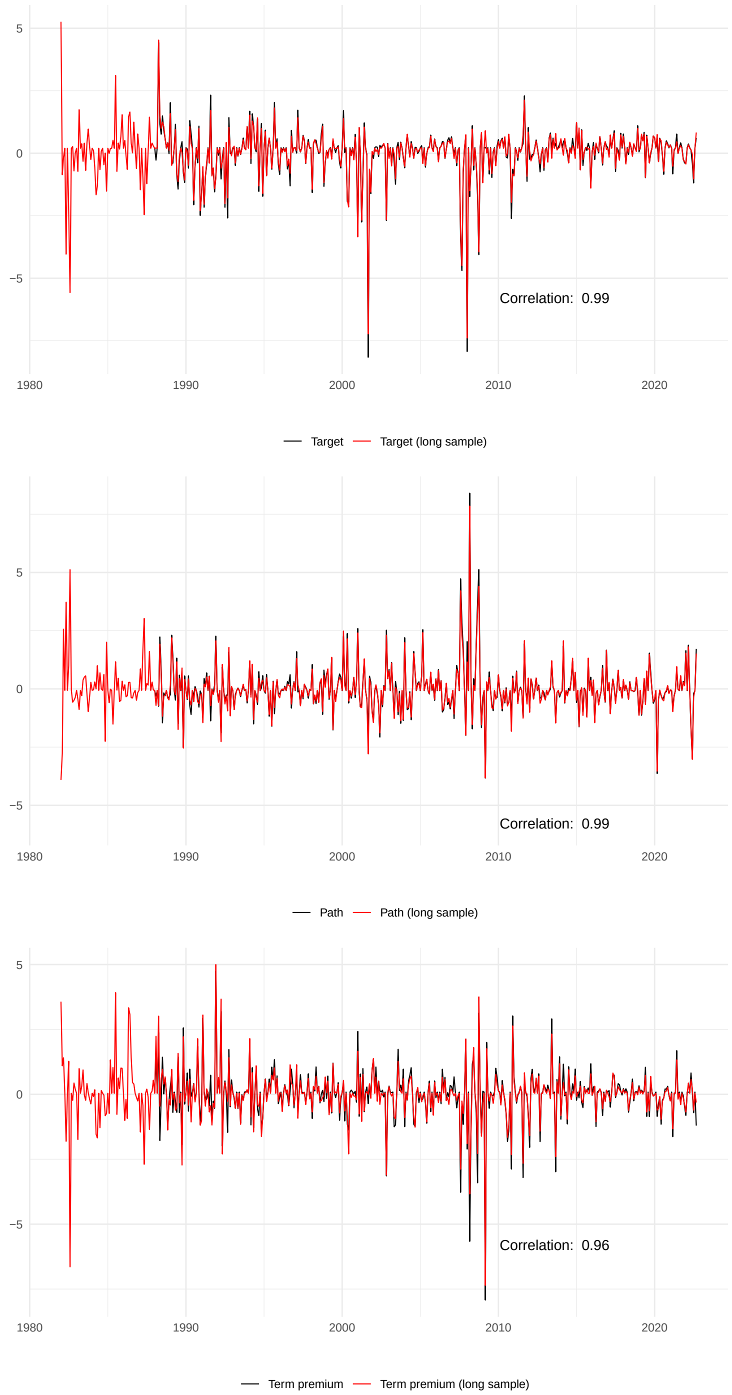
Notes: The table shows regressions of our estimated monetary policy shocks on the shocks provided by Swanson (2021) and Bu et al. (2021). Before estimating the regressions, the shocks have been aggregated to monthly by summing them up within the same month. Significance levels are given by *p<0.1; **p<0.05; ***p<0.01. HAC robust standard errors are in parentheses.

Figure 2.H.1 — Heteroscedasticity-based compared to high-frequency shocks



Notes: These graphs compare our estimated three-dimensional monetary policy shock series to the high-frequency series by Swanson (2021) on FOMC announcement dates from July 1991 to June 2019. For readability, the series have been normalized to have a mean of zero and a standard deviation of one.

Figure 2.H.2 — Heteroscedasticity-based shocks over long and short sample



Notes: These graphs compare our estimated three-dimensional daily monetary policy shock series once estimated on a sample from 1982 to 2019 and once on a sample from 1988 to 2019. For readability, the series have been normalized to have a mean of zero and a standard deviation of one.

Figure 2.H.3 — Heteroscedasticity-based shocks with alternative interest rate data

Notes: These graphs compare our three-dimensional monetary policy shock series to the high-frequency series by Swanson (2021) on FOMC announcement dates from July 1991 to June 2019. The estimates are based on alternative interest rate data from the Federal Reserve Board. For readability, the series have been normalized to have a mean of zero and a standard deviation of one.

Table 2.H.2 — Comparison to existing shocks using alternative interest rate data

	Swanson FFR	Swanson Path	Swanson LSAP	Bu et al. (2021)
Target	1.852***	0.316	−0.273*	0.150***
	(0.313)	(0.307)	(0.146)	(0.014)
Path	0.567**	2.334***	0.297	0.148***
	(0.245)	(0.417)	(0.208)	(0.010)
Term premium	−0.630*	0.745**	2.070***	0.003
	(0.382)	(0.331)	(0.593)	(0.011)
Observations	241	241	241	215
R^2	0.385	0.376	0.360	0.788
Adjusted R^2	0.378	0.368	0.352	0.785

Notes: The table shows regressions of our estimated monetary policy shocks using alternative interest rate data by the Federal Reserve Board on the shocks provided by Swanson (2021) and Bu et al. (2021). Significance levels are given by *p<0.1; **p<0.05; ***p<0.01. HAC robust standard errors are in parentheses.

Table 2.H.3 — Correlation matrix of monetary policy shocks

	Target	Path	Term premium	FFR	Path	LSAP	BRW
Target	1.00						
Path	-0.23	1.00					
Term premium	0.21	-0.11	1.00				
Swanson FFR	0.52	0.03	-0.10	1.00			
Swanson Path	-0.12	0.57	0.03	-0.00	1.00		
Swanson LSAP	-0.14	0.18	0.46	-0.00	0.00	1.00	
BRW	0.42	0.73	0.06	0.33	0.49	0.03	1.00

Notes: The table shows the correlation matrix of the monetary policy shocks. For comparison, we also add the shocks by Swanson (2021) and Bu et al. (2021).

Table 2.H.4 — Correlation matrix of monetary policy shocks censoring target shock at effective lower bound

	Target	Path	Term premium	FFR	Path	LSAP	BRW
Target	1.00						
Path	-0.27	1.00					
Term premium	0.28	-0.11	1.00				
Swanson FFR	0.58	0.03	-0.10	1.00			
Swanson Path	-0.14	0.57	0.03	-0.00	1.00		
Swanson LSAP	-0.09	0.18	0.46	-0.00	0.00	1.00	
BRW	0.28	0.73	0.06	0.33	0.49	0.03	1.00

Notes: The table shows the correlation matrix of the monetary policy shocks. For comparison, we also add the shocks by Swanson (2021) and Bu et al. (2021). The target shock is set to zero at the effective lower bound.

3 Three centuries of Swiss economic sentiment†

3.1 Introduction

"The farther backward you can look, the farther forward you are likely to see."

— Winston Churchill

Churchill's insight resonates profoundly with the objective of this chapter, which takes on the challenge of uncovering Switzerland's business cycle history. A story not entirely told due to a lack of well-measured economic data for the 19th and early 20th centuries. In the first instance, consistent and well-measured gross domestic product (GDP) statistics, which are often used to describe business cycle fluctuations, are missing. There is only a tentative estimate of annual GDP since 1851 and no information before that (see Historische Statistik der Schweiz HSSO, 2012k; Stohr, 2016). Moreover, real GDP estimates suffer from error-prone deflators that bias econometric estimates and descriptive statistics (Kaufmann, 2020). The lack of well-measured historical economic data is not unique to Switzerland.

†For helpful comments and discussions, I thank David Ardia, Elliott Ash, Jean-Marie Grether, Martin Huber, Malin Jensen, Daniel Kaufmann, Bruno Lanz, Jason Lennard, Jannis Stefanopulos, Christian Stohr, Rebecca Stuart, Leif Anders Thorsrud, Philipp Wegmüller, Mark W. Watson, and participants at the IRENE PhD Meeting in Neuchâtel, the PhD Macro Workshop in Basel, the SSES Annual Congress in Lucerne, the IAAE Conference in Thessaloniki, and the SECO Brown Bag Seminar. I am thankful to Regina Gloor (Tamedia), Patrick Halbeisen (SNB), Martin Lüpold (SWA), Théophile Naito (scriptorium.bcu-lausanne.ch), Jürg Rütimann (AWP), Florian Steffen (e-newspaperarchives.ch) and Mathias Wiesmann (ZKB) for providing data. Nicola Francescutto and Idy Abdoul N'Dao provided excellent research assistance. I gratefully acknowledge the support from a UniNE Doc.Mobility grant to visit the Norwegian Business School (BI).

Because hard data is difficult to measure retrospectively, I use written narratives in this chapter to fill in the gaps where traditional economic data is missing or inaccurate. Using textual data, I develop a quarterly business cycle indicator for Switzerland spanning three centuries: the 19^{th}, 20^{th}, and 21^{st}. Given the impossibility of asking people and businesses from the past about their financial and economic situation, I rely on alternative sources like business association reports, business reports, and newspapers. Following recent contributions, I construct qualitative indicators for several economic topics related to the business cycle (see e.g. Chapter 1, Thorsrud (2020) Bybee et al. (2023)). I assemble and digitize a unique dataset of historical documents relevant to business cycle fluctuations to accomplish this. One of the challenges associated with historical text sources is that they are very noisy, primarily due to issues related to quality. Therefore, I develop methods beyond natural language processing (NLP) techniques to extract relevant information. Finally, I aggregate the topic-specific indicators into a composite business cycle indicator.

The indicator effectively tracks major economic downturns, including the recessions of the late 20^{th} and early 21^{st} centuries and crises like the two World Wars and the Great Depression. Moreover, the indicator also sheds light on lesser-known economic disturbances in the 19^{th} century, such as the Sonderbund War and the Franco-Prussian War. The indicator's ability to reflect business cycle fluctuations in Switzerland, particularly in recent decades, validates its accuracy and underscores its utility for economic analysis.

Furthermore, the chapter provides a first business cycle chronology for Switzerland in the 19^{th} and early 20^{th} centuries, a period previously lacking systematic economic analysis. My research reveals that Swiss recessions have become less frequent, aligning with broader European economic trends (Broadberry & Lennard, 2023). However, contrary to the European trend, I do not find that the duration of Swiss recessions exhibits a significant decrease over the sample periods (Broadberry & Lennard, 2023). This finding challenges some prevailing narratives about the nature of business cycles.

The foundational work of Thorp (1926) and Burns and Mitchell (1946) serves as a basis for my approach of using textual data to create a business cycle indicator and chronology. Thorp (1926) used narrative accounts to describe the business cycle, highlighting the importance of qualitative data when hard data is missing. Burns and Mitchell (1946) further developed this field by introducing a more systematic and empirical approach to identifying business cycles, emphasizing the analysis of a wide range of economic indicators. This chapter bridges these methodologies by integrating advanced computational techniques to convert qualitative textual records into a quantifiable time series akin to Burns and Mitchell's (1946) methodology. An advantage of my approach is that it involves less judgment and more data-driven analysis, making it more objective and replicable without reading the documents. Moreover, it can be applied consistently across different periods.[1]

More recently, Shiller (2017, 2019) highlighted that stories and public discourse shape economic trends and, possibly, cause economic fluctuations. Narratives, often rooted in credible business and media sources, serve dual purposes: they reflect the economic conditions of their time and have the potential to influence future economic decisions and policies. Shiller's findings about the importance of stories highlight how written information can offer valuable economic insights, suggesting its broad applicability in various economic studies.

The chapter is related to a growing body of research that uses textual data to measure economic activity and sentiment.[2] The study by Van Binsbergen et al. (2024), which develops a 170-year-long measure of economic sentiment for the US, is most closely aligned with this chapter in its historical approach. Similarly, Kabiri et al. (2023) developed a monthly sentiment index for the United States from 1920 to 1934. For more contemporary periods, Burri and Kaufmann (2020), Bybee et al. (2023), Larsen and Thorsrud (2019), Shapiro et al. (2022), and Thorsrud (2020) have utilized text mining and machine learning to analyze newspaper articles, creating leading indicators and business cycle measures. Complementing these efforts, studies by Ardia et al. (2019), Barbaglia et al. (2023), Burri (2023), Ellingsen et al. (2022), and Kalamara et al. (2022) demonstrate the effectiveness of textual data in predicting various economic variables.

I make several contributions to this body of research. First, while previous research primarily focused on contemporary data over short periods, this chapter applies similar methodologies to long historical episodes. I introduce a quarterly business cycle indicator that is the most extensive record of Swiss business cycle fluctuations. Second,

[1]This contrasts the methodology used for determining NBER recession dates, which varies over time (see e.g. Romer, 1994; Romer & Romer, 2020).

[2]Textual data are also used for a variety of other purposes in economics, such as predicting stock returns (Hanna et al., 2020; Hirshleifer et al., 2023; Tetlock, 2007), measuring economic policy uncertainty (Ardia et al., 2021; Baker et al., 2016; Larsen, 2021), identifying monetary policy shocks (Aruoba & Drechsel, 2024; Ter Ellen et al., 2022), and measuring inflation expectations (Angelico et al., 2022; Binder, 2016; Larsen et al., 2021). Ash and Hansen (2023) and Beach and Hanlon (2022) provide an extensive review.

I use a broad range of historical documents relevant to business cycle fluctuations, including business reports, association documents, and archival material, diverging from related studies that primarily rely on well-structured and often categorized newspaper articles. This unique and diverse textual dataset enriches the analysis and makes it more complex, thus requiring adequate methodology. Therefore, I present a method for extracting meaningful insights from noisy and heterogeneous historical text sources. Finally, I establish the first business cycle dating for Switzerland in the 19th and early 20th centuries.

The remainder of this chapter is organized as follows. In the next section, I describe the textual data in more detail. Section 3.3 explains the methodology to create the business cycle indicator. In section 3.4, I evaluate the indicators and discuss the business cycle chronology. Section 3.5 conducts a series of robustness checks. The last section concludes.

3.2 Data

This chapter aims to construct a Swiss business cycle indicator using textual data for the 19th and early 20th century, a period for which accurate hard data is necessarily scarce. Therefore, I collected and digitized many historical documents, including company records and business association reports that potentially comprise useful information on business cycle fluctuations.[3] They comprise written information on a company's individual performance or the economy's aggregate performance. I complement this information with digitized newspaper articles. Newspapers often write about the economy or related topics like the labor market. Therefore, they can provide valuable information on the business cycle. In total, I collected 106 sources in German and French language.[4] Table 3.1 provides an overview of the number of sources per type, language, and frequency.[5] Overall, there are more newspapers than reports available. French business reports are more numerous than German ones, and newspapers and business association reports are more numerous in German. Moreover, most sources are available at a higher frequency than quarterly.

Figure 3.1 provides information on how many sources are available for each period. The

[3]See table 3.A.1 in the Appendix for a comprehensive list of all data sources. Appendix 3.B provides a detailed description of the data collection process.

[4]Even though it is one of the Swiss national languages, I did not include textual data in Italian in the analysis. The reason is that there are too few sources available.

[5]A source is the same publication available over a more extended period. This can be a newspaper (e.g., NZZ, daily 1820 - 2020) or a business report (e.g., Credit Suisse, annually 1895 - 2016)

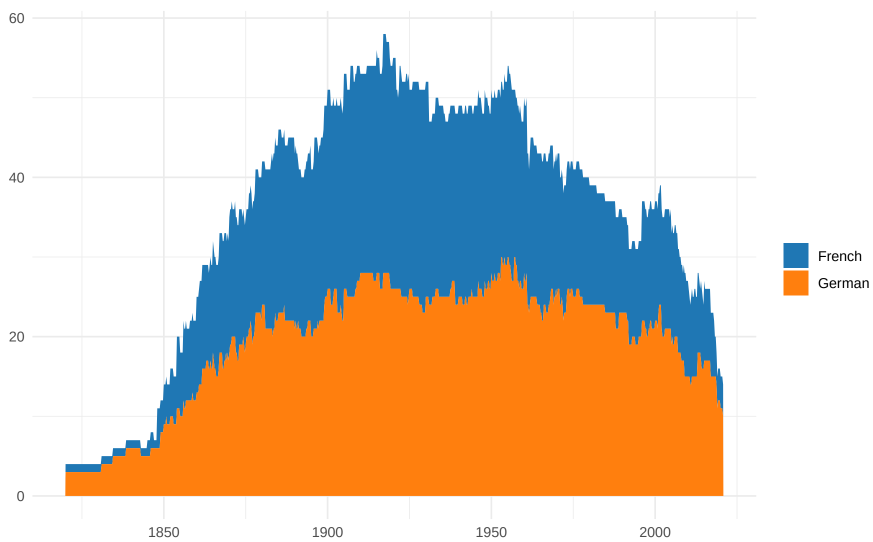
Table 3.1 — Number of text sources per type, language, and frequency

	Newspapers	Business reports	Business association reports
German	44	8	7
French	28	16	3
Frequency	$> Q$	$< Q$	$< Q$

Notes: The table shows the number of sources per type, language, and frequency. $> Q$ means the sources are available at a higher frequency than quarterly, and $< Q$ means the sources are available at a lower frequency than quarterly.

number of sources increases over time, especially after 1850. It reaches its peak around 1925 and then decreases again. Due to the historical focus of this project, I did not include more recent sources in the analysis. The Figure shows that, on average, around 40 sources are available. Moreover, the data is balanced between French and German sources. The number of sources before 1850 is below 10. However, this should not be a problem as two important Swiss newspapers are available for this period. Namely, the *Neue Zürcher Zeitung* (NZZ) from the German part of Switzerland and *L'Express* from the French part of Switzerland.

Figure 3.1 — Number of sources



Notes: The graph shows the number of available sources per quarter for both languages, French and German.

3.3 Development of a historical business cycle indicator

In this section, I describe the methodology used to extract information from text, create topic-specific indicators, and ultimately create a composite indicator of the Swiss business cycle. The method of constructing the business cycle indicator combines tools from several fields. First, I use natural language processing (NLP) techniques to extract information from the textual data. Second, I use time series econometrics methods to correct for data anomalies. Finally, I use machine learning techniques to combine the different indicators into a composite business cycle indicator. The mapping from the individual text sources to a business cycle indicator involves several steps. To provide an overview, I illustrate these steps in a schematic diagram in Figure 3.A.1 in the Appendix.

3.3.1 Extracting economic signals from text

The texts contain much information, such as advertisements, tables, and figures, irrelevant to the business cycle indicator. Therefore, I use layout parsing techniques to identify and remove these elements (see section 3.B in the Appendix). Moreover, as is standard in the literature, I remove punctuation marks, numbers, and special characters like Extensible Markup Language (XML) and Hypertext Markup Language (HTML) tags. I also remove stopwords, that is, words that are not informative, typically conjunctions such as "or" and "if" (see, e.g. Thorsrud, 2020). The stop words for both French and German are provided by Feinerer and Hornik (2019). Moreover, for each text, I count terms appearing in a German-French lexicon available from dict.cc, containing roughly 60'000 words. If a text is supposed to be in French, I only keep it if 20% of its terms are included in the French lexicon and more French than German words are identified. This filters texts with poor Oprical Character Recognition (OCR) quality or those not in German or French.

Only a fraction of the remaining texts, however, contain information about the business cycle. In the literature, it is common to use a topic modeling algorithm to classify texts into topics to further refine the selection (see, e.g. Thorsrud, 2020). However, with the amount of data available for this study, it would require a lot of computational power. Therefore, I use a simple keyword-based method to create indicators for hand-selected economic topics (see Chapter 1). This method creates indicators by using two different approaches. The first is a count-based approach, where the indicator is given by simply counting terms related to specific topics. And second, a keyword-in-context (KWIC) approach in which topics are defined by keywords and sentiment is extracted from a few words surrounding the keywords (Luhn, 1960).[6] I apply these two approaches to create indicators for each topic and each source.

The keywords for eleven topics (three count-based and eight KWIC-based) are selected by reading through business association reports for all available periods.[7] This is important, especially for the 19th century when language might have been very different from today and different terms described the same topics. Reading the texts makes sure I capture potentially changing language over time. The French keywords are obtained by translating the German keywords using ChatGPT and reading French texts (OpenAI, 2023a). Using these keywords, I create a new text corpus in which each document consists of a keyword and its 15 preceding and 15 following words.

The count-based indicator measures the number of documents associated with a specific topic. I use a straightforward methodology to calculate sentiment scores that resembles business cycle indicators based on firm surveys. Using existing dictionaries, I classify all words into positive, neutral, and negative words. Then, I compute the sentiment score for each document, that is, the share of positive minus the share of negative words. Finally, I create quarterly (for sources with quarterly or higher publication frequency) and yearly (for sources with lower than quarterly publication frequency) indicators for every topic and source. The count-based indicators are calculated as the sum of all identified keywords. The KWIC-based indicators are calculated as the average sentiment score over all documents.

To define positive and negative words, I follow Shapiro et al. (2022) and combine existing dictionaries that are proven to capture economic sentiment. The German lexicon combines the dictionaries developed by Remus et al. (2010) (see, e.g. Burri & Kaufmann, 2020) and a translation of the Loughran and McDonald (2011) lexicon (see, e.g. Ardia et al., 2019). The French lexicon consists of translations of the same two dictionaries and the dictionary developed by Abdaoui et al. (2017).

More formally, let $\mathcal{K}_j$ be the list of keywords for topic j, $\mathcal{P}$ the list of positive and $\mathcal{N}$ the list of negative words. Then $w_{j,i,t,d} = (w_{j,i,t,d,-15}, w_{j,i,t,d,-14}, \ldots, w_{j,i,t,d,0} \in \mathcal{K}_j, \ldots, w_{j,i,t,d,15})$ denotes the list of terms in document d at date t for topic j and source i. The count-based indicators are then calculated as

[6] Some papers also refer to this method as aspect-based sentiment analysis (see e.g Barbaglia et al., 2023).

[7] The complete list of topics with keywords chosen is available in Tables 3.A.3 and 3.A.4 the Appendix. Moreover, I provide information about the number of identified keywords over time in Figure 3.A.6 in the Appendix.

$$s_{j,i,t} = \sum_d \mathbb{1}(w_{j,i,t,d,0} \in \mathcal{K}_j) \tag{3.1}$$

The document-level sentiment score is given by

$$s_{j,i,t,d} = \frac{\sum_n \mathbb{1}(w_{j,i,t,d,n} \in \mathcal{P}) - \sum_n \mathbb{1}(w_{j,i,t,d,n} \in \mathcal{N})}{|w_{j,i,t,d}|}. \tag{3.2}$$

Figure 3.A.5 in the Appendix provides a more intuitive example of calculating the document-level sentiment score. Finally, KWIC-based indicators, $s_{j,i,t}$, for a given topic j and source i are calculated as a simple average of the sentiment scores

$$s_{j,i,t} = \frac{\sum_d s_{j,i,t,d}}{|s_{j,i,t,d}|}. \tag{3.3}$$

3.3.2 Creating topic-specific indicators

The source-level indicators suffer from several deficiencies. First, the quality of the OCR often changes over time. This is particularly relevant for the 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries, for which the archives often received the records from various sources of varying quality. This can also be interpreted as a change in the measurement error – the higher the quality of the OCR, the lower the measurement error. Second, the publication frequencies of the sources change over time – increased publication frequency means a higher information density and, therefore, lower measurement error. Third, due to technological advancements, the length of the articles and reports tends to increase over time. Finally, some indicators contain missing values, and most sources are only available for a limited period. These points potentially lead to indicators with trends and structural breaks in mean and variance. To address these issues, a six-step procedure is proposed to rectify anomalies in the data.[8] A comprehensive description of the

[8] Of course, other possibilities exist to interpolate or detrend the data. For example, Schorfheide and Song (2015) use mixed-frequency VAR models to account for varying frequencies of the underlying data. Moreover, Canova (1994, 1998) show that business cycle facts vary widely between detrending methods. Exploring these possibilities is beyond the scope of this chapter, and I leave these possibilities for future research.

procedure is provided in Section 3.C of the Appendix.

1. If the frequency is annual, interpolate missing observations using Stineman's method (Stineman, 1980). This ensures that the indicator can be temporally disaggregated to quarterly frequency in step 5.

2. Detrend the indicator using Locally Estimated Scatterplot Smoothing (LOESS) with a bandwidth of 0.7 (Cleveland, 1979).

3. Detect structural breaks in mean and variance using a binary segmentation algorithm (Killick & Eckley, 2014). Split the indicator into segments at the detected breakpoints.

4. Normalize each segment, that is subtract mean and divide by standard deviation.[9]

5. If the frequency is annual, temporally disaggregate the indicator to quarterly frequency (Dagum & Cholette, 2006).

6. Remove outliers (observations more than three standard deviations away from the mean).

Finally, I aggregate the corrected source-level indicators into an overall indicator for each topic. To effectively summarize the information content of the data and eliminate idiosyncratic noise, I estimate a factor model in static form. Therefore, the source-level indicators follow

$$s_{j,i,t} = \lambda_{j,i} f_{j,t} + e_{j,i,t} \tag{3.4}$$

where $s_{j,i,t}$ is the corrected source-level indicator for topic j and source i, $\lambda_{j,i}$ is the factor loading for source i and topic j, $f_{j,t}$ is the common factor for topic j at time t, and $e_{j,i,t}$ is the idiosyncratic component. Finally, I use the first principal component as the topic-specific indicator, $S_{j,t}$ for topic j.

The advantage of using a factor model is that it allows for summarizing the information in a large data matrix with a small number of common factors. As in Chapter 1, factors and loadings can be estimated through principal components, under the assumption

[9] I show in Appendix 3.D that with time-varying measurement error, this is preferable to not normalizing.

that the idiosyncratic components are only weakly serially and cross-sectionally correlated (Bai & Ng, 2013; Stock & Watson, 2002).[10]

3.3.3 Estimation of composite indicator

The final step is to combine the topic-specific indicators into one composite indicator for the business cycle. For this, I fit the topic-specific indicators to a measure of the business cycle on a very recent sample where it is widely acknowledged that the business cycle is captured well. Then, I use the estimated coefficients to backcast the business cycle. As the measure for the business cycle, I use the output gap.[11] The topic-specific indicators, as well as four lags and four leads, are used as predictors.

Following Bybee et al. (2023), shrinkage methods are employed to estimate the coefficients. Specifically, I use elastic net regression. A primary strength of an elastic net model lies in its ability to handle collinearity effectively. Traditional regression methods can struggle to provide reliable coefficient estimates in such cases. The elastic net, however, combines L1 (Lasso) and L2 (Ridge) regularization techniques, creating a balance that addresses collinearity and encourages sparsity in the model. This means it can automatically select a subset of important predictors while shrinking the coefficients of less important ones toward zero. By doing so, the elastic net enhances the interpretability of the model and improves its predictive performance by reducing overfitting. Before estimating the model, I normalize all variables. The objective function of the elastic net is given by

$$\text{Minimize:} \quad \frac{1}{2T_{Train}} \sum_{t=1990Q1}^{T} (y_t - S'_{j,t}\beta)^2 + \lambda \sum_{k=1}^{p} \left(\alpha|\beta_k| + \frac{1-\alpha}{2}\beta_k^2 \right) \tag{3.5}$$

where T_{Train} is the number of observations in the training set, T is the date of the last observation in the data, p the number of predictors, y_t is the output gap, β is the vector of coefficients, $S_{j,t}$ is the vector of topic-specific indicators, λ is the penalty parameter,

[10] I interpolate missing values using an EM algorithm (Stock & Watson, 2002), after standardizing the data to have zero mean and unit variance. For interpolating the data, I use one factor ($r = 1$) as there were converging issues with $r > 1$. Finally, I use the first principal component of the interpolated data set. Using least squares instead of the EM algorithm, I estimate the factor model with missing values as a robustness test. Even though the method is simpler than the EM method, it struggled with the sparsity of the data.

[11] To create a quarterly output gap, I use Hamilton (2018) filtered real GDP estimates from the State Secretariat for Economic Affairs (SECO). I also use estimates of the output gap provided by the SECO and the SNB for robustness. The results are very similar.

Figure 3.2 — Quarterly output gap split into training and test set

Notes: The red line represents quarterly output gap estimated using the method proposed by Hamilton (2018). The training set is used to estimate elastic net models with different combinations of hyperparameters. The gray shaded area is the test set used to evaluate the models and select the best combination of hyperparameters. The blue line is the fitted output gap using the best model.

and α is the mixing parameter. The mixing parameter, α, controls the relative weight of the L1 and L2 penalties. When $\alpha = 0$, the penalty is an L2 penalty (ridge regression); when $\alpha = 1$, it is an L1 penalty (lasso regression). The elastic net penalty is a convex combination of the L1 and L2 penalties. Figure 3.2 illustrates how optimal values of λ and α are chosen. I perform a grid search over a range of λ and α values, estimate a model on a training set, and select the combination that minimizes the mean squared error (MSE) on a test set. The test set spans 1980 - 1989, and the training set 1990-2022. Using the best-performing model, the composite indicator is calculated as the fitted output gap from 1821 to 2021.

$$S_t = S'_{j,t}\hat{\beta}. \tag{3.6}$$

The blue line in Figure 3.2 shows the model fit on the training set and the prediction on the test set. The model performs reasonably well – the blue line lies close to the red line, which represents the actual output gap.

3.4 Evaluation of the indicators

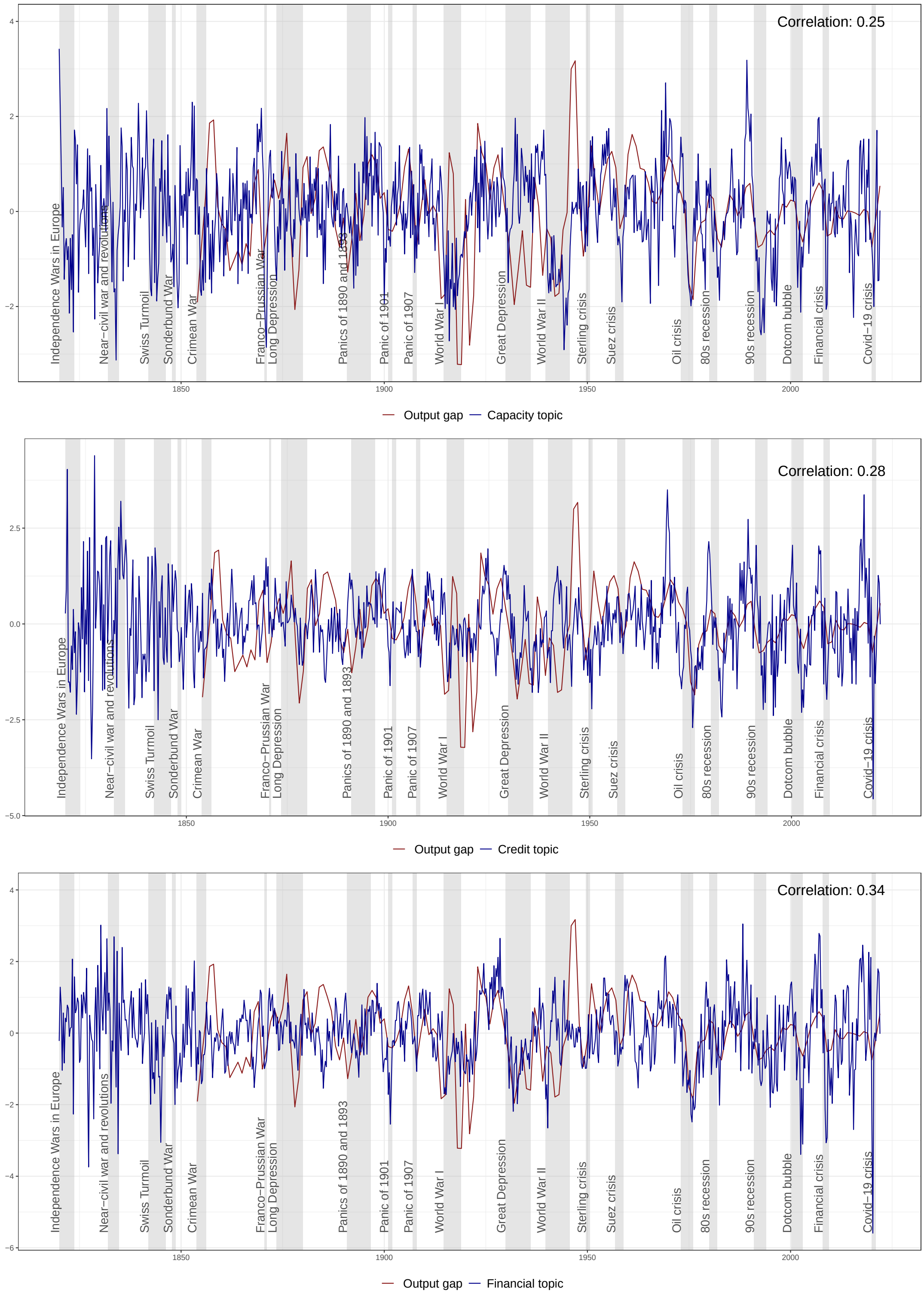
In this section, I evaluate the business cycle indicator. Because reliable data is missing in the 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries, I rely on descriptive analysis and compare downswings in the indicator to narratives of economic crises and wars. First, I present the sentiment-based and count-based topic-specific indicators. Second, I show the composite indicator and discuss which topic-specific indicators contribute the most to the indicator. Then, I show the results from a correlation analysis. Finally, I present a business cycle chronology for Switzerland in the 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries.

3.4.1 Characteristics of topic-specific indicators

The sentiment-based topic-specific indicators are depicted in Figure 3.3. Gray-shaded areas highlight wars, crises, and recessions related to the Swiss or the global economy. Two observations stand out. First, the indicators are highly correlated with each other. This is unsurprising as the topics are related to the business cycle and, therefore, should move together. The correlation matrix in Table 3.A.7 in the Appendix confirms this. These correlations underline the importance of using a model that can handle collinearity, like the elastic net, to combine the indicators.

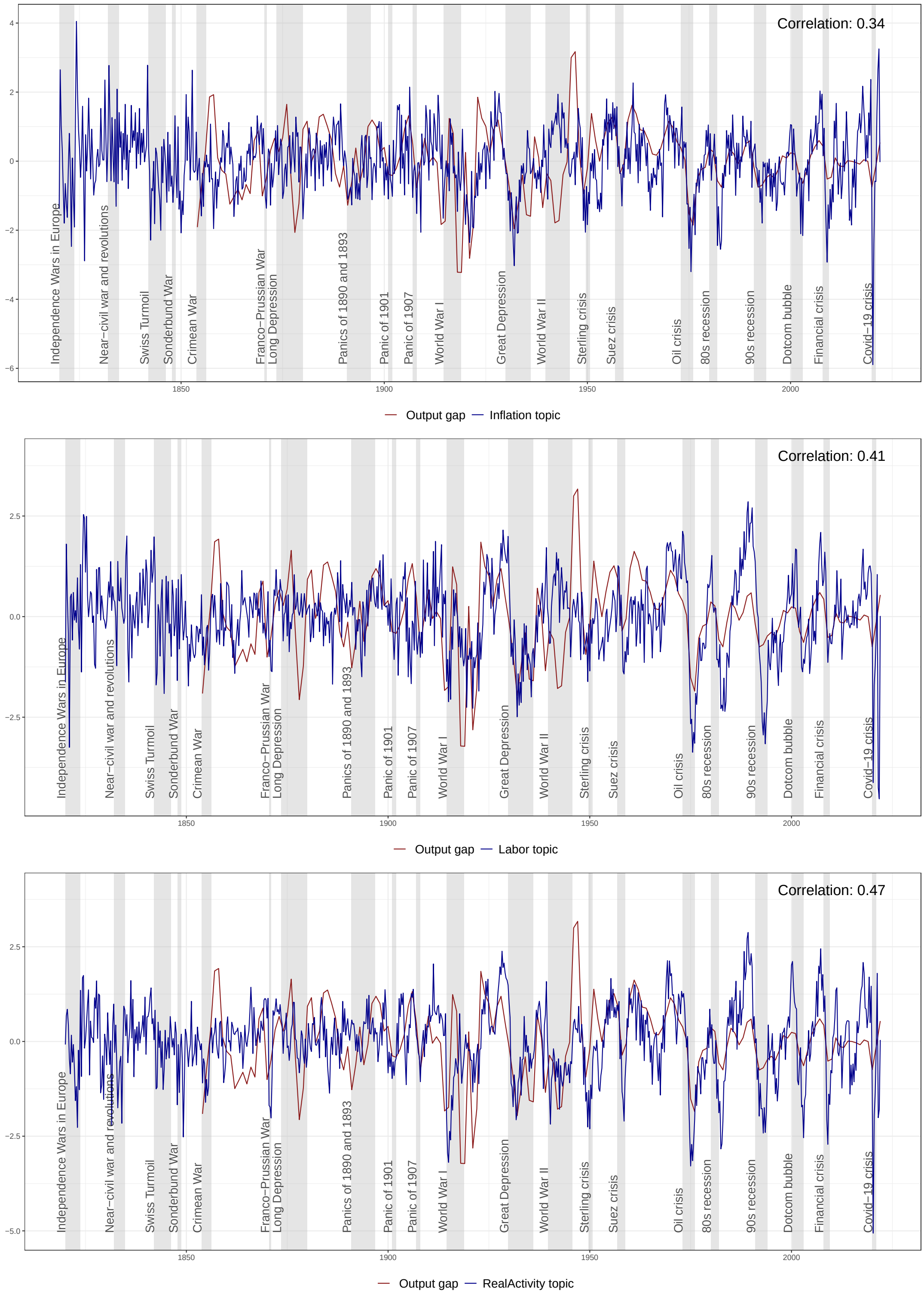
Second, the variance of the indicators seems to increase over time. This may seem counterintuitive given that the overall economy, including GDP growth, exhibited reduced volatility during the Great Moderation period. The quarterly Net Economic Sentiment for the US developed by Van Binsbergen et al. (2024) covering 1850 - 2020 shows a similar increase in volatility over time. However, they do not discuss this behavior in their paper. Nevertheless, there are possible explanations for it. First, it may be related to increased media sensationalism. This refers to the practice of presenting news stories or events in a way that exaggerates their importance or sensationalizes aspects of the story to attract attention and generate public interest. Media sensationalism might have intensified over the past century due to technological, economic, and social factors. Second, the data from earlier periods may contain higher measurement errors, leading to lower volatility by construction. In practical terms, when normalizing the identified segments of the source-level indicators, a higher noise-to-signal ratio in the early data can lead to a more pronounced signal reduction during normalization. This results in a dampened signal in the aggregated indicator,

Figure 3.3 — Sentiment-based topic-specific indicators



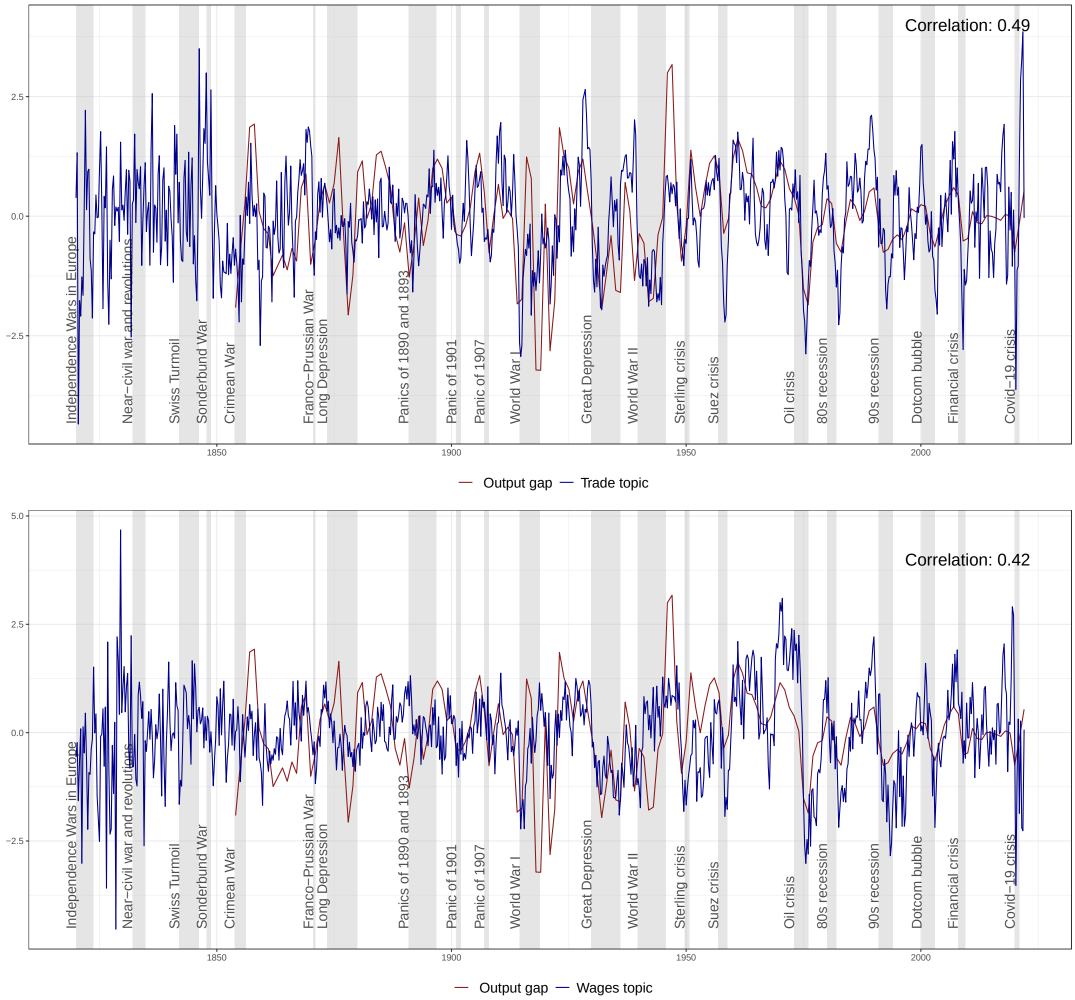
Notes: These graphs show sentiment-based topic-specific indicators and the output gap. Gray-shaded areas highlight wars, crises, and recessions related to the Swiss or the global economy.

Figure 3.3 — Sentiment-based topic-specific indicators, continued from previous page



Notes: These graphs show sentiment-based topic-specific indicators and the output gap. Gray-shaded areas highlight wars, crises, and recessions related to the Swiss or the global economy.

Figure 3.3 — Sentiment-based topic-specific indicators, continued from previous page

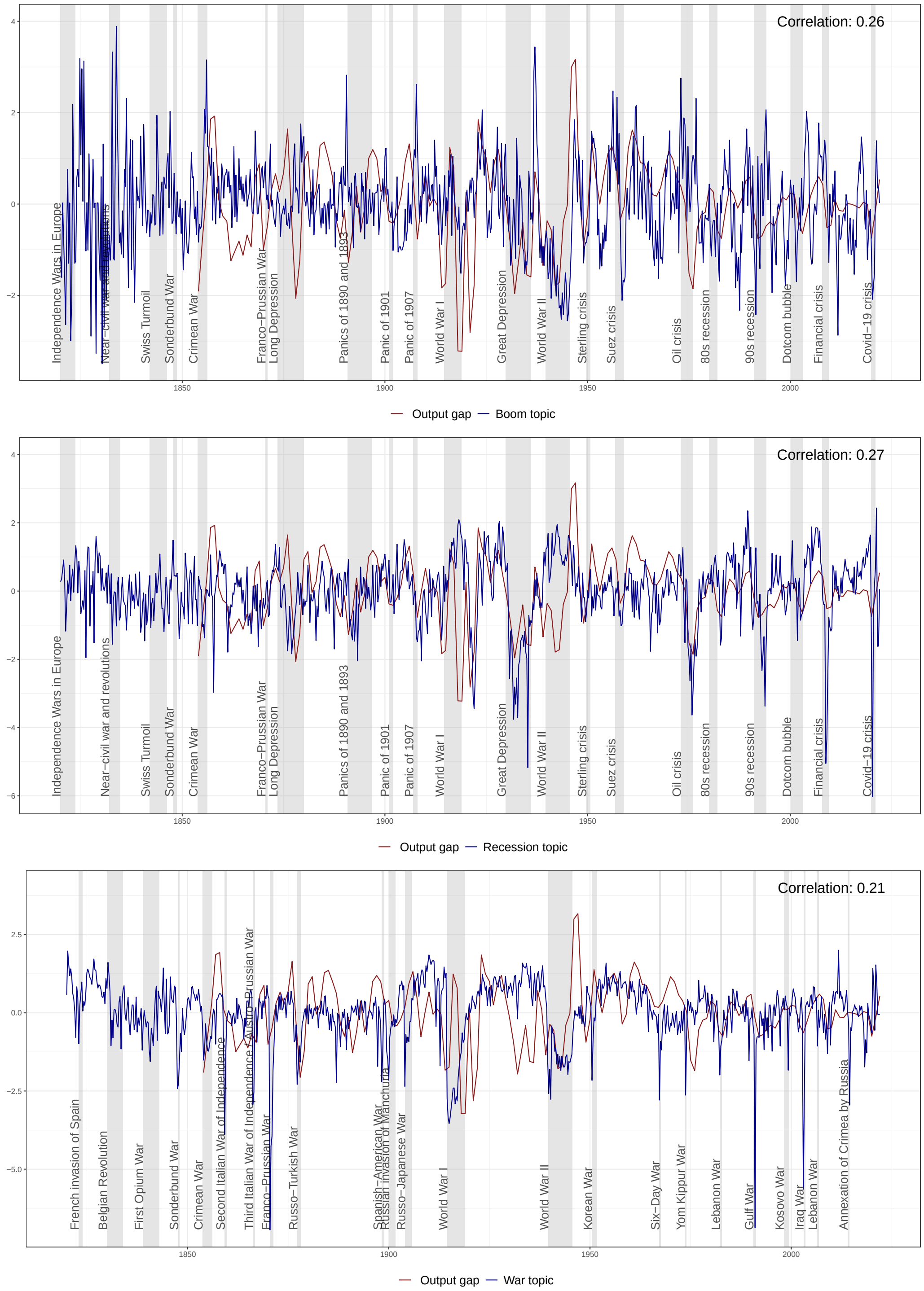


Notes: These graphs show sentiment-based topic-specific indicators and the output gap. Gray-shaded areas highlight wars, crises, and recessions related to the Swiss or the global economy.

where some noise is effectively canceled. Appendix 3.D provides a more detailed explanation of this phenomenon.

The count-based topic-specific indicators are depicted in Figure 3.4. They are scaled to have a positive correlation with the business cycle. Overall, the coincident correlations of these indicators are weaker than those of the sentiment-based indicators. Moreover, the increase in volatility over time is not present for the Boom and War indicators. For the Recession indicator, it is there. Although I also use other keywords like "crisis" for this indicator, this might be explained because the word "recession" has been less used

Figure 3.4 — Count-based topic-specific indicators

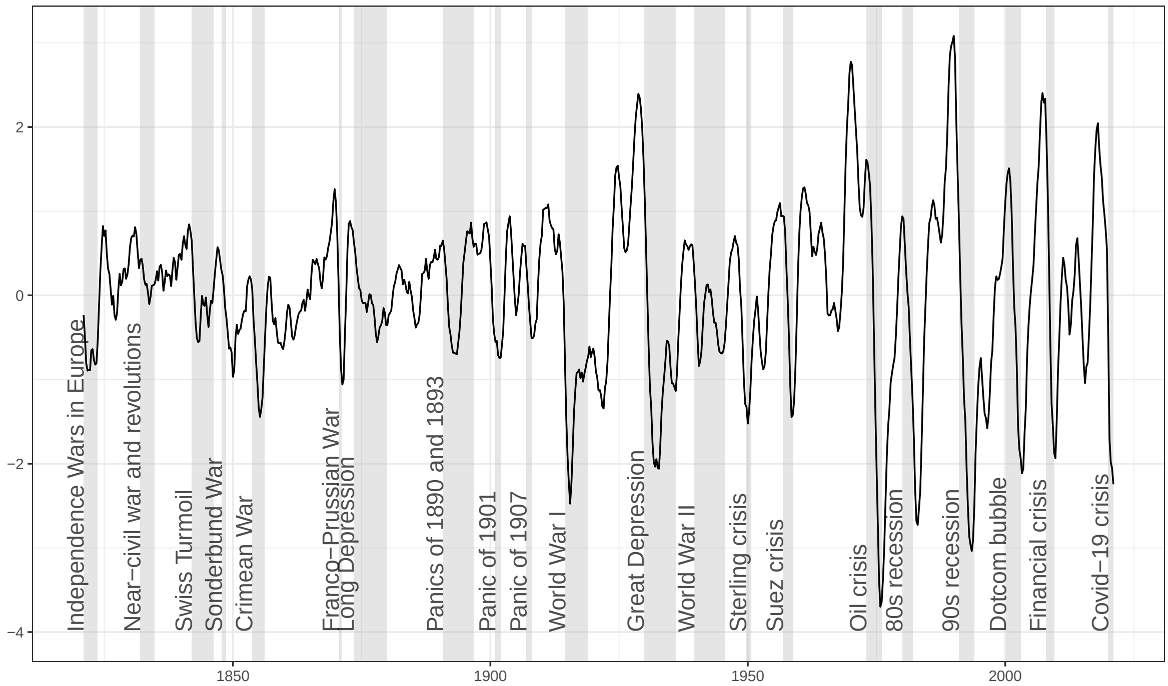


Notes: These graphs show count-based topic-specific indicators and the output gap. Indicators are scaled to have a positive correlation with the business cycle. Gray-shaded areas highlight wars, crises, and recessions related to the Swiss or the global economy.

(or did not exist in this context) in the earlier periods.[12] Gray-shaded areas in the graph for the war indicator highlight important wars. The indicator captures most of these combat actions well.

3.4.2 Characteristics of composite indicator

Figure 3.5 — Historical business cycle indicator



Notes: The black line represents the composite historical business cycle indicator. Gray-shaded areas indicate global and Swiss crises, recessions, and wars.

Figure 3.5 depicts the composite indicator with the black solid line. A few observations stand out. First, the indicator tracks economic downswings, given by the gray-shaded areas, very well.[13] It captures all the major recessions the Swiss economy faced in the last decades, like the Oil crisis from 1974, the recessions from the beginning of the 80s and 90s, the Great Financial Crisis (GFC) and the Covid–19 crisis. Moreover, the indicator responds to the two World Wars and the Great Depression. It is also able to capture crises in the 19th century like the Sonderbund War from 1847, the Franco-Prussian War from 1870, the Long Depression from 1873, and the Panic from 1890 (also known as

[12]According to the Oxford English Dictionary, the word "recession" was only increasingly used in the second half of the 20th century.

[13]For Switzerland, no official business cycle dating like from the NBER for the US exists. Therefore, these downswings are based on national and international narratives of crises and wars (see, e.g. Church & Head, 2013)

Table 3.2 — Ten most important predictors

Predictor	Coefficient
Wages	0.0212
Labor	0.0211
Capacity lag 3	0.0203
Capacity lag 2	0.0194
Trade lag 3	0.0193
Capacity lag 1	0.0186
Labor lead 1	0.0180
Real activity	0.0178
Inflation	0.0177
Capacity	0.0169

Notes: The table shows the ten most important predictors (highest coefficients in absolute values) from the elastic net regression.

the Baring crisis).[14] The indicator also captures Swiss-specific turmoil from 1842 as Switzerland experienced significant tensions between Liberals and Catholics. Second, as the underlying indicators, the composite indicator also gets more volatile over time. This is unsurprising as the indicator is a linear combination of the topic-specific indicators.

Which topic-specific indicators contribute the most to the composite indicator? To answer this question, Table 3.2 shows the ten most important predictors (highest coefficients in absolute values) from the elastic net regression. The most important predictors are the indicators for Wages, coincident and leading Labor, lagged Capacity utilization, lagged Trade, as well as Real activity and Inflation. Because Capacity utilization is directly related to the output gap, it is not surprising that it is one of the most important predictors. The count-based indicators seem less important.

Figure 3.6 shows the rolling correlations between the indicator and a selected set of variables with a fixed window size of 20 years. By construction, the indicator is highly correlated with the output gap, with an overall correlation of around 0.5. The correlation is highest during the late 20$^{\text{th}}$ and 21$^{\text{st}}$ centuries, with a correlation coefficient of around

[14]The Sonderbund War was a civil conflict in Switzerland. It happened because seven Catholic cantons formed the Sonderbund in 1845 to protect their interests from a centralization of power. The war ended with the Sonderbund's defeat. This led to Switzerland becoming a federal state and marked the end of a period of political change in the country (Church & Head, 2013). The Franco-Prussian War was a conflict between the Second French Empire and the North German Confederation. As a neutral state, Switzerland was not directly involved in the conflict. However, due to its proximity, the war significantly impacted the Swiss economy (Church & Head, 2013). The Baring crisis was a financial crisis that occurred in 1890. It was caused by the near-collapse of Barings Bank, one of the most important financial institutions in the world at the time. The crisis had a significant impact on the global economy.

0.9. The correlation is lower during the early 20^{th} century but with around 0.5 still substantial. There is still some correlation for the 19^{th} century, but it is not statistically different from zero. A similar picture emerges for labor market data, trade, and real and nominal GDP growth. The correlation is high during the late 20^{th} and 21^{st} centuries and lower for the 19^{th} and early 20^{th} centuries. The lower correlation in the early sample does not necessarily imply that the indicator is less accurate. It may be due to higher measurement errors in 19^{th}-century data, which pushes the coefficients towards zero.

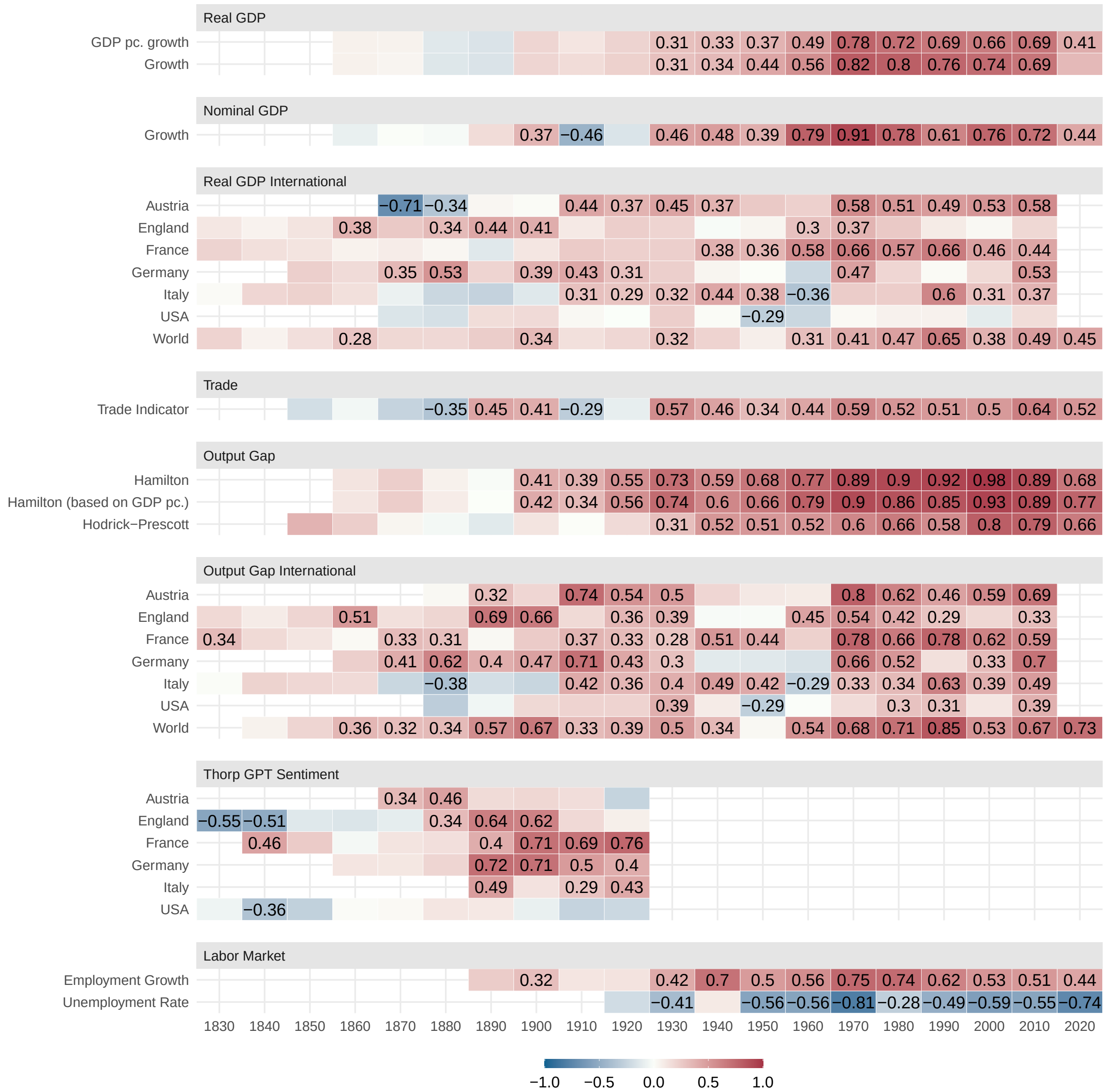
In the late 20^{th} and early 21^{st} centuries, the indicator shows high correlations between the output gap and real GDP growth of neighboring countries. This correlation is less pronounced in the USA and England yet remains substantial. In contrast, during the 19^{th} century, the correlation diminishes for all countries but maintains statistical significance in France, Germany, and England. This pattern indicates a higher synchronization of the Swiss business cycle with international trends in the 20^{th} century compared to the 19^{th} century. These observations align with the findings of Broadberry and Lennard (2023), who report increasing synchronization of business cycles over time. Specifically, they find that 25% of the potential correlations between European countries' GDP growth were significantly positive from 1870 to 1950. This proportion rose to 66% from 1950 to 2000. However, these findings could also be driven by measurement errors.

The broad picture is confirmed by the correlations with the sentiment indicators based on Thorp's (1926) description of the state of the economy. These indicators were generated using OpenAI's (2023b) GPT-3.5, as no sentiment indicators exist for the 19^{th} century.[15] For all countries except the USA, there is a substantial correlation with the business cycle indicator. The strongest correlations are observed for Germany, France, and England.

Before 1850 there are a few significant positive correlations with variables from France. This makes sense since Switzerland was trading a lot with its larger neighbor sharing a common language. However, Switzerland was also highly dependent on southern Germany. However, there are no positive significant correlations with variables from other countries. This divergence could be explained by unique internal political and social conflicts following the Congress of Vienna in 1815. The period until the foundation of modern Switzerland in 1848 was characterized by pronounced divisions between liberal and conservative cantons, culminating in the Sonderbund War of

[15] Section 3.B.2 in the Appendix provides a detailed description of the methodology to create the Thorp GPT Sentiment indicators.

Figure 3.6 — Rolling correlations with 20 year window



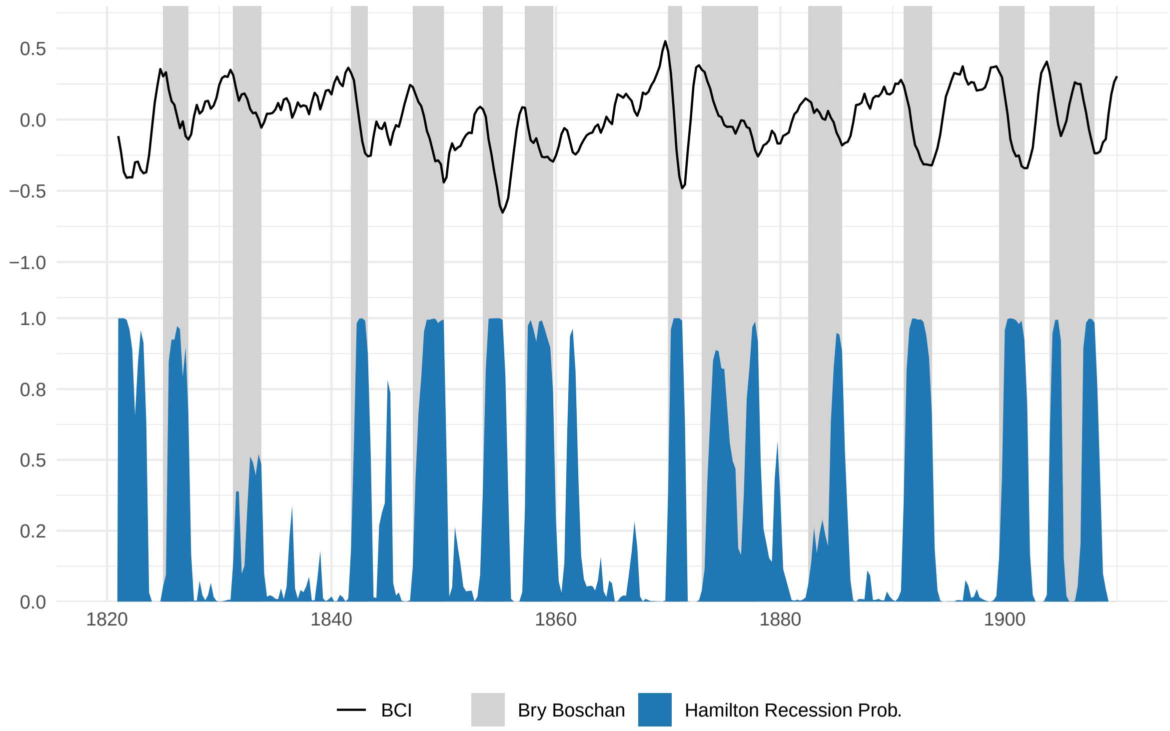
Notes: The graph shows rolling coincident correlations with a fixed window size of 20 years between the indicator on the left and the composite indicator. The period considered is given by the year on the x-axis plus and minus ten years. Only statistically significant (on a 10% level) correlations are labeled. Correlations with fewer than ten observations are not shown.

1847. These internal disputes, centered on governance, federal authority, and religious differences, fostered a climate of economic uncertainty and a focus on domestic concerns. Consequently, economic sentiment in Switzerland might have been insulated from broader European trends during this period (see, e.g. Church & Head, 2013).

3.4.3 Business cycle dating

It is generally acknowledged that the modern business cycle features long expansions combined with short recessions (see, e.g. Romer & Romer, 2020). Due to data limitations, systematic data-driven analysis of business cycles could, until recently, be only conducted on very modern data (Broadberry & Lennard, 2023). However, with the progress in the quantification of economic activity, Broadberry and Lennard (2023) analyzed the nature of business cycles of nine European countries. Their main finding is that recessions got less frequent and shorter over time.

Figure 3.7 — Indicator and recession classification 1820 - 1910

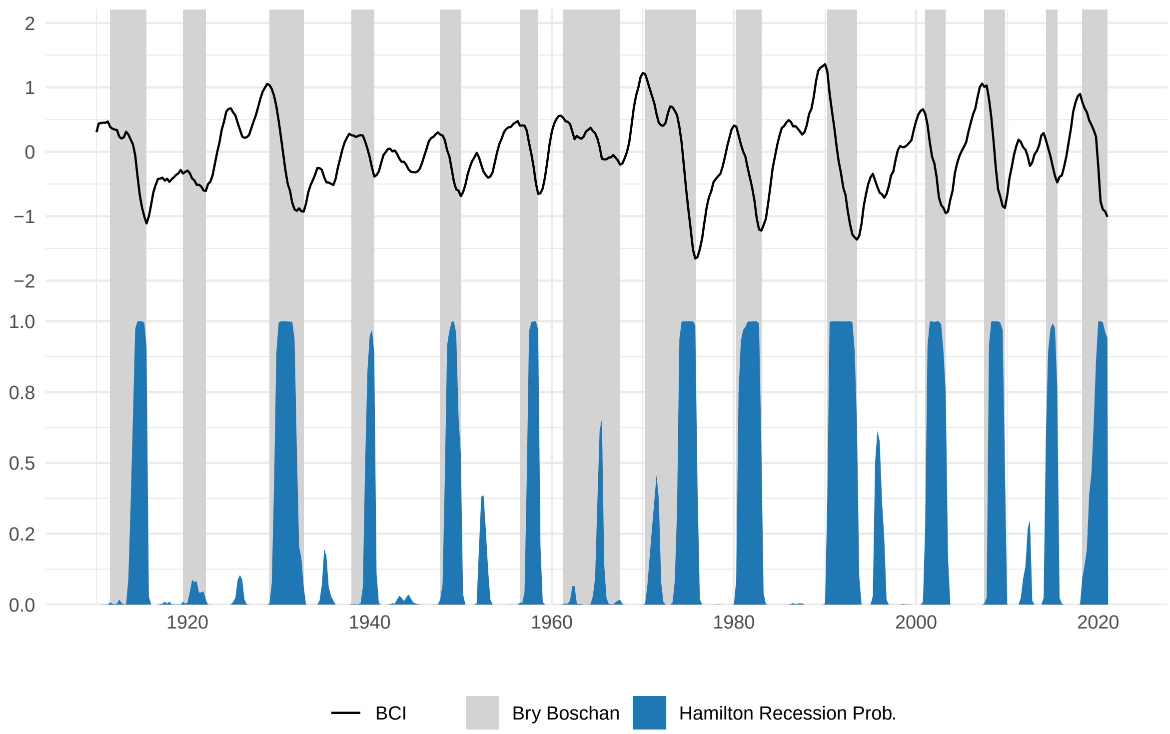


Notes: The graph shows the composite indicator given by the black solid line together with smoothed recession probabilities obtained by Hamilton's (1989) Markow-Switching autoregression model. The gray shaded areas indicate recessions obtained by an adapted Bry and Boschan (1971) algorithm.

How do these findings compare to the Swiss business cycle? For the 19th and early 20th centuries, there is, to the best of my knowledge, no business cycle chronology available. The Economic Cycle Research Institute (ECRI) provides a chronology of business cycle turning points for Switzerland starting in 1956. The ECRI chronology is more judgemental, considering several indicators, including output, employment, income, and sales (Glocker & Wegmueller, 2020). Moreover, Siliverstovs (2011) uses an approach based on Markov-switching models, and Glocker and Wegmueller (2020) use a Markov-switching dynamic factor model (MS-DFM) to date business cycle turning points for Switzerland starting in 1980. Both identify turning points that are broadly consistent with those determined by the ECRI. Therefore, I mainly use the ECRI chronology for comparison.

To address the lack of a business cycle chronology for Switzerland during the 19th and early 20th centuries, this section utilizes the developed business cycle indicator to identify and date business cycle turning points. As Romer and Romer (2020) explain, recessions are not simply random categorizations of macroeconomic outcomes. Instead, they represent critical macroeconomic moments characterized by a rapid and significant deviation of economic activity from its normal state. Establishing a business cycle chronology for the 19th and early 20th centuries is thus crucial for understanding these economic shifts.

Figure 3.8 — Indicator and recession classification 1910 - 2021



Notes: The graph shows the composite indicator given by the black solid line together with smoothed recession probabilities obtained by Hamilton's (1989) Markow-Switching autoregression model. The gray shaded areas indicate recessions obtained by an adapted Bry and Boschan (1971) algorithm.

I use the Markov-Switching autoregression model proposed by Hamilton (1989) to estimate recession probabilities. This aligns with Romer and Romer (2020) who show that "recession periods emerge clearly from a Markov-switching model". The model is given by

$$y_t = \mu_{s_t} + \phi y_{t-1} + \varepsilon_t \tag{3.7}$$

with $\varepsilon_t \sim N(0, \sigma^2)$ and where s_t is the realization of a two-state Markov chain with

$$\Pr(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots, y_{t-1}, y_{t-2}, \ldots) = \Pr(s_t = j | s_{t-1} = i) = p_{ij} \tag{3.8}$$

where $i, j = 0, 1$. Within this framework, one can label s_t = 0 and s_t = 1 as the expansion and recession states at time t. I estimate the model on two samples because the indicator is less volatile in the 19th century. First, from 1820 to 1910, and second, from 1910 to 2021.[16]

For robustness, I also use the nonparametric algorithm proposed by Harding and Pagan (2002) to date business cycle turning points. The algorithm is a quarterly adoption of Bry and Boschan's (1971) algorithm for monthly data. However, since the algorithm is designed for data in levels, I increased the number of quarters within which a local maximum or minimum has to occur to 10 quarters.[17]

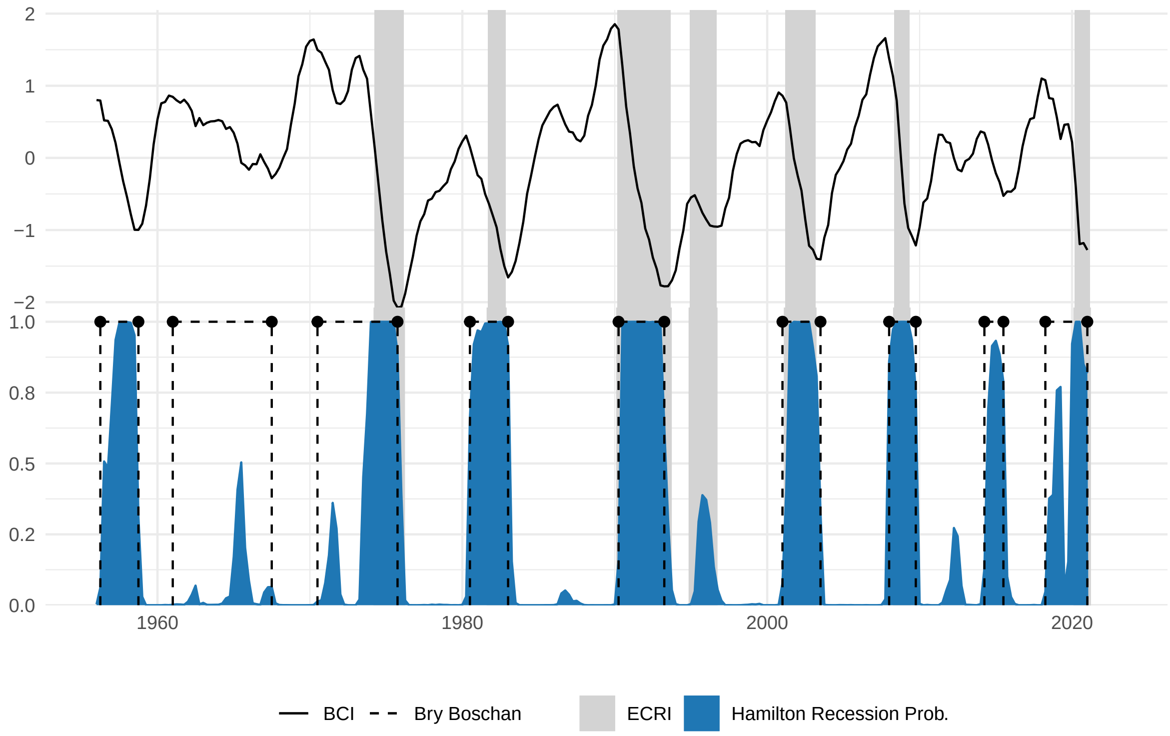
An issue when utilizing detrended data arises because it tends to categorize so-called growth recessions as contractions. In the literature, these are not generally recognized as actual recessions (Broadberry & Lennard, 2023). To address this, I refine the identified recession periods, ensuring that their duration aligns with negative values in the business cycle indicator. The censored business cycle dating is very similar to the original one. All recessionary episodes identified using the Markov-switching model are listed in the Appendix in table 3.A.8.

The results of the estimated recessionary episodes are depicted in Figures 3.7 and 3.8.

[16] Estimating the model on different samples is not uncommon. For example, Romer and Romer (2020) estimate the model with data from 1929 to 1947 and 1948 to 2019.

[17] This aligns with the general understanding that business cycles can vary but often fall within a 2 to 10-year range. See Harding and Pagan (2002) for more details on the algorithm.

Figure 3.9 — Comparison with ECRI dating



Notes: The graph shows the composite indicator given by the black solid line together with smoothed recession probabilities obtained by Hamilton's (1989) Markow-Switching autoregression model. The black dashed lines indicate recessions obtained by an adapted Bry and Boschan (1971) algorithm. The gray-shaded areas are recession dates obtained from ECRI.

I associate values of the regime probability above 0.5 with a recession.[18] The two regimes identified appear to correspond with phases of recessions and expansions, with the model's high recession probability periods showing considerable alignment with the recessions delineated by the Bry-Boschan algorithm. However, the Bry-Boschan procedure tends to overestimate recessions' duration.

Figure 3.9 compares my classification and the recession episodes identified by the ECRI. There is a significant overlap between the recessionary regime probabilities (illustrated in blue) and the ECRI's recession dating (indicated by the gray-shaded area). The Hamilton model tends to classify recessions more often. It additionally identifies with high probability recessions in 1958, known as the Eisenhower recession, and in 2015, coinciding with the SNB's termination of the minimum exchange rate. Additionally,

[18] The recession threshold of 0.5 is arbitrary and chosen to be in line with Hamilton (1989). However, using different thresholds like, e.g., 0.8 proposed by Romer and Romer (2020) would not change the results significantly because the estimated probability is typically close to either 0 or 1.

Table 3.3 — Frequency, duration, and number of Swiss recessions

	Algorithm	Recession (Peak to trough) 1820-1910	1911-1950	1951-2022
Frequency	Hamilton (1989)	32.5	17.5	25.3
	Bry and Boschan (1971)	50.6	40.0	38.8
	BL2023 (Swiss GDP)	25.4	45.0	15.3
	ECRI			19.3
	OECD			39.0
	Technical recession			15.7
Duration	Hamilton (1989)	6.5	7.0	7.1
	Bry and Boschan (1971)	14.0	12.8	12.1
	BL2023 (Swiss GDP)	5.0	8.0	6.3
	ECRI			7.4
	OECD			22.5
	Technical recession			2.7
Number	Hamilton (1989)	18.0	4.0	10.0
	Bry and Boschan (1971)	13.0	5.0	9.0
	BL2023 (Swiss GDP)	12.0	9.0	7.0
	ECRI			7.0
	OECD			13.0
	Technical recession			10.0

Notes: Frequency is the share in percent of quarters in a given phase. Duration is the average number of quarters in a given phase. The ECRI dating starts in 1956, OECD dating starts in 1960, and quarterly GDP to calculate technical recessions is available from 1980. Broadberry and Lennard (2023) (BL2023 Swiss GDP) classify recessions as negative real annual GDP growth. Annual real Swiss GDP growth is available from 1852.

the model suggests with less certainty another brief recessionary period in the 1960s.[19] The fact that we can attribute the additional recessions to well-known economic events suggests that the Markov-switching model identifies recessions more reliably than the existing ECRI classification.

How does the modern Swiss business cycle compare to the business cycle in the 19th century? Tables 3.3 and 3.4 show the frequency, duration, and number of identified recessions and expansions for periods 1820 to 1910, 1911 to 1950, and 1951 to 2022. Recessions in the 19th century were more frequent than in the 20th and 21st centuries.

[19] I also compare my classification to the recession dates provided by the Organisation for Economic Co-operation and Development (OECD), recessionary episodes based on the technical definition for recessions (two consecutive quarters of negative GDP growth), and the definition by Broadberry and Lennard (2023) - classifying periods with negative annual GDP growth as recessions. The classifications show a high degree of overlap and are depicted in Figures 3.A.7, 3.A.8 and 3.A.9 in the Appendix.

Table 3.4 — Frequency, duration, and number of Swiss expansions

	Algorithm	Expansion (Trough to peak) 1820-1910	1911-1950	1951-2022
Frequency	Hamilton (1989)	67.5	82.5	74.7
	Bry and Boschan (1971)	49.4	60.0	61.2
	BL2023 (Swiss GDP)	74.6	55.0	84.7
	ECRI			80.7
	OECD			61.0
	Technical recession			84.3
Duration	Hamilton (1989)	13.8	34.5	21.3
	Bry and Boschan (1971)	14.1	20.4	19.4
	BL2023 (Swiss GDP)	13.5	10.5	34.3
	ECRI			29.9
	OECD			33.2
	Technical recession			13.5
Number	Hamilton (1989)	18.0	4.0	10.0
	Bry and Boschan (1971)	13.0	5.0	9.0
	BL2023 (Swiss GDP)	12.0	9.0	7.0
	ECRI			7.0
	OECD			13.0
	Technical recession			10.0

Notes: Frequency is the share in percent of quarters in a given phase. Duration is the average number of quarters in a given phase. The ECRI dating starts in 1956, OECD dating starts in 1960, and quarterly GDP to calculate technical recessions is available from 1980. Broadberry and Lennard (2023) (BL2023 Swiss GDP) classify recessions as negative real annual GDP growth. Annual real Swiss GDP growth is available from 1852.

The share of quarters in recession decreases no matter which classification is used. However, the duration of recessions does not change significantly over the periods considered. Expansions, in contrast, get longer over time. The average duration of expansions increases from 13.8 quarters in the period up to 1910 to 21.3 quarters since 1951. Tables 3.A.5 and 3.A.6 in the Appendix compare these numbers to the European (Broadberry & Lennard, 2023) and the British business cycle (Broadberry et al., 2023). The results on the frequency and the expansions are in line with the findings on the European business cycle, albeit less pronounced. However, the duration of recessions is not. Broadberry and Lennard (2023) find that the duration of recessions in Europe decreases over time. On the other hand, Broadberry et al. (2023) confirm that the duration of recessions in the UK stays constant over time. Therefore, the widespread belief that recessions got shorter is not fully supported by the data.

3.5 Robustness

In this section, I perform a series of robustness checks. In particular, I show that the keyword-based algorithm delivers similar results as a rating based on OpenAI's (2023b) LLM GPT-3.5. Moreover, I analyze the sensitivity of the indicator concerning different aggregation techniques and different subsets of the data.

3.5.1 Keyword-based algorithm

Given the emergence of LLMs, the question of why these models are not employed for extracting information from the text corpus is naturally raised. There are two main reasons. First, the cost of utilizing these models is a significant factor. For instance, OpenAI's (2023b) GPT-3.5 model incurs a charge of 0.001 USD per 1000 tokens[20]. With a conservative estimate of the average text length being only 20 tokens, processing 100 million texts would approximate a total expense of around 2 million USD. Second, it would take significant time to process all the texts. Nevertheless, it would be interesting to compare the keyword-based algorithm with LLMs for future research.

Due to the high costs associated with LLMs, their use for analyzing my text corpus is not feasible. Nonetheless, to ascertain the efficacy of the keyword-based algorithm, I also apply the algorithm to Thorp's (1926) work on business cycles (i.e., his descriptions of the state of the economy). I then compare the indicators created this way to those derived with GPT-3.5 as described in section 3.B.2 in the Appendix. To be precise, GPT-3.5-based indicators use the texts in English. For the keyword-based algorithm, the English texts are first translated to German and French.[21]

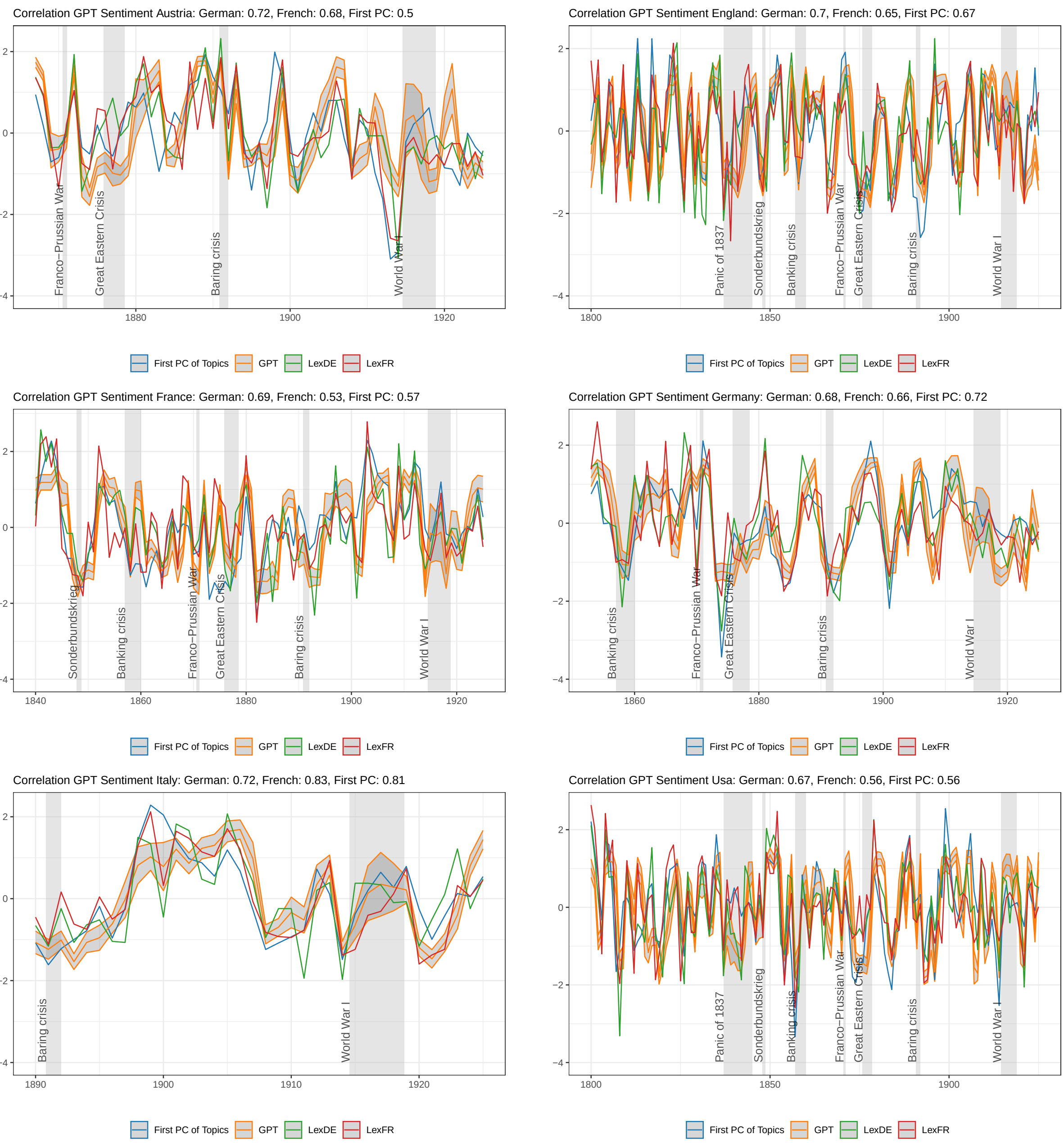
As shown in Figure 3.10, the correlations between the GPT-3.5-based and keyword-based indicators range between 0.5 and 0.8 for the countries under consideration. This suggests that the keyword-based algorithm delivers similar results as LLMs. Whether one or the other method is superior is beyond the scope of this study and should be addressed in future research. However, the results suggest that the keyword-based algorithm is a viable alternative to LLMs for constructing business cycle indicators.

3.5.2 Aggregation techniques

There are countless possibilities for constructing a business cycle indicator from textual data. I compare the baseline indicator to several alternatives to check that it is not sensitive to the aggregation technique.

[20]This corresponds to approximately 750 words.
[21]To translate the texts, I use ChatGPT (OpenAI, 2023a).

Figure 3.10 — Comparison of lexicons and methods



Notes: These graphs show several text-based sentiment indicators based on Thorp's (1926) description. In orange, the normalized sentiment indicator based on OpenAI's (2023b) GPT-3.5 model and one standard deviation confidence bands. In green, the sentiment indicator based on a German translation of Thorp (1926) and the German lexicon. In red, the sentiment indicator based on a French translation of Thorp (1926) and the French lexicon. In blue, the first principal component from indicators based on both translations and the keyword-based algorithm outlined in section 3.3. Gray-shaded areas represent crises.

Other aggregation methods to composite indicator. Instead of using an elastic net model to aggregate the topic-specific indicators into the composite indicator, I use a

- Weighted average based on keyword importance in each topic:

$$S_t = \sum_{j=1}^{P} \omega_{j,t} s_{j,t} \tag{3.9}$$

  where S_t denotes the composite indicator in quarter t, P is the total number of topics. $\omega_{j,t}$ and $s_{j,t}$ represent the weight derived by the number of identified keywords and the topic-specific indicator, respectively, for topic j in quarter t.

- Markov-Switching Dynamic Factor Model (MS-DFM) brought forward by Kim (1994), Diebold and Rudebusch (1996), Kim and Yoo (1995) and Chauvet (1998).[22] In these models, an unobservable regime-switching variable governs the common component.[23] The model comprises one factor (based on screeplot in Figure 3.A.10) and five lags (based on Bayesian Information Criterion (BIC)).[24] Using an MS-DFM allows to check the sensitivity of the indicator and the sensitivity of the business cycle dating.

Other aggregation methods to topic-specific indicators. Instead of using a static factor model to aggregate the sources to a topic-specific indicator, I

- use a weighted average based on keyword importance in each source:

$$S_{j,t} = \sum_{i=1}^{N_t} \omega_{j,i,t} s_{j,i,t} \tag{3.10}$$

  where $S_{j,t}$ denotes the topic-specific indicators. N_t is the total number of sources available in quarter t. $\omega_{j,i,t}$ and $s_{j,i,t}$ represent the weight derived by the number of identified keywords and the source-level indicator, respectively, for a specific topic j and source i in quarter t.

[22] I thank Philipp Wegmüller for providing his programs to estimate MS-DFMs.
[23] Camacho et al. (2015) show that performing the estimation in one step is superior to estimating a Markov-switching process on the factor in a sequential step.
[24] See section 3.E in the Appendix for more details.

- use a fixed effects regression model à la Shapiro et al. (2022). I estimate the quarter fixed effects and use these as the indicator ($S_{j,t} = \hat{f}_{j,t}$) from the following regression

$$s_{j,i,t} = f_{j,t} + f_{j,i} + f_{j,l} + f_{j,f} + f_{j,i} \times f_{j,p} + \varepsilon_{j,i,t} \tag{3.11}$$

  where $s_{j,i,t}$ is the sentiment score in quarter t for a specific topic j and source i. $f_{j,t}$ is a sample quarter fixed effect, $f_{j,i}$ a source fixed effect, $f_{j,l}$ a language fixed effect, $f_{j,f}$ a frequency fixed effect, $f_{j,i} \times f_{j,p}$ a source×type fixed effect. The type is either a report or a news article. Allowing for all these other fixed effects besides the quarter fixed effects ensures that the index is independent of changes over time in the sample's composition across newspapers and reports versus regular articles. Moreover, it controls for differences in the sentiment scores across sources, languages, and frequencies. This might be important because the sentiment scores differ considerably across sources, types, frequencies, and languages.

- treat the identified segments as separate indicators in the static factor model. The reason for this robustness test is that in the baseline specification, the source-level sentiments are assumed to follow

$$s_{j,i,t} = (\lambda_{j,i}/\sigma_{j,i,t}) f_{j,t} + (e_{j,i,t}/\sigma_{j,i,t}). \tag{3.12}$$

  The factor loadings $\lambda_{j,i}$ have been scaled by $1/\sigma_{j,i,t}$. If $\sigma_{j,i,t}$ is large in the early part of the sample, then principle components (which assumes a time-invariant value of $\lambda_{j,i}$) potentially will under-estimate the variance of $f_{j,t}$ in the early part of the sample.

Other adjustment methods. Instead of adjusting the source-level indicators according to the procedure described in section 3.3.2, I

- do not scale the sentiment-based indicators. I still identify structural breaks but only demean each segment. I do not change the calculation of the count-based indicators. See section 3.C in the Appendix for more details.

- do not subtract a trend.

Subsets of data. Instead of using all available data, I calculate the indicator

- only using German texts.
- only using French texts.
- only using sentiment-based indicators.
- only using count-based indicators.

The alternative indicators are very similar, as shown in Figures 3.11 and 3.A.12 in the Appendix. However, there is some more dispersion before World War I. Nevertheless, all alternative indicators capture major downturns like the Franco-Prussian war from 1870 or the Baring crisis from 1890. There is a substantial correlation between the alternative indicators and the baseline indicator. The correlation between the alternative indicators and the output gap is similar to that between the baseline indicator and the output gap. Finally, the recession probabilities obtained with the MS-DFM in Figure 3.A.11 in the Appendix show substantial overlap with the dating obtained using Hamilton's (1989) model. These analyses demonstrate that the indicator is not overly sensitive to specific aggregation techniques or data limitations.

3.6 Concluding remarks

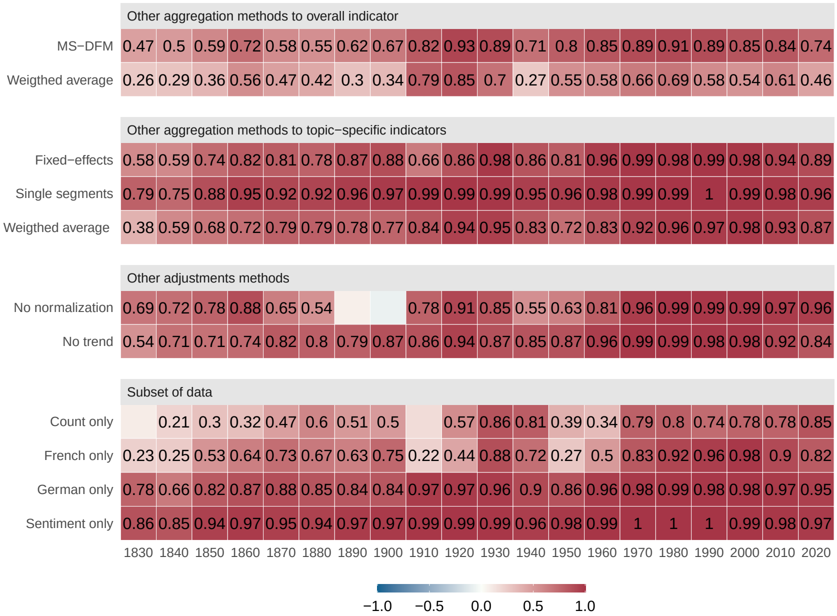
It is well known that measurement problems in hard economic data distort business cycle facts in the 19$^{\text{th}}$ century. For Switzerland, existing GDP series have "the characteristics of a rough estimate at best" (Historische Statistik der Schweiz HSSO, 2012k) and are available only at an annual frequency.

To overcome this problem, this chapter proposes a novel approach to measuring business cycle fluctuations over long historical episodes using textual data. I collect a large body of texts relevant to business cycle fluctuations. Using textual analysis, I create a quarterly business cycle indicator for Switzerland from 1821 to 2021. Based on this indicator, a business cycle chronology for Switzerland is established.

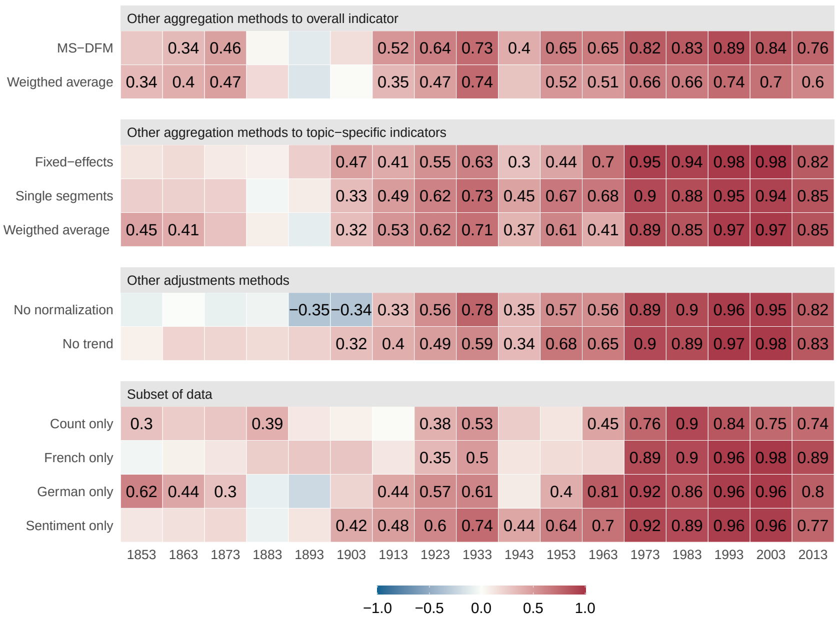
The business cycle indicator successfully captures major economic downturns, including the two World Wars, the Great Depression, and various financial crises. It strongly correlates with existing hard economic data in Switzerland, particularly in the second half of the 20$^{\text{th}}$ century and the early 21$^{\text{st}}$ century. This correlation, although somewhat

Figure 3.11 — Correlations of alternative indicators

a) Correlation with baseline composite indicator



b) Correlation with output gap



Notes: The graph shows rolling correlations with a fixed window size of 20 years between the indicator on the left and the composite indicator (in panel a) and the output gap (in panel b). The period considered is given by the year on the x-axis plus minus ten years. Only statistically significant (on a 10% level) correlations are labeled.

weaker, extends to earlier periods. These lower correlations do not necessarily imply that the indicator does not accurately represent economic fluctuations in the 19th and early 20th centuries. Instead, they may reflect higher measurement errors in the data from earlier periods, which push the correlation toward zero. Therefore, a low correlation is what we would expect under the assumption that hard economic data in the 19th century is measured with higher measurement error. Comparing the indicator with potentially uncollected hard data to verify its accuracy would be an interesting avenue for future research.

The business cycle dating is consistent with the one by a dating committee for the 20th century. For the 19th century, the dating is consistent with narratives about wars and crises. Moreover, the results show that Swiss recessions have become less frequent, aligning with trends observed in other European nations. In contrast to other European countries, the duration of Swiss recessions does not show a significant reduction.

Because hard data and qualitative data are subject to measurement error, it is crucial to combine different data sources to obtain a reliable measure of the business cycle. Therefore, a promising line for future research would be to combine error-prone hard and qualitative data to obtain a more accurate measure of the business cycle.

References

Abdaoui, A., Azé, J., Bringay, S., & Poncelet, P. (2017). FEEL: A french expanded emotion lexicon. *Language Resources and Evaluation*, *51*(3), 833–855. https://doi.org/10.1007/s10579-016-9364-5

Angelico, C., Marcucci, J., Miccoli, M., & Quarta, F. (2022). Can we measure inflation expectations using twitter? *Journal of Econometrics*, *228*(2), 259–277. https://doi.org/10.1016/j.jeconom.2021.12.008

Ardia, D., Bluteau, K., & 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. https://doi.org/10.1016/j.ijforecast.2018.10.010

Ardia, D., Bluteau, K., & Kassem, A. (2021). A century of economic policy uncertainty through the french–canadian lens. *Economics Letters*, *205*, 109938. https://doi.org/10.1016/j.econlet.2021.109938

Aruoba, S. B., & Drechsel, T. (2024). Identifying monetary policy shocks: A natural language approach. *NBER Working Paper No. 32417*. https://doi.org/10.3386/w32417

Ash, E., & Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, *15*(1), 659–688. https://doi.org/10.1146/annurev-economics-082222-074352

Bai, J., & Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, *176*(1), 18–29. https://doi.org/10.1016/j.jeconom.2013.03.007

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Barbaglia, L., Consoli, S., & Manzan, S. (2023). Forecasting with economic news. *Journal of Business & Economic Statistics*, *41*(3), 708–719. https://doi.org/10.1080/07350015.2022.2060988

Beach, B., & Hanlon, W. W. (2022). Historical newspaper data: A researcher's guide and toolkit. *NBER Working Paper No. 30135*.

Binder, C. C. (2016). Estimation of historical inflation expectations. *Explorations in Economic History*, *61*, 1–31. https://doi.org/10.1016/j.eeh.2016.01.002

Bolt, J., & van Zanden, J. L. (2020). Maddison style estimates of the evolution of the world economy. a new 2020 update. *Maddison-Project Working Paper*.

Broadberry, S., Chadha, J. S., Lennard, J., & Thomas, R. (2023). Dating business cycles in the united kingdom, 1700–2010. *The Economic History Review*, *76*(4), 1141–1162. https://doi.org/10.1111/ehr.13238

Broadberry, S., & Lennard, J. (2023). European business cycles and economic growth, 1300-2000. *Economic History Working Papers 120364*. https://ideas.repec.org/p/ehl/wpaper/120364.html

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020, July 22). Language models are few-shot learners. http://arxiv.org/abs/2005.14165

Bry, G., & Boschan, C. (1971). *Cyclical analysis of time series: Selected procedures and computer programs*. National Bureau of Economic Research; Columbia University Press.

Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. National Bureau of Economic Research, Inc.

Burri, M. (2023). Do daily lead texts help nowcasting GDP growth? *IRENE Working Papers 23-02*.

Burri, M., & Kaufmann, D. (2020). A daily fever curve for the swiss economy. *Swiss Journal of Economics and Statistics*, *156*(1), 6. https://doi.org/10.1186/s41937-020-00051-z

Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2023). Business news and business cycles. *Journal of Finance, Forthcoming*. http://dx.doi.org/10.2139/ssrn.3446225

Camacho, M., Perez-Quiros, G., & Poncela, P. (2015). Extracting nonlinear signals from several economic indicators. *Journal of Applied Econometrics*, *30*(7), 1073–1089. https://doi.org/10.1002/jae.2416

Canova, F. (1994). Detrending and turning points. *European Economic Review*, *38*(3), 614–623. https://doi.org/10.1016/0014-2921(94)90097-3

Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics*, *41*(3), 475–512. https://doi.org/10.1016/S0304-3932(98)00006-3

Chauvet, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching. *International Economic Review*, *39*(4), 969. https://doi.org/10.2307/2527348

Church, C. H., & Head, R. C. (2013, May 23). *A concise history of switzerland* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139013765

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Dagum, E. B., & Cholette, P. A. (2006). *Benchmarking, temporal distribution, and reconciliation methods for time series* (Vol. Lecture notes in statistics). Springer.

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, *66*(333), 99–102. https://doi.org/10.1080/01621459.1971.10482227

Diebold, F. X., & Rudebusch, G. D. (1996). Measuring business cycles: A modern perspective. *The Review of Economics and Statistics*, *78*(1), pp. 67–77.

Edwards, A. W. F., & Cavalli-Sforza, L. L. (1965). A method for cluster analysis. *Biometrics*, *21*(2), 362. https://doi.org/10.2307/2528096

Ellingsen, J., Larsen, V. H., & Thorsrud, L. A. (2022). News media versus FRED-MD for macroeconomic forecasting. *Journal of Applied Econometrics*, *37*(1), 63–81. https://doi.org/10.1002/jae.2859

Feinerer, I., & Hornik, K. (2019). *Tm: Text mining package*. https://CRAN.R-project.org/package=tm

Glocker, C., & Wegmueller, P. (2020). Business cycle dating and forecasting with real-time swiss GDP data. *Empirical Economics*, *58*(1), 73–105. https://doi.org/10.1007/s00181-019-01666-9

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*(2), 357. https://doi.org/10.2307/1912559

Hamilton, J. D. (2018). Why you should never use the hodrick-prescott filter. *The Review of Economics and Statistics*, *100*(5), 831–843. https://doi.org/10.1162/rest_a_00706

Hanna, A. J., Turner, J. D., & Walker, C. B. (2020). News media and investor sentiment during bull and bear markets. *The European Journal of Finance*, *26*(14), 1377–1395. https://doi.org/10.1080/1351847X.2020.1743734

Harding, D., & Pagan, A. (2002). Dissecting the cycle: A methodological investigation. *Journal of Monetary Economics*, *49*(2), 365–381. https://doi.org/10.1016/S0304-3932(01)00108-8

Hirshleifer, D., Mai, D., & Pukthuanthong, K. (2023). War discourse and the cross section of expected stock returns. *NBER Working Paper No. 31348*. https://doi.org/10.3386/w31348

Historische Statistik der Schweiz HSSO. (2012a). Ausfuhrmengen nach warenarten und ausfuhrmengenindizes 1851–1913. Retrieved September 19, 2023, from https://hsso.ch/de/2012/l/2

Historische Statistik der Schweiz HSSO. (2012b). Bilanz des aussenhandels und zollerträgnisse nach warenarten 1886–1992. Retrieved September 19, 2023, from https://hsso.ch/de/2012/l/3

Historische Statistik der Schweiz HSSO. (2012c). Bruttoinlandprodukt nach verwendungsarten in preisen von 1929 und nominal, 1890-1948. Retrieved September 19, 2023, from https://hsso.ch/de/2012/q/16a

Historische Statistik der Schweiz HSSO. (2012d). Bruttoinlandprodukt nach verwendungsarten zu preisen von 1990 und nominal, 1948-2005. Retrieved September 19, 2023, from https://hsso.ch/de/2012/q/16b

Historische Statistik der Schweiz HSSO. (2012e). Einfuhrmengen nach warenarten und einfuhrmengenindizes 1851–1913. Retrieved September 19, 2023, from https://hsso.ch/de/2012/l/1a

Historische Statistik der Schweiz HSSO. (2012f). Erwerbstätige, wochen-, jahresarbeitszeit 1890-2005. Retrieved September 19, 2023, from https://hsso.ch/de/2012/f/29a

Historische Statistik der Schweiz HSSO. (2012g). Monatliche und vierteljährliche ausfuhrmengen januar 1924 bis april 1967 und monatlicher und vierteljährlicher ausfuhrmengenindex von august 1974 bis oktober 1987. Retrieved September 19, 2023, from https://hsso.ch/de/2012/l/5a

Historische Statistik der Schweiz HSSO. (2012h). Monatliche und vierteljährliche ausfuhrwerte 1924–1992. Retrieved September 19, 2023, from https://hsso.ch/de/2012/l/8

Historische Statistik der Schweiz HSSO. (2012i). Monatliche und vierteljährliche einfuhrmengen januar 1924 bis dezember 1987. Retrieved September 19, 2023, from https://hsso.ch/de/2012/l/4

Historische Statistik der Schweiz HSSO. (2012j). Monatliche und vierteljährliche einfuhrwerte 1924–1992. Retrieved September 19, 2023, from https://hsso.ch/de/2012/l/7

Historische Statistik der Schweiz HSSO. (2012k). Nominales und reales BIP und bruttowertschöpfung nach branchen 1851-1913. Retrieved September 19, 2023, from https://hsso.ch/2012/q/1a

Historische Statistik der Schweiz HSSO. (2012l). Stellensuchende und arbeitslosenquote nach geschlecht im jahresmittel 1913-1995. Retrieved September 19, 2023, from https://hsso.ch/de/2012/f/18a

Horvath, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *The Annals of Statistics*, *21*(2). https://doi.org/10.1214/aos/1176349143

Kabiri, A., James, H., Landon-Lane, J., Tuckett, D., & Nyman, R. (2023). The role of sentiment in the US economy: 1920 to 1934. *The Economic History Review*, 3–30. https://doi.org/10.1111/ehr.13160

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2022). Making text count: Economic forecasting using newspaper text. *Journal of Applied Econometrics*, *37*(5), 896–919. https://doi.org/10.1002/jae.2907

Kaufmann, D. (2020). Is deflation costly after all? the perils of erroneous historical classifications. *Journal of Applied Econometrics*, *35*(5), 614–628. https://doi.org/10.1002/jae.2762

Killick, R., & Eckley, I. A. (2014). **changepoint** : An *R* package for changepoint analysis. *Journal of Statistical Software*, *58*(3). https://doi.org/10.18637/jss.v058.i03

Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, *60*(1), 1–22. https://doi.org/10.1016/0304-4076(94)90036-1

Kim, M.-J., & Yoo, J.-S. (1995). New index of coincident indicators: A multivariate markov switching factor model approach. *Journal of Monetary Economics*, *36*(3), 607–630. https://doi.org/10.1016/0304-3932(95)01229-X

Larsen, V. H., & Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, *210*(1), 203–218. https://doi.org/10.1016/j.jeconom.2018.11.013

Larsen, V. H., Thorsrud, L. A., & Zhulanova, J. (2021). News-driven inflation expectations and information rigidities. *Journal of Monetary Economics*, *117*, 507–520. https://doi.org/10.1016/j.jmoneco.2020.03.004

Larsen, V. H. (2021). Components of uncertainty. *International Economic Review*, *62*(2), 769–788. https://doi.org/10.1111/iere.12499

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Luhn, H. P. (1960). Key word-in-context index for technical literature. *American Documentation*, *11*(4), 288–295. https://doi.org/10.1002/asi.5090110403

Mariano, R. S., & Murasawa, Y. (2010). A coincident index, common factors, and monthly real GDP. *Oxford Bulletin of Economics and Statistics*, *72*(1), 27–46. https://doi.org/10.1111/j.1468-0084.2009.00567.x

OpenAI. (2023a). ChatGPT. Retrieved October 1, 2023, from https://www.chatgpt.com/

OpenAI. (2023b). GPT-3.5 API. Retrieved October 1, 2023, from https://www.platform.openai.com/

Picard, F., Robin, S., Lavielle, M., Vaisse, C., & Daudin, J.-J. (2005). A statistical approach for array CGH data analysis. *BMC Bioinformatics*, *6*(1), 27. https://doi.org/10.1186/1471-2105-6-27

Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS - a publicly available german-language resource for sentiment analysis. *Proceedings of the seventh international conference on language resources and evaluation (LREC'10)*.

Romer, C. D. (1994). Remeasuring business cycles. *The Journal of Economic History*, *54*(3), 573–609.

Romer, C. D., & Romer, D. H. (2020). NBER recession dates: Strengths, weaknesses, and a modern upgrade. *mimeo*.

Sax, C., & Steiner, P. (2013). Temporal disaggregation of time series. *The R Journal*, *5*(2), 80. https://doi.org/10.32614/RJ-2013-028

Schorfheide, F., & Song, D. (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, *33*(3), 366–380. https://doi.org/10.1080/07350015.2014.954707

Scott, A. J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, *30*(3), 507. https://doi.org/10.2307/2529204

Sen, A., & Srivastava, M. S. (1975). On tests for detecting change in mean. *The Annals of Statistics*, *3*(1). https://doi.org/10.1214/aos/1176343001

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, *228*(2), 221–243. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.07.053

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004. https://doi.org/10.1257/aer.107.4.967

Shiller, R. J. (2019). *Narrative economics: How stories go viral & drive major economic events*. Princeton University Press.

Siliverstovs, B. (2011). Dating business cycles in a historical perspective: Evidence for switzerland. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1886347

Stineman, R. W. (1980). A consistently well behaved method of interpolation. *Creative Computing*, *6*(7), 54–57.

Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, *4*, 351–394. https://doi.org/10.1086/654119

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162. https://doi.org/10.1198/073500102317351921

Stock, J. H., & Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. *Handbook of macroeconomics* (pp. 415–525). Elsevier. https://doi.org/10.1016/bs.hesmac.2016.04.002

Stohr, C. (2016). Trading gains: New estimates of swiss GDP,1851 to 2008. *Economic History Working Papers No. 67032*.

Stohr, C. (2017). Das schweizer bruttoinlandprodukt: Methoden, daten und internationale vergleiche. *mimeo*. https://doi.org/10.13140/RG.2.2.29030.11848

Ter Ellen, S., Larsen, V. H., & Thorsrud, L. A. (2022). Narrative monetary policy surprises and the media. *Journal of Money, Credit and Banking*, *54*(5), 1525–1549. https://doi.org/10.1111/jmcb.12868

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Thorp, W. L. (1926). *Business annals*. National Bureau of Economic Research, Inc.

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. https://doi.org/10.1080/07350015.2018.1506344

Van Binsbergen, J., Bryzgalova, S., Mukhopadhyay, M., & Sharma, V. (2024). (almost) 200 years of news-based economic sentiment. *NBER Working Paper No. 32026*. https://doi.org/10.3386/w32026

Appendix 3.A Supplementary material

Table 3.A.1 — Textual data sources

Publication	Name	Source	Availability	Lang.	Freq.	# texts
AGB	Zeitschrift Schweizer Arbeitgeber	SWAB	1907 - 2014	de	M	124.94
ARC	ArcInfo	e-newspaper archives.ch	2018 - 2021	fr	D	143.05
AWP	AWP Financial News	AWP	2001 - 2020	de	D	1860.85
BAN	Briger Anzeiger	e-newspaper archives.ch	1899 - 1933	de	BW	204.22
BAZ	Basler Zeitung	Tamedia	2000 - 2021	de	D	323.94
BEOL	Berner Oberländer	Tamedia	2018 - 2021	de	D	76.22
BILA	Bilan	Tamedia	1996 - 2021	fr	D	18.08
BNN	Bündner Nachrichten	e-newspaper archives.ch	1885 - 1892	de	D	50.42
BR_DE_GB	Annual Report Swiss Confederation	Swiss Confederation	1849 - 2018	de	Y	306.57
BR_de_SR	Annual Report Government Account	Swiss Confederation	1849 - 2021	de	Y	3193.07
BR_FR_GB	Annual Report Swiss Confederation	Swiss Confederation	1848 - 2018	fr	Y	3170.38
BU	Der Bund	Tamedia	1995 - 2021	de	D	800.47
BUR	Bote vom Untersee und Rhein	e-newspaper archives.ch	1900 - 2018	de	BW	620.46
BZ	Berner Zeitung	Tamedia	1996 - 2021	de	D	1051.75
CMV	Aktiv : CMV/FCOM : Gemeinsames Gewerkschaftsmagazin	e-newspaper archives.ch	1930 - 1998	de	BM	40.83
CS_BUL	Credit Suisse Bulletin	Swiss National Library	1895 - 2016	de	Y	3121.80
DBB	Der Bund	e-newspaper archives.ch	1850 - 1994	de	D	2865.65
Démocrate	Le Démocrate	Scriptorium	1855 - 1999	fr	D	4141.31
DMR	Der Murtenbieter	e-newspaper archives.ch	1854 - 2010	de	BW	651.18
EDP	Engadiner Post	e-newspaper archives.ch	1893 - 1930	de	TW	92.23

Continued on next page

Table 3.A.1 – continued from previous page

Publication	Name	Source	Availability	Lang.	Freq.	# texts
EIZE	Eidgenössische Zeitung	e-newspaper archives.ch	1838 - 1864	de	D	73.19
EXR	FAN - L'express : feuille d'avis de Neuchâtel	e-newspaper archives.ch	1738 - 2018	fr	D	8410.07
FAAV	Feuille d'avis du district d'Avenches	Scriptorium	1882 - 2010	fr	D	5750.28
FCS	Feuille commerciale de Sierre et du district	e-newspaper archives.ch	1918 - 1930	fr	BW	50.74
FDV	Journal et feuille d'avis du Valais	e-newspaper archives.ch	1903 - 1968	fr	TW	1311.36
FUW	Finanz und Wirtschaft	Tamedia	2000 - 2021	de	D	93.47
FZG	Freiburger Nachrichten	e-newspaper archives.ch	1864 - 2006	de	D	2638.32
GAV	Gazette du Valais	e-newspaper archives.ch	1855 - 1922	fr	TW	258.07
GBL	Geschäftsblatt für den obern Teil des Kantons Bern	e-newspaper archives.ch	1876 - 1938	de	BW	339.40
GDB	Gazette de Berne	e-newspaper archives.ch	1692 - 1797	fr	BW	69.08
GTR	Grütlianer	e-newspaper archives.ch	1852 - 1925	de	BW	229.89
HAZ	Handels-Zeitung	SWAB	1861 - 1894	de	D	36.38
HEU	24 Heures	Tamedia	1996 - 2021	fr	D	535.01
IMP	L'impartial	e-newspaper archives.ch	1881 - 2018	fr	D	7102.04
IND	L'indicateur = Der Anzeiger	e-newspaper archives.ch	1914 - 1923	fr	W	48.32
ISB	Intelligenzblatt für die Stadt Bern	e-newspaper archives.ch	1834 - 1922	de	D	579.98
JC	Journal du district de Cossonay	Scriptorium	1899 - 2007	fr	D	3353.52
JM	Journal de Morges	Scriptorium	1894 - 2017	fr	D	9032.77
KK_CA	Reports Kommission für Konjunkturbeobachtung	SNB	1953 - 1983	de	Y	7.81
KK_ES	Econmic Situation report Kommision für Konjunkturbeobachtung	SNB	1932 - 2001	de	Y	47.15

Continued on next page

Table 3.A.1 – continued from previous page

Publication	Name	Source	Availability	Lang.	Freq.	# texts
KK_MIT	Mitteilungen Kommission für Konjunkturbeobachtung	SNB	1933 - 2007	de	Y	145.69
LAT	Langenthaler Tagblatt	Tamedia	2019 - 2021	de	D	26.99
LB	Der Landbote	Tamedia	1998 - 2021	de	D	446.08
LBP	Le bien public	e-newspaper archives.ch	1879 - 1888	fr	TW	70.05
LCE	Le confédéré	e-newspaper archives.ch	1861 - 2009	fr	W/BW	701.46
LCG	Le confédéré de Fribourg	e-newspaper archives.ch	1848 - 1907	fr	BW	169.57
LCR	Le chroniqueur	e-newspaper archives.ch	1854 - 1881	fr	TW	96.29
LEJ	Le Jura	e-newspaper archives.ch	1852 - 1970	fr	BW	182.35
LES	L'essor	e-newspaper archives.ch	1906 - 2015	fr	M	51.82
LFM	Le Franc-Montagnard	e-newspaper archives.ch	1898 - 2020	fr	BW/TW	533.68
LGE	La Gruyère	e-newspaper archives.ch	1882 - 1930	fr	TW	158.79
LLE	La liberté	e-newspaper archives.ch	1871 - 2012	fr	D	6975.28
LSL	La Suisse libérale	e-newspaper archives.ch	1881 - 1982	fr	W	304.98
MIG	Construire : hebdomadaire du capital à but social	e-newspaper archives.ch	1944 - 2004	fr	W	399.06
MIM	Wir Brückenbauer : Wochenblatt des sozialen Kapitals	e-newspaper archives.ch	1943 - 2004	de	W	479.30
NVB	Nidwaldner Volksblatt	e-newspaper archives.ch	1866 - 1991	de	BW	236.52
NVE	Le nouvelliste	e-newspaper archives.ch	1904 - 2016	fr	D	7505.93
NZG	Neue Zuger Zeitung	e-newspaper archives.ch	1846 - 1891	de	BW	25.10
NZN	Neue Zürcher Nachrichten	e-newspaper archives.ch	1895 - 1991	de	D	1412.94
NZZ	Neue Zürcher Zeitung	e-newspaper archives.ch	1780 - 2020	de	D	3698.76

Continued on next page

Table 3.A.1 – continued from previous page

Publication	Name	Source	Availability	Lang.	Freq.	# texts
OIZ	Die Gewerkschaft : das Magazin der Gewerkschaft Textil, Chemie, Papier	e-newspaper archives.ch	1901 - 1992	de	BM	51.08
OTB	Oberländer Tagblatt	e-newspaper archives.ch	1909 - 1961	de	D	371.66
SGZ	St. Galler Zeitung	e-newspaper archives.ch	1831 - 1881	de	D	173.81
SLB	Seeländer Bote	e-newspaper archives.ch	1850 - 1904	de	TW	75.30
SMZ	SMUV-Zeitung / Schweizerischer Metall- und Uhrenarbeitnehmer-Verband	e-newspaper archives.ch	1902 - 2001	de	M	44.69
SNB_GB	Annual Report SNB	SNB	1908 - 2018	de	Y	736.58
SNB_GMB	Money Market Report SNB	SNB	1945 - 1985	de	Y	58.09
SNB_QB	Quarterly Report SNB	SNB	1967 - 2019	de	Q	46.70
SWA_BER_Vorort	Annual Report Vorort	SWAB	1878 - 1976	de	Y	2063.75
SWA_GB_BCG	Annual Report Banque du Commerce Geneve	SWAB	1845 - 1907	fr	Y	85.54
SWA_GB_BCN	Annual Report Banque cantonale Neuchâteloise	SWAB	1883 - 1989	fr	Y	166.83
SWA_GB_BCV	Annual Report Banque cantonale vaudoise	SWAB	1854 - 1949	fr	Y	25.48
SWA_GB_BCVL	Annual Report Banque cantonale du valais	SWAB	1856 - 1871	fr	Y	11.27
SWA_GB_BDG	Annual Report Banque de Genève	SWAB	1858 - 1930	fr	Y	36.24
SWA_GB_BGG	Annual Report Banque Glane/Gruyere	SWAB	1888 - 1960	fr	Y	4.45
SWA_GB_BHCG	Annual Report Banque hypothécaire du canton de Genève	SWAB	1848 - 1950	fr	Y	108.56
SWA_GB_BPG	Annual Report Banque populaire de la Gruyère	SWAB	1864 - 1968	fr	Y	58.07

Continued on next page

Table 3.A.1 – continued from previous page

Publication	Name	Source	Availability	Lang.	Freq.	# texts
SWA_GB_CAIB	Annual Report Crédit agricole et industriel de la Broye	SWAB	1867 - 1965	fr	Y	103.68
SWA_GB_CCIG	Annual Report Chambre de Commerce de Genève	SWAB	1865 - 1930	fr	Y	80.12
SWA_GB_CEB	Annual Report Caisse d'epargne de Bassecourt	SWAB	1883 - 1988	fr	Y	59.58
SWA_GB_CECL	Annual Report Caisse d'epargne et de crédit Lausanne	SWAB	1878 - 1956	fr	Y	35.91
SWA_GB_CEPL	Annual Report Caisse d'Epargne et de Prévoyance de Lausanne	SWAB	1905 - 1960	fr	Y	7.80
SWA_GB_CF	Annual Report Credit foncier vaudois	SWAB	1901 - 1956	fr	Y	41.84
SWA_GB_CHCF	Annual Report Caisse Hypo Fribourg	SWAB	1854 - 1909	fr	Y	4.58
SWA_GB_CHCV	Annual Report Caisse hypothécaire Cantonale vaudoise	SWAB	1860 - 1900	fr	Y	34.26
SWA_GB_CME	Annual Report Caisse Mutuel pour l'Epargne	SWAB	1874 - 1920	fr	Y	17.42
SWA_GB_CVCI	Annual Report Chambre vaudoise du commerce et de l'industrie	SWAB	1904 - 1929	fr	Y	2.90
SWA_GB_Helveti	Annual Report Helvetia	SWAB	1859 - 1957	de	Y	4.18
SWA_GB_SLKB	Annual Report Spar- und Leihkasse Bern	SWAB	1858 - 1960	de	Y	46.59
SWA_GB_SwissR	Annual Report Swiss RE	SWAB	1864 - 1960	de	Y	33.26
SWA_GB_ZVAO	Annual Report Arbeitgeberverband	SWAB	1910 - 1960	de	Y	94.63
TA	Tages-Anzeiger	Tamedia	1996 - 2021	de	D	994.33
TAA	Täglicher Anzeiger für Thun und das Berner Oberland	e-newspaper archives.ch	1877 - 1907	de	D	68.06

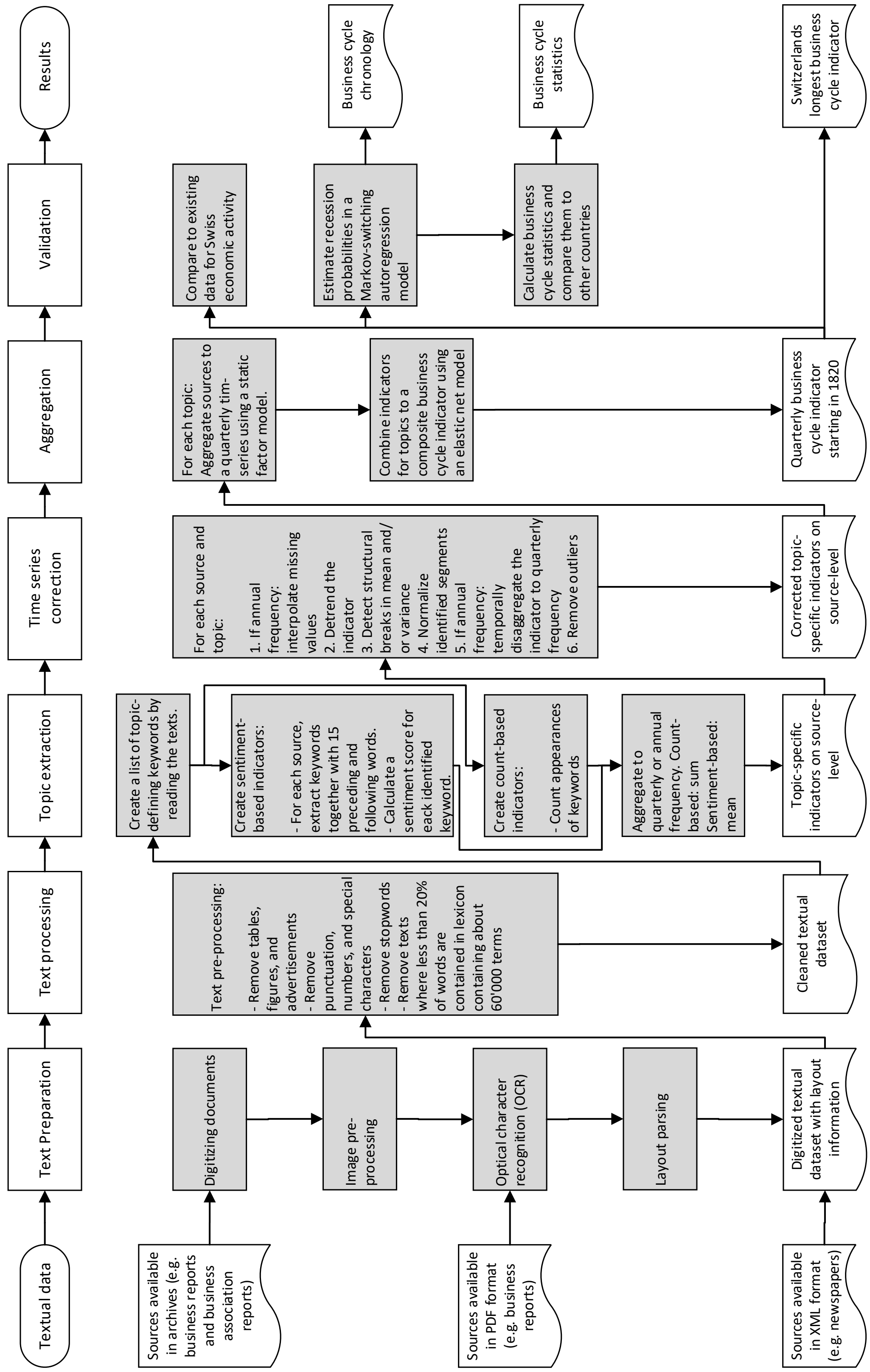
Continued on next page

Table 3.A.1 – continued from previous page

Publication	Name	Source	Availability	Lang.	Freq.	# texts
TCP	FTCP : organe de la Fédération du personnel du textile, de la chimie et du papier	e-newspaper archives.ch	1951 - 1979	fr	BM	23.40
TDG	Tribune de Genève	Tamedia	1996 - 2021	fr	D	592.38
THT	Thuner Tagblatt	Tamedia	2019 - 2021	de	D	28.61
TSB	Tagblatt der Stadt Biel	e-newspaper archives.ch	1865 - 1900	de	D	115.44
VHT	VHTL-Zeitung / Gewerkschaft Verkauf, Handel, Transport, Lebensmittel	e-newspaper archives.ch	1904 - 2004	de	W/BM	21.98
WAB	Walliser Bote	e-newspaper archives.ch	1861 - 2008	de	D	2494.40
ZFZ	Zürcherische Freitagszeitung	e-newspaper archives.ch	1705 - 1914	de	W	53.95
ZGN	Zuger Nachrichten	e-newspaper archives.ch	1886 - 1900	de	TW	18.80
ZHUL	Zürcher Unterländer	Tamedia	2011 - 2021	de	D	179.52
ZKB_GB	Annual Reports Zürcher Kantonalbank	ZKB	1870 - 2016	de	Y	733.79
ZSZ	Zürichsee-Zeitung	Tamedia	2001 - 2021	de	D	549.34
ZVB	Zuger Volksblatt	e-newspaper archives.ch	1861 - 1900	de	BW	32.11
ZWB	Zürcherisches Wochenblatt	e-newspaper archives.ch	1801 - 1842	de	BW	33.43

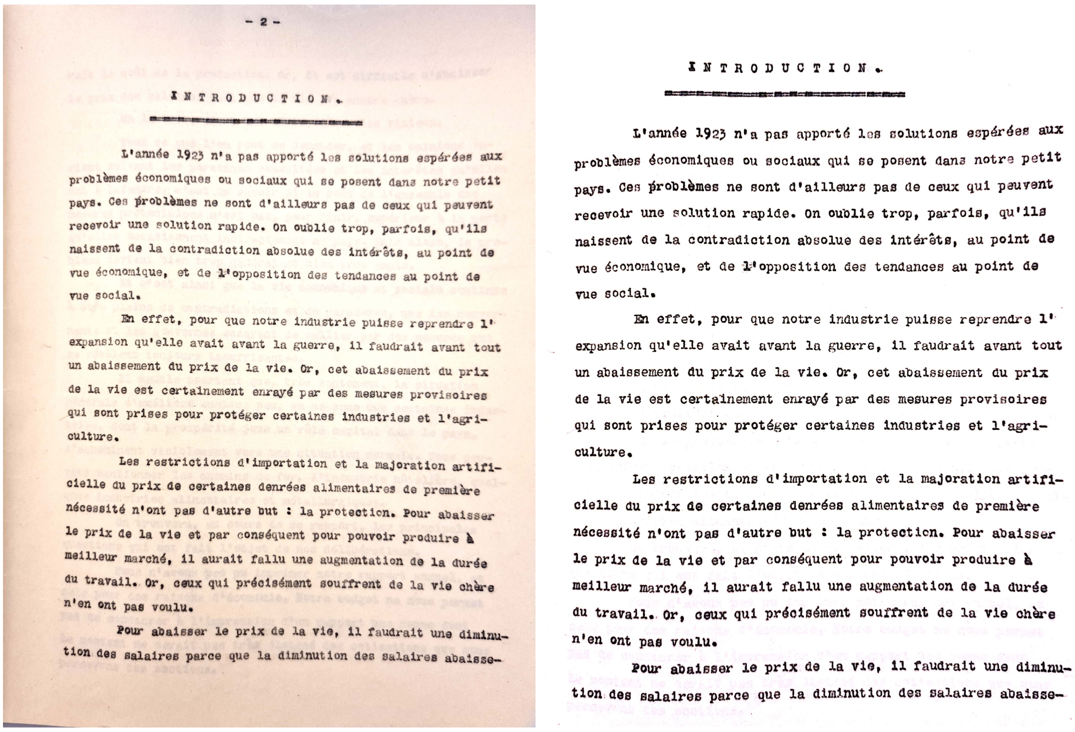
Notes: Publications with source SWAB were digitized by ourselves (except AGB and HAZ).
Frequencies: D: daily, TW: three a week, BW: twice a week, W: weekly, BM: twice a month, M: monthly, Q: quarterly, Y: yearly.
Number of texts is in thousands.

Figure 3.A.1 — Schematic diagram of the study's workflow



Notes: The diagram shows the schematic workflow of the study. Wavy boxes denote different states of the textual data. Gray-shaded boxes represent tasks that process the data.

Figure 3.A.2 — Scanned page, a pre-processed image, and the OCR text: Chambre vaudoise du commerce et de l'industrie, 1923.



\- 2 -

INTRODUCTION.

L'année 1923 n'a pas apporté les solutions espérées aux problèmes économiques ou sociaux qui se posent dans notre petit pays. Ces problèmes ne sont d'ailleurs pas de ceux qui peuvent recevoir une solution rapide. On oublie trop, parfois, qu'ils naissent de la contradiction absolue des intérêts, au point de vue économique, et de l'opposition des tendances au point de vue social.

En effet, pour que notre industrie puisse reprendre l'expansion qu'elle avait avant la guerre, il faudrait avant tout un abaissement du prix de la vie. Or, cet abaissement du prix de la vie est certainement enrayé par des mesures provisoires qui sont prises pour protéger certaines industries et l'agriculture.

Les restrictions d'importation et la majoration artificielle du prix de certaines denrées alimentaires de première nécessité n'ont pas d'autre but : la protection. Pour abaisser le prix de la vie et par conséquent pour pouvoir produire à meilleur marché, il aurait fallu une augmentation de la durée du travail. Or, ceux qui précisément souffrent de la vie chère n'en ont pas voulu.

Pour abaisser le prix de la vie, il faudrait une diminution des salaires parce que la diminution des salaires abaisse-

Notes: Source: Schweizerisches Wirtschaftsarchiv Basel.

Result from OCR of figure 3.A.2
INTRODUCTION*.

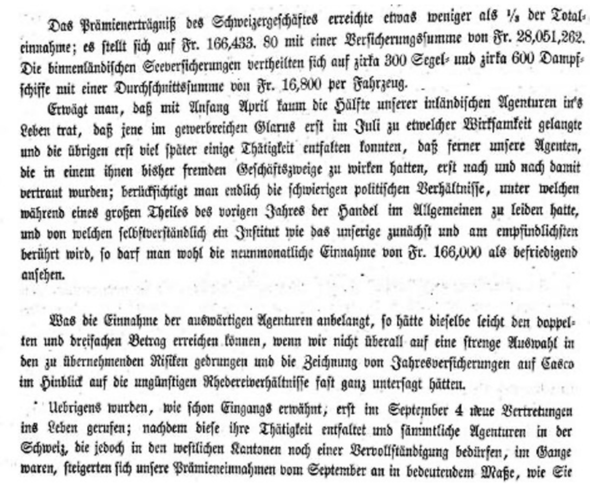
L'année 1923 n'a pan apporté les solutions espérées aux problèmes économiques ou sociaux qui so posent dans notre petit pays* Ces problèmes ne sont d'ailleurs pas de ceux qui peuvent recevoir une solution rapide, on ou'olie trop, parfois, qu'Ils naissent de la contradiction absolue des intérêts, au point de vue économique, et de *3?'opposition des tendances au point de vue social»

En effet, pour que notre industrie puisse reprendre l'expansion qu'elle avait avant la guerre, il faudrait avant tout un abaissement du prix de la vie. Or, cet abaissement du prix de la vie est certainement enrayé par des mesures provisoires qui sont prises pour protéger certaines industries et l'agriculture.

Les restrictions d'importation et la majoration artificielle du prix de certaines denrées alimentaires de première nécessité n'ont pas d'autre but : la protection. Pour abaisser le prix de la vie et par conséquent pour pouvoir produire à meilleur marché, il aurait fallu une augmentation de la durée du travail. Or, ceux qui précisément souffrent de la vie chère n'en ont pas voulu.

Pour abaisser le prix de la vie, il faudrait une diminution des salaires parce que la diminution des salaires abaisse-

Figure 3.A.3 — Scanned page, a pre-processed image, and the OCR text: Erster Geschäftsbericht, Helvetia, 1860.

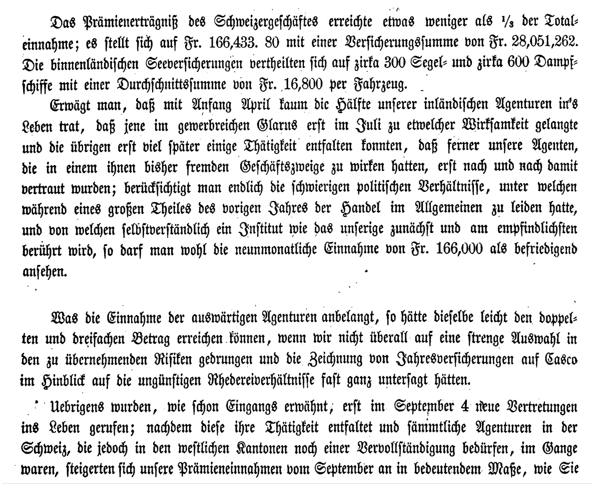


Das Prämienerträgniß des Schweizergeschäftes erreichte etwas weniger als ⅓ der Totaleinnahme; es stellt sich auf Fr. 166,433. 80 mit einer Versicherungssumme von Fr. 28,051,262. Die binnenländischen Seeversicherungen vertheilten sich auf zirka 300 Segel- und zirka 600 Dampfschiffe mit einer Durchschnittssumme von Fr. 16,800 per Fahrzeug.

Erwägt man, daß mit Anfang April kaum die Hälfte unserer inländischen Agenturen in's Leben trat, daß jene im gewerbreichen Glarus erst im Juli zu etwelcher Wirksamkeit gelangte und die übrigen erst viel später einige Thätigkeit entfalten konnten, daß ferner unsere Agenten, die in einem ihnen bisher fremden Geschäftszweige zu wirken hatten, erst nach und nach damit vertraut wurden; berücksichtigt man endlich die schwierigen politischen Verhältnisse, unter welchen während eines großen Theiles des vorigen Jahres der Handel im Allgemeinen zu leiden hatte, und von welchen selbstverständlich ein Institut wie das unserige zunächst und am empfindlichsten berührt wird, so darf man wohl die neunmonatliche Einnahme von Fr. 166,000 als befriedigend ansehen.

Was die Einnahme der auswärtigen Agenturen anbelangt, so hätte dieselbe leicht den doppelten und dreifachen Betrag erreichen können, wenn wir nicht überall auf eine strenge Auswahl in den zu übernehmenden Risiken gedrungen und die Zeichnung von Jahresversicherungen auf Casco im Hinblick auf die ungünstigen Rhedereiverhältnisse fast ganz untersagt hätten.

Uebrigens wurden, wie schon Eingangs erwähnt, erst im September 4 neue Vertretungen ins Leben gerufen; nachdem diese ihre Thätigkeit entfaltet und sämmtliche Agenturen in der Schweiz, die jedoch in den westlichen Kantonen noch einer Vervollständigung bedürfen, im Gange waren, steigerten sich unsere Prämieneinnahmen vom September an in bedeutendem Maße, wie Sie



Das Prämienerträgniß des Schweizergeschäftes erreichte etwas weniger als ⅓ der Totaleinnahme; es stellt sich auf Fr. 166,433. 80 mit einer Versicherungssumme von Fr. 28,051,262. Die binnenländischen Seeversicherungen vertheilten sich auf zirka 300 Segel- und zirka 600 Dampfschiffe mit einer Durchschnittssumme von Fr. 16,800 per Fahrzeug.

Erwägt man, daß mit Anfang April kaum die Hälfte unserer inländischen Agenturen in's Leben trat, daß jene im gewerbreichen Glarus erst im Juli zu etwelcher Wirksamkeit gelangte und die übrigen erst viel später einige Thätigkeit entfalten konnten, daß ferner unsere Agenten, die in einem ihnen bisher fremden Geschäftszweige zu wirken hatten, erst nach und nach damit vertraut wurden; berücksichtigt man endlich die schwierigen politischen Verhältnisse, unter welchen während eines großen Theiles des vorigen Jahres der Handel im Allgemeinen zu leiden hatte, und von welchen selbstverständlich ein Institut wie das unserige zunächst und am empfindlichsten berührt wird, so darf man wohl die neunmonatliche Einnahme von Fr. 166,000 als befriedigend ansehen.

Was die Einnahme der auswärtigen Agenturen anbelangt, so hätte dieselbe leicht den doppelten und dreifachen Betrag erreichen können, wenn wir nicht überall auf eine strenge Auswahl in den zu übernehmenden Risiken gedrungen und die Zeichnung von Jahresversicherungen auf Casco im Hinblick auf die ungünstigen Rhedereiverhältnisse fast ganz untersagt hätten.

Uebrigens wurden, wie schon Eingangs erwähnt, erst im September 4 neue Vertretungen ins Leben gerufen; nachdem diese ihre Thätigkeit entfaltet und sämmtliche Agenturen in der Schweiz, die jedoch in den westlichen Kantonen noch einer Vervollständigung bedürfen, im Gange waren, steigerten sich unsere Prämieneinnahmen vom September an in bedeutendem Maße, wie Sie

Notes: Source: Schweizerisches Wirtschaftsarchiv Basel.

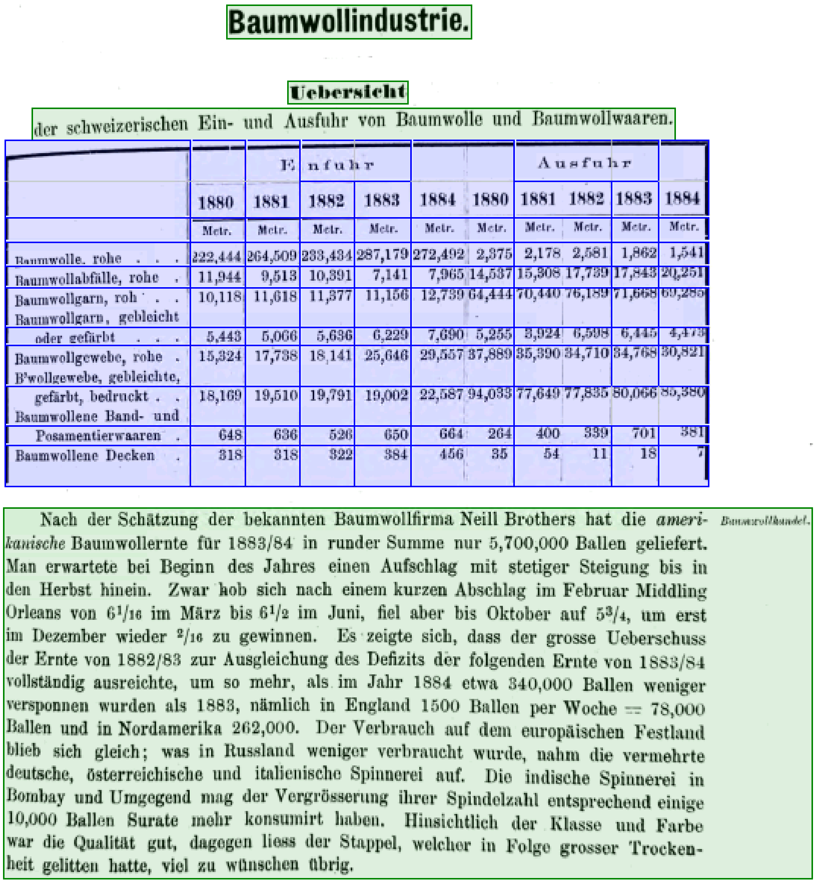
Result from OCR of figure 3.A.3

Das Prämienerträgniß des Schweizergeschäftes erreichte etwas weniger als !/; der Totaleinnahme; es stellt sich auf Fr. 166,433. 80 mit einer Versicherungssumme von Fr. 28,051,262. Die binnenländis hen Seeversicherungen vertheilten sich auf zirka 300 Segel- und zirka 600 Dampfschiffe mit einer Durchschnittssumme von Fr. 16,800 per Fahrzeug.

Erwägt man, daß mit Anfang April kaum die Hälfte unserer inländischen Agenturen in's Leben trat, daß jene im gewerbreichen Glarus erst im Juli zu etwelcher Wirksamkeit gelangte und die übrigen erst viel später einige Thätigkeit entfalten konnten, daß ferner unsere Agenten, die in einem ihnen bisher fremden Geschäftszweige zu wirken hatten, erst nach und nach damit vertraut wurden; berücsichtigt man endlich die schwierigen politischen Verhältnisse, unter welchen während eines großen Theiles des vorigen Jahres der Handel im Allgemeinen zu leiden hatte, und von welchen selbstverständlich ein Jnstitut wie das unserige zunächst und am empfindlichsten berührt wird, so darf man wohl die neunmonatliche Einnahme von Fr. 166,000 als befriedigend anjehen.

Was die Einnahme der auswärtigen Agenturen anbelangt, so hätte dieselbe leicht den doppelten und dreifachen Betrag erreichen können, wenn wir nicht überall auf eine strenge Auswahl in den zu übernehmenden Risiken gedrungen und die Zeichnung von Jahresversicherungen auf Casco im Hinbli> auf die ungünstigen Rhedereiverhältnisse fast ganz untersagt hätten. * "Vebrigens wurden, wie shon Eingangs erwähnt; erst im September 4 neue Vertretungen ins Leben gerufen; nachdem diese ihre Thätigkeit entfaltet und sämmtliche Agenturen in der Schweiz, die jedoch in den westlichen Kantonen no< einer Vervollständigung bedürfen, im Gange waren, steigerten sich unsere Prämieneinnahmen vom September an in bedeutendem Maße, wie Sie

Figure 3.A.4 — Parsing the layout: Handels- und Industrieverein Vorort, 1923



Baumwollindustrie.

Uebersicht

der schweizerischen Ein- und Ausfuhr von Baumwolle und Baumwollwaaren.

	Einfuhr					Ausfuhr				
	1880	1881	1882	1883	1884	1880	1881	1882	1883	1884
	Metr.	Metr.	Metr.	Metr.	Metr.	Metr.	Metr.	Metr.	Metr.	Metr.
Baumwolle, rohe . . .	222,444	264,509	233,434	287,179	272,492	2,375	2,178	2,581	1,862	1,541
Baumwollabfälle, rohe .	11,944	9,513	10,391	7,141	7,965	14,537	15,308	17,739	17,843	20,251
Baumwollgarn, roh . .	10,118	11,618	11,377	11,156	12,739	64,444	70,440	76,189	71,668	69,285
Baumwollgarn, gebleicht oder gefärbt . . .	5,443	5,066	5,636	6,229	7,690	5,255	3,924	6,598	6,445	4,473
Baumwollgewebe, rohe .	15,324	17,738	18,141	25,646	29,557	37,889	35,390	34,710	34,768	30,821
B'wollgewebe, gebleichte, gefärbt, bedruckt . .	18,169	19,510	19,791	19,002	22,587	94,033	77,649	77,835	80,066	85,380
Baumwollene Band- und Posamentierwaaren .	648	636	526	650	664	264	400	339	701	381
Baumwollene Decken .	318	318	322	384	456	35	54	11	18	7

Nach der Schätzung der bekannten Baumwollfirma Neill Brothers hat die *amerikanische* Baumwollernte für 1883/84 in runder Summe nur 5,700,000 Ballen geliefert. Man erwartete bei Beginn des Jahres einen Aufschlag mit stetiger Steigung bis in den Herbst hinein. Zwar hob sich nach einem kurzen Abschlag im Februar Middling Orleans von $6^1/_{16}$ im März bis $6^1/_2$ im Juni, fiel aber bis Oktober auf $5^3/_4$, um erst im Dezember wieder $^2/_{16}$ zu gewinnen. Es zeigte sich, dass der grosse Ueberschuss der Ernte von 1882/83 zur Ausgleichung des Defizits der folgenden Ernte von 1883/84 vollständig ausreichte, um so mehr, als im Jahr 1884 etwa 340,000 Ballen weniger versponnen wurden als 1883, nämlich in England 1500 Ballen per Woche = 78,000 Ballen und in Nordamerika 262,000. Der Verbrauch auf dem europäischen Festland blieb sich gleich; was in Russland weniger verbraucht wurde, nahm die vermehrte deutsche, österreichische und italienische Spinnerei auf. Die indische Spinnerei in Bombay und Umgegend mag der Vergrösserung ihrer Spindelzahl entsprechend einige 10,000 Ballen Surate mehr konsumirt haben. Hinsichtlich der Klasse und Farbe war die Qualität gut, dagegen liess der Stappel, welcher in Folge grosser Trockenheit gelitten hatte, viel zu wünschen übrig. *Baumwollhandel.*

Notes: This figure illustrates that it is possible to parse the layout (i.e., titles, tables and paragraphs) of a given text. Source: Schweizerisches Wirtschaftsarchiv Basel.

Table 3.A.2 — Validation data sources

Publication	Name	Source	Availability	Freq.
HSSODEF	Real GDP (deflator-adjusted)	Historische Statistik der Schweiz HSSO (2012k)	1851 - 1890	Y
HSSOCPI	Real GDP (CPI-adjusted)	Historische Statistik der Schweiz HSSO (2012k)	1851 - 1890	Y
HSSO	Nominal GDP	Historische Statistik der Schweiz HSSO (2012k)	1851 - 1890	Y
HSSO16A	Real and nominal GDP	Historische Statistik der Schweiz HSSO (2012c)	1890 - 1948	Y
HSSO16B	Real and nominal GDP	Historische Statistik der Schweiz HSSO (2012d)	1948 - 2005	Y
MADDISON	Real GDP	Bolt and van Zanden (2020)	1820 - 2018	Y
FSOGDP	Real and nominal GDP	Federal Statistical Office (FSO)	1948 - 2021	Y
SECOGDP	Real and nominal GDP	State Secretariat for Economic Affairs (SECO)	1980 - 2022	Q
CHAINR	Chained real GDP	Maddison, HSSO, FSO, SECO	1851 - 2021	Y
CHAINN	Chained nominal GDP	HSSO, FSO, SECO	1851 - 2021	Y
RGDPWORLD	Real GDP world		1830 - 2022	Y
RGDPCapita	Real GDP per capita	HSSO	1851 - 2021	Y
OutputGapHP	HP-filtered CHAINR	see CHAINR	1851 - 2021	Y
OutputGapHam	Hamilton-filtered CHAINR	see CHAINR	1851 - 2021	Y
GDPSTOHR	Real and nominal GDP	Stohr (2016)	1851 - 2008	Y
RGDPINT	Real GDP international (Austria, England, France, Germany, Italy, USA)	Bolt and van Zanden (2020)	1851 - 2020	Y
IMPVOL	Import volumes	Historische Statistik der Schweiz HSSO (2012e)	1851 - 1913	Y
EXPVOLBAI	Export volumes index (Bairoch)	Historische Statistik der Schweiz HSSO (2012a)	1851 - 1913	Y
EXPVOLBER	Export volumes index (Bernegger)	Historische Statistik der Schweiz HSSO (2012a)	1851 - 1913	Y
IMPVAL	Import values	Historische Statistik der Schweiz HSSO (2012b)	1886 - 1992	Y
EXPVAL	Export values	Historische Statistik der Schweiz HSSO (2012b)	1886 - 1992	Y

Continued on next page

Table 3.A.2 – continued from previous page

Publication	Name	Source	Availability	Freq.
IMPVOLM	Import volumes	Historische Statistik der Schweiz HSSO (2012i)	1924 - 1987	M
EXPVOLM	Export volumes	Historische Statistik der Schweiz HSSO (2012g)	1924 - 1967	M
IMPVALM	Import values	Historische Statistik der Schweiz HSSO (2012j)	1924 - 1992	M
EXPVALM	Export values	Historische Statistik der Schweiz HSSO (2012h)	1924 - 1992	M
IMPVALEZV	Import values	Federal Office for Customs and Border Security (FOCBS)	1988 - 2022	M
EXPVALEZV	Export values	Federal Office for Customs and Border Security (FOCBS)	1988 - 2022	M
TradeIndicator	First principal component of trade related indexes	HSSO, FOCBS	1851 - 2022	Y
THORPGPT	Created sentiment indicators (Austria, England, France, Germany, Italy, USA)	Thorp (1926), GPT-3.5	1800 - 1925	Y
HSSOUNEMP	Unemployment rate	Historische Statistik der Schweiz HSSO (2012l)	1913 - 1995	Y
HSSOEMP	Employment growth	Historische Statistik der Schweiz HSSO (2012f)	1890 - 2005	Y
FSOUNEMP	Unemployment rate	FSO	1970 - 2021	M
FSOEMP	Employment growth	FSO	1998 - 2021	M
ECRIREC	Recession indicator	Economic Cycle Research Institute (ECRI)	1956 - 2022	M
OECDREC	Recession indicator	Organisation for Economic Co-operation and Development (OECD), Fred: CHEREC	1960 - 2022	M

Notes: This table shows sources, available periods and frequency of the validation data. Most series are further spliced to a long time series spanning as many periods as possible.

Table 3.A.3 — Keywords defining economic concepts in German

Topic	Keywords (based on readings)	Method
Real activity	wirtschaft ware absatz nachfrage geschäft konsum waare fabrikant erlös umsatz markt industrie branche käufer unternehmer ernte ergebniss konjunktur kundschaft verkauf produktion dienstleistung verarbeitung gewerbe ertrag einnahmen ausgaben fabrikation bestellung versorgung materialbeschaffung einkäufer verlust konkurrenten fabrizieren fabrikat besteller werth neugründung materialien betriebsmittel materialeinkäufe jahresresultat geschäftsperiode werkstätten erfolg bestellungen eigenkosten produkte rohmaterialien einbusse fabric fabrik herstellung geschäftsgang wirtschaftsleben wirtschaftlich nachfrage geschäftslage marge angebot erträgnis rendite produzent vertrieb volkswirtschaft konjunkturrückgang verkaufsziffern kauflust kaufunlust geschäftsjahr kleinbetrieb bautätigkeit verbraucher konkurrenz erzeugnis konsum profit fremdenverkehr dienste dienstleistungen investition versorgungsmöglichkeit versorgungslage versorgungsschwierig bruttoinlandprodukt bruttosozialprodukt realwachstum wertschöpfung	KWIC
Trade	eingangszölle eingangszoll konkurrenzverhältnisse konkurrenzverhältnis einfuhr ausfuhr export import sendungen aussendung importeure exporteure handelsstatistik absatzfeld wettbewerb absatzgebiet zwischenhandel handel importhaus zollverhältnisse zollverhältnis handelsbilanz waarenverkehr warenverkehr waarenausfuhr warenausfuhr waareneinfuhr wareneinfuhr importhandel handelsverkehr zoll zolleinnahmen zölle weltbedarf fracht exportziffern gesammtexport gesamtexport taxen verkehrserleichterung bezugsquelle ausland generaltarif tarif einfuhrverbote zufuhren zufuhr grosshändler seefracht wasserweg welthandel weltverkehr güterstrom güteraustausch güterumschlag umschlagverkehr wagenverkehr	KWIC
Capacity	lager kornspeicher speicher ueberproduktion überproduktion vorräthe vorräte liefertermine lieferfrist vorrat vorrath aufträge lieferfristen lagerware lager depots bestellungen lieferungen wagenmangel	KWIC
Labor	arbeit erwerb beruf erwerbende arbeiter aufsichtspersonal arbeiterin angestellte arbeitskräfte beschäftigung arbeitszeit arbeitgeber arbeiterschaft ueberzeit überzeit arbeitstag arbeitsleistung ausbildung lehrlinge ueberzeitarbeit überzeitarbeit streik arbeitseinstellung arbeitsfeld thätigkeit tätigkeit arbeiterinnen personal	KWIC

Continued on next page

Table 3.A.3 – continued from previous page

Topic	Keywords (based on readings)	Method
Inflation	preis getreidepreise theuer teuer preisfall kostenpreise wechselkurs silberpreis preisaufschlag silberkurs preisbasis kurs preissturz baumwollpreise preisnotierung preisnotirung waarenpreise warenpreise kursschwankungen kostenpreis preissteigerung schleuderpreise abschlag materialpreise maschinenpreise unterbietung preisvorteile preisvortheile verkaufspreise rohpreise preiserhöhung entwerthung entwertung goldkurs geldvertheuerung geldverteuerung geldwerth geldwert vertheuerung verteuerung wechselkurse pari preise teuerung kaufkraft aufschlag inflation inflatorisch	KWIC
Wages	entgelt lohnerhöhung arbeitslöhne löhne arbeitslohn kaufkraft lohnverhältnisse lohnsätze lohnansätze stundenlöhne stundenlohn akkordlöhne akkordlohn akkord tagesverdienst löhnungen löhnung lohnverhältnis verdienstverhältnisse verdienstverältnis einkommen lohnniveau lohnstopp lohnpolitik	KWIC
Credit	kreditverhältnisse banknoten münzen einleger guthaben einzahlung rückzahlung prämien renten kapitalien verzinsung amortisation wechsel geldmarkt zinsen zins rendite diskontsätze disconto discontsätze diskontosätze diskontosatz diskontsatz geldkraft zinsfuss wechselgeschäft wechselkverkehr zinssätze zinssatz diskontopolitik geldinstitute lombardvorschüsse notenemission vorschüsse kontokorrent kreditwirtschaft kredit schuldbriefe kreditwesen leihmarkt emissionsbanken diskonto geldbedarf diskonti diskontoverkehr geldsuchenden emissionen geldverteuerung liquidität geldknappheit geldstand lombardsatz geldleihpreis geldmärkte kapitalmarkt anleihen hypothek depositen darlehen pfandbrief geldpolitik geldmenge	KWIC
Financial	kapitalvermehrung emissionskurs tageskurs agio kurse dividenden emittirt emittiert obligationen rentabilität konversion börsengeschäft kapital werthpapiere wertpapiere titelverkäufe papiere kurssteigerung portefeuille tratten ueberspekulation überspekulation entwerthung entwertung finanzcrisis finanzkrisis börse emission gründung aktie actie aktien kurs effekten märkte wechselkurs devisen valoren dividende wertschriften	KWIC
Recession	crisis krise rezession recession finanzcrisis finanzkrisis spekulationskrise krach krisis zahlungsfähigkeit fallimente bankerottir liegenschaftenkrisis konkurse schaden konjunkturrückgang weltbrand notstand valutasturz depression valutaschwierigkeit wirtschaftskatastrophe schäden liquidation liquidier hemmnisse zusammenbruch notlage katastrophe baisse	Count
Boom	hausse aufschwung hochkonjunktur prosperität erholungsperiode boom	Count

Continued on next page

Table 3.A.3 – continued from previous page

Topic	Keywords (based on readings)	Method
War	krieg konflikt putsch	Count

Notes: The table contains keywords used to identify economic concepts in German. The most right column indicates the method used to identify the keywords. KWIC stands for keyword-in-context, which means that a sentiment score is calculated from the words surrounding the keyword. Count indicates that the indicator is created by counting the appearances of these keywords. Terms embedded in \b are used to avoid counting words that contain the keyword, but are not the keyword itself. For example, \bertrag\b (in english: revenue) is used to avoid counting the word “ertragen” (endure) as a keyword. For the other words, wildcards are used, which means that all words containing e.g. “import” are counted as a keyword.

Table 3.A.4 — Keywords defining economic concepts in French

Topic	Keywords (based on readings)	Method
Real activity	économie marchandise vente \bdemande\b affaire consommation marchandis fabricant revenu horlogerie marchand industrie branche acheteur entrepreneur récolte marcheé conjoncture clientèle production \bservice\b \bachat\b rendement recettes dépenses fabrication commande approvisionnement perte concurrents fabriquer produit commanditaire matériaux achats commandes \bcoûts\b produits fabriq usine fabric économique demande récoltes marge \boffre\b \bmarché\b producteur distribution economie ventes economique \bacheter\b fiscal entreprise consommateur concurrence profit tourisme revenus services investissement produite pib \bcroissance\b	KWIC
Trade	importation exportation export import importateurs exportateurs commerc commerce douani marchandises douane fret taxes tarif approvisionnements approvisionnement grossistes marchandise transbordement	KWIC
Capacity	entrepôt silo stockage surproduction réserves livraison réserve commandes marchandise dépôts livraisons \bstock\b	KWIC
Labor	travail gain métier travailleurs ouvrier personnel ouvrière employée main-d'œuvre emploi employeur formation apprentis grève activité ouvrières	KWIC
Inflation	prix taux cours rabais dévalorisation monétaire renchérissement parité majoration inflation inflationniste	KWIC
Wages	rémunération salaire salaires salariaux accord rémunérations revenu	KWIC
Credit	crédit banque monnaie déposants solde versement remboursement primes rentes capitaux intérêt amortissement change monétaire intérêts rendement taux escompte lombard émission \bavances\b créance argent émissions liquidité capital emprunts hypothèque dépôts prêts	KWIC
Financial	capital spéculation \bémission\b cours agio dividendes émis obligations rentabilité conversion bourse titres papiers portefeuille trattes sur-spéculation dévalorisation financière financi émission fondation action actions \bcours\b effets \bmarchés\b devises valeurs dividende	KWIC
Recession	crise récession krach solvabilité faillites faillite dommages conflagration dépression catastrophe liquidation liquider effondrement baisse	Count
Boom	hausse expansion prospérité rétablissement boom	Count

War	guerre conflit putsch	Count

Notes: The table contains keywords used to identify economic concepts in French. The most right column indicates the method used to identify the keywords. KWIC stands for keyword-in-context, which means that a sentiment score is calculated from the words surrounding the keyword. Count indicates that the indicator is created by counting the appearances of these keywords. Terms embedded in \b are used to avoid counting words that contain the keyword, but are not the keyword itself. For example, \bdemande\b (in english: demand) is used to avoid counting the word "demander" (asking) as a keyword. For the other words, wildcards are used, which means that all words containing e.g. "import" are counted as a keyword.

Figure 3.A.5 — 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

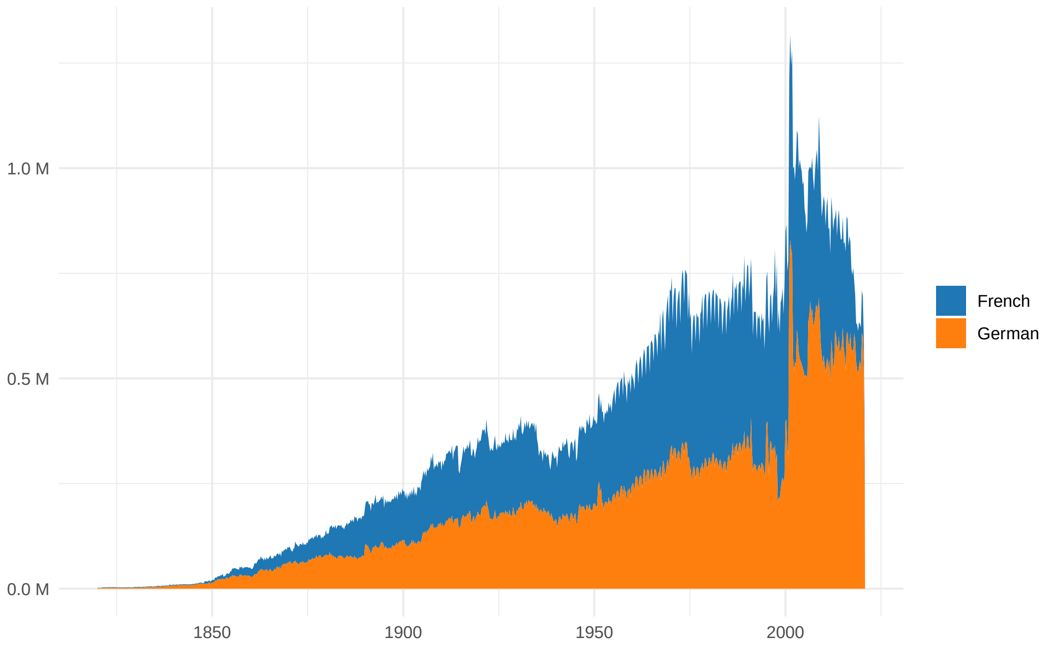
Before cleaning
The corona virus is hitting the Swiss economy with full force from the spring. Economists are predicting a weak second quarter. Consumption, tourism and the export industry are already suffering. The Federal Council should extend short-time work.

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

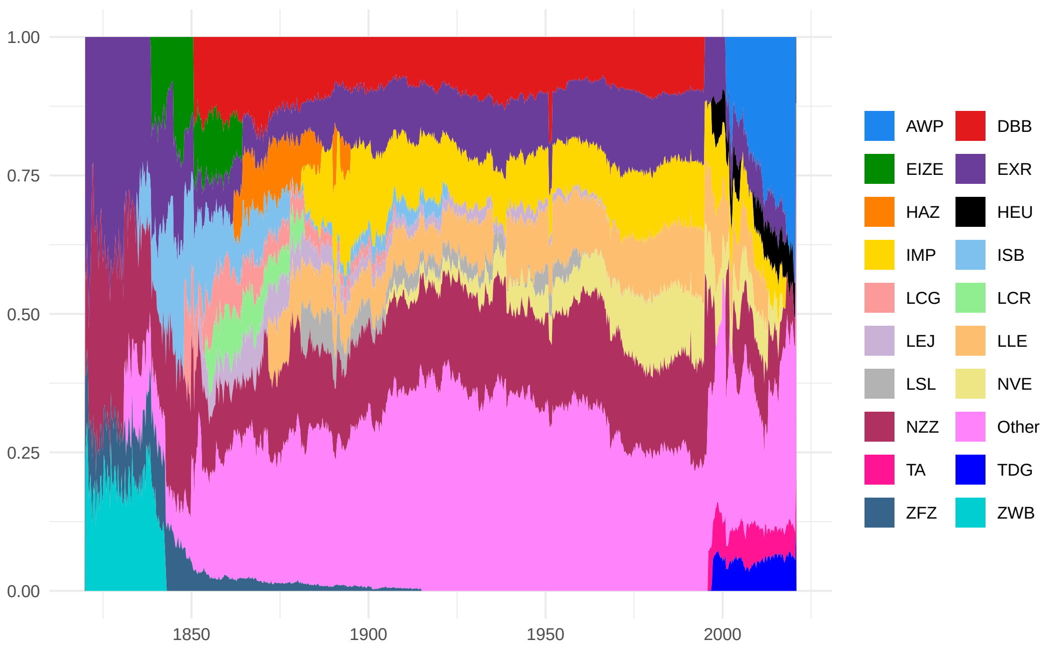
Notes: Example of how document-level sentiment scores for two topics are calculated based on a article teaser from FUW. For the general economy topic that is defined by the keyword (in blue) *wirtschaft*, the number of negative words (in red) is subtracted from the number of positive words (in green) 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 $S_{t,d,\text{economy}} = (1-1)/14 = 0$. Note that there are only 14 words in the denominator because the keyword is close to the beginning of the text. The same method is applied to calculate the sentiment score for other topics, such as the industry topic, which in this example is given by $S_{t,d,\text{industry}} = (1-2)/19 = -0.05$.

Figure 3.A.6 — Number of identified keywords and important sources

a) Number of identified keywords



b) Most important sources



Notes: The graph shows the number of identified keywords for each language in Panel a). Panel b) shows the most important (by most identified keywords) sources and their share.

Table 3.A.5 — Frequency, duration, and number of Swiss compared to European business cycles

	Algorithm	Recession (Peak to trough) 1800-1870	1870-1950	1950-2000	Expansion (Trough to peak) 1800-1870	1870-1950	1950-2000
Frequency	Hamilton (1989)	32.3	25.3	21.5	67.7	74.7	78.5
	Bry and Boschan (1971)	49.3	45.9	36.5	50.7	54.1	63.5
	BL2023 (Swiss GDP)	35.0	32.5	14.0	65.0	67.5	86.0
	ECRI			19.2			80.8
	BL2023 (European GDP)	38.0	28.4	5.9	62.0	71.6	94.1
Duration	Hamilton (1989)	7.6	6.5	7.2	17.0	19.6	23.0
	Bry and Boschan (1971)	11.4	15.1	14.4	14.6	16.9	21.8
	BL2023 (Swiss GDP)	5.6	6.5	7.0	9.6	13.2	36.0
	ECRI			8.7			32.1
	BL2023 (European GDP)	6.4	7.2	4.0	9.6	17.2	54.8
Number	Hamilton (1989)	8.0	13.0	6.0	9.0	12.0	6.0
	Bry and Boschan (1971)	7.0	11.0	5.0	8.0	10.0	6.0
	BL2023 (Swiss GDP)	5.0	16.0	4.0	5.0	16.0	4.0
	ECRI			4.0			4.0

Notes: Frequency is the share of quarters in a given phase in percent. Duration is the average number of quarters in a given phase. Dating based on Hamilton (1989) and Bry and Boschan (1971) starts in 1820, real Swiss GDP growth is available from 1852, and the ECRI dating starts in 1956. Broadberry and Lennard (2023) (BL2023 European GDP) classifiy recessions as negative GDP per capita growth.

Table 3.A.6 — Frequency, duration, and number of Swiss compared to UK business cycles

	Algorithm	Recession (Peak to trough) 1817-1908	1909-1947	1948-2009	Expansion (Trough to peak) 1817-1908	1909-1947	1948-2009
Frequency	Hamilton (1989)	33.4	13.5	27.4	66.6	86.5	72.6
	Bry and Boschan (1971)	51.6	35.9	39.9	48.4	64.1	60.1
	BL2023 (Swiss GDP)	25.9	43.6	17.7	74.1	56.4	82.3
	ECRI			21.9			78.1
	B2023 (UK GDP)	22.8	25.6	11.3	77.2	74.4	88.7
Duration	Hamilton (1989)	6.9	7.0	7.6	14.6	35.7	23.2
	Bry and Boschan (1971)	14.0	13.5	12.6	14.1	20.4	21.1
	BL2023 (Swiss GDP)	5.0	8.5	6.3	13.7	10.7	33.3
	ECRI			7.9			27.7
	B2023 (UK GDP)	5.2	10.0	7.2	17.6	29.2	55.2
Number	Hamilton (1989)	17.0	3.0	9.0	17.0	3.0	9.0
	Bry and Boschan (1971)	13.0	4.0	8.0	13.0	5.0	7.0
	BL2023 (Swiss GDP)	12.0	8.0	7.0	12.0	8.0	7.0
	ECRI			6.0			6.0
	B2023 (UK GDP)	16.0	4.0	4.0	16.0	4.0	4.0

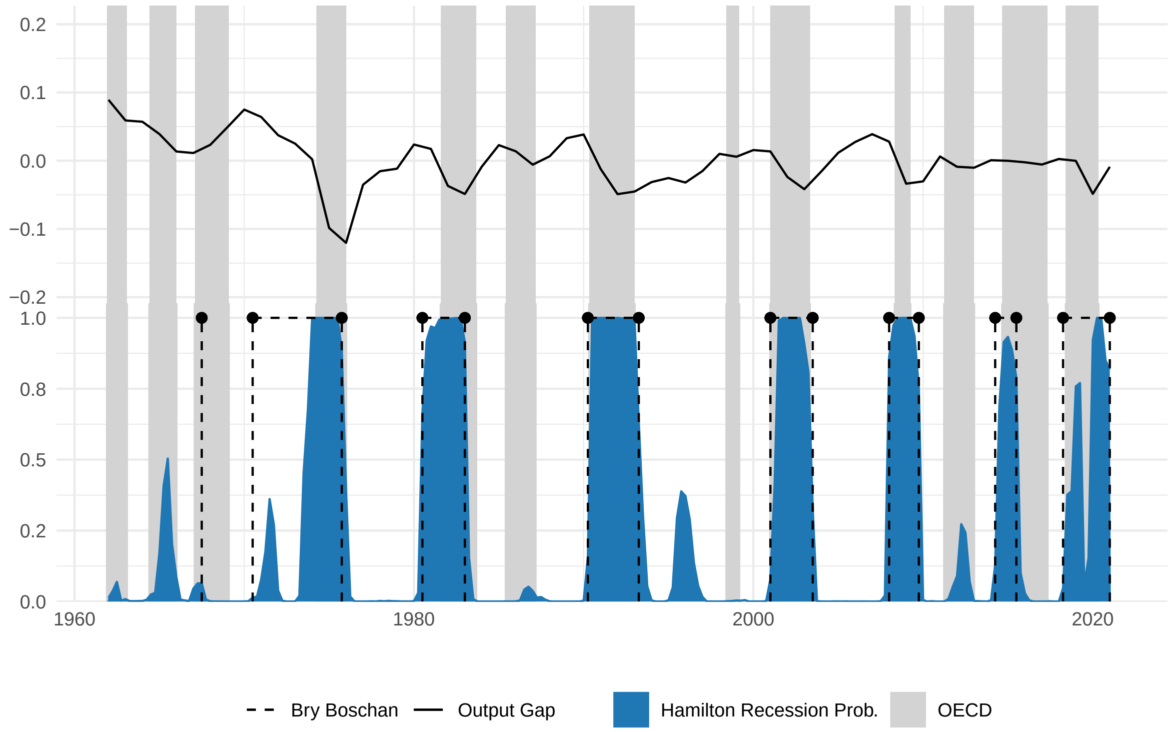
Notes: Frequency is the share of quarters in a given phase in percent. Duration is the average number of quarters in a given phase. Dating based on Hamilton (1989) and Bry and Boschan (1971) starts in 1820, real Swiss GDP growth is available from 1852, and the ECRI dating starts in 1956. Broadberry et al. (2023) (B2023 UK GDP) use discretion to classify phases in economic activity. BL2023 classifies recessions as negative GDP per capita growth.

Table 3.A.7 — Coincident correlations of topic-specific indicators

	Boom	Capacity	Credit	Financial	Inflation	Labor	RealActivity	Recession	Trade	Wages
Boom										
Capacity	0.14***									
Credit	0.08**	0.30***								
Financial	0.02	0.37***	0.46***							
Inflation	0.11***	0.31***	0.43***	0.55***						
Labor	0.09**	0.36***	0.43***	0.43***	0.47***					
RealActivity	0.23***	0.60***	0.50***	0.56***	0.58***	0.67***				
Recession	-0.11***	0.06*	0.20***	0.35***	0.33***	0.40***	0.35***			
Trade	0.20***	0.42***	0.37***	0.46***	0.44***	0.43***	0.62***	0.23***		
Wages	0.11***	0.26***	0.34***	0.41***	0.35***	0.53***	0.51***	0.31***	0.34***	
War	0.06	0.30***	0.09**	0.18***	0.05	0.02	0.29***	-0.13***	0.26***	0.03

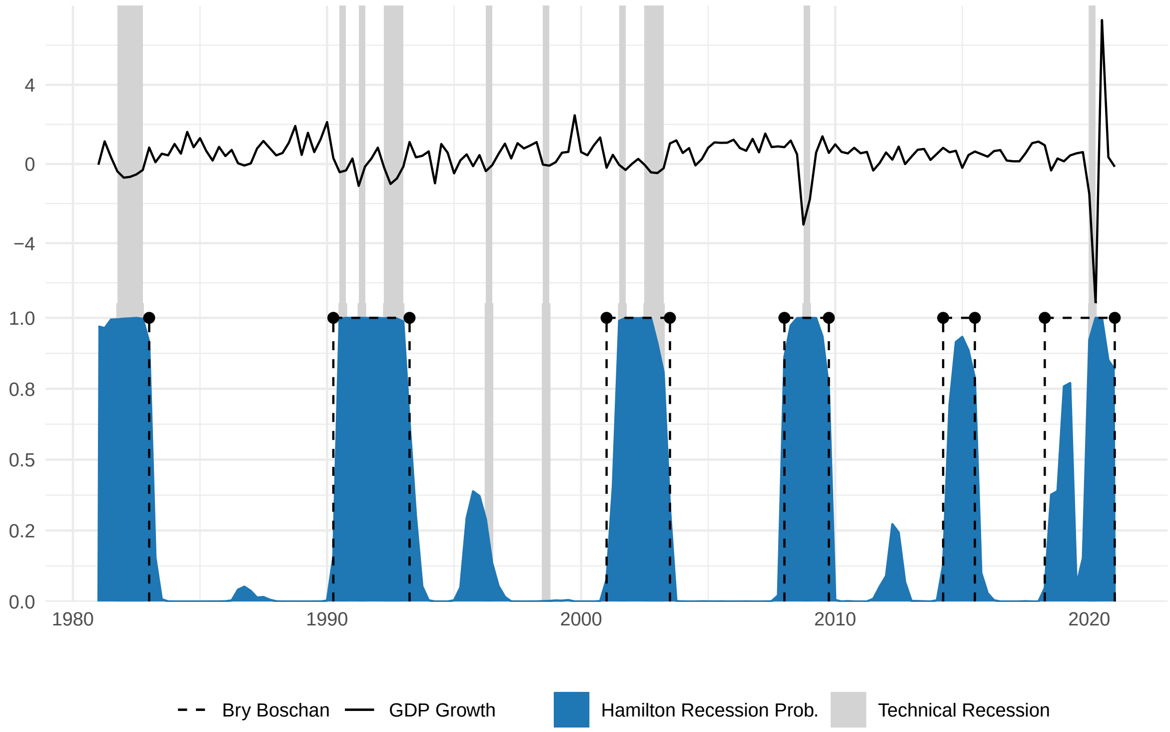
Notes: The table shows the coincident correlations between the topic-specific indicators. All indicators are scaled to have a positive correlation with the output gap. Significance levels are as follows: ***: 1%, **: 5%, *: 10%.

Figure 3.A.7 — Comparison with OECD dating



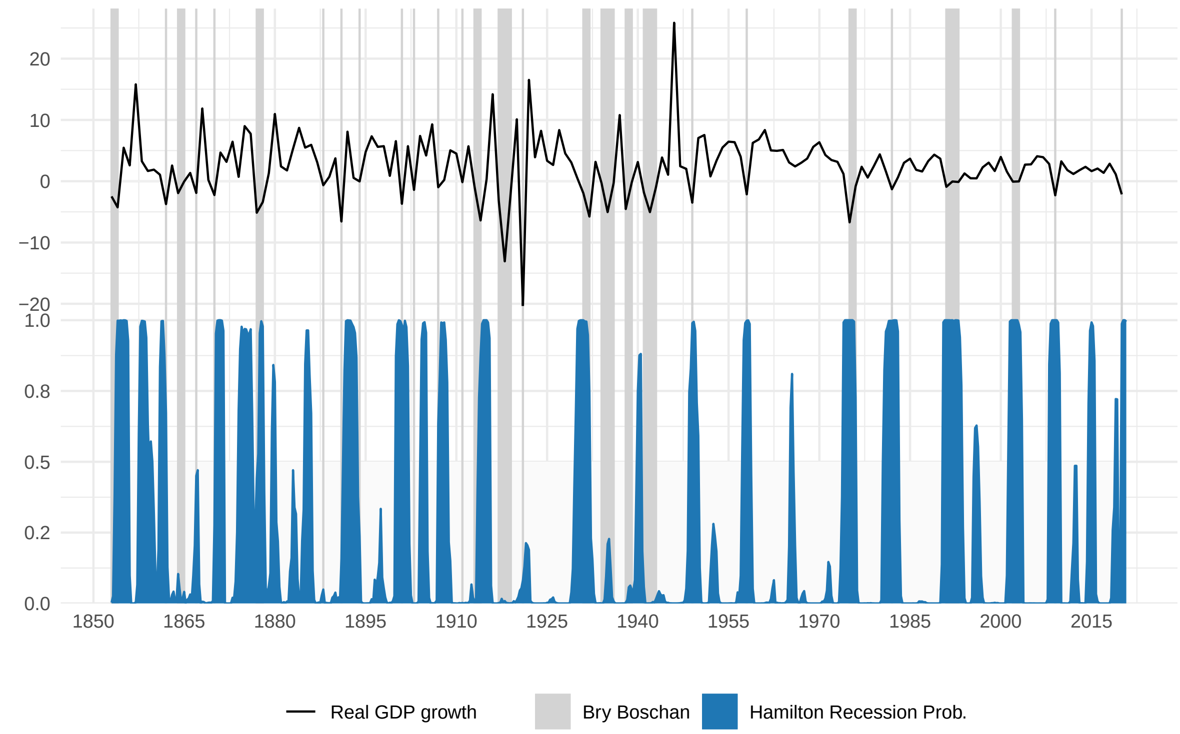
Notes: This graph shows the output gap given by the black solid line together with smoothed recession probabilities obtained by Hamilton's (1989) Markow-Switching autoregression model. The black dashed lines indicate recessions obtained by an adapted Bry and Boschan (1971) algorithm. The gray-shaded areas are recession dates obtained from the OECD.

Figure 3.A.8 — Comparison with technical recession



Notes: This graph shows quarterly real GDP growth given by the black solid line together with smoothed recession probabilities obtained by Hamilton's (1989) Markow-Switching autoregression model. The black dashed lines indicate recessions obtained by an adapted Bry and Boschan (1971) algorithm. The gray shaded areas indicate technical recessions.

Figure 3.A.9 — Comparison with BL2023 algorithm based on annual GDP growth



Notes: This graph shows annual real GDP growth given by the black solid line and smoothed recession probabilities obtained by Hamilton's (1989) Markow-Switching autoregression model. The gray shaded areas indicate contractions obtained with the algorithm proposed by Broadberry and Lennard (2023).

Table 3.A.8 — Quarterly turning points in Switzerland, 1821 - 2021

Censored

Peak	Trough
1821 Q1	1823 Q3
1825 Q3	1827 Q2
1842 Q1	1843 Q3
1845 Q1	1845 Q2
1847 Q4	1850 Q2
1853 Q4	1855 Q3
1857 Q3	1859 Q4
1861 Q1	1861 Q4
1870 Q2	1871 Q3
1873 Q4	1875 Q3
1877 Q1	1878 Q1
1879 Q4	1879 Q4
1884 Q3	1885 Q4
1891 Q2	1893 Q3
1900 Q1	1902 Q1
1904 Q1	1905 Q1
1907 Q1	1908 Q2
1914 Q1	1915 Q3
1929 Q4	1932 Q1
1939 Q4	1940 Q3
1948 Q3	1950 Q1
1957 Q3	1958 Q3
1965 Q2	1965 Q3
1974 Q1	1975 Q4
1980 Q3	1983 Q1
1990 Q3	1993 Q3
1995 Q3	1996 Q1
2001 Q2	2003 Q2
2008 Q1	2009 Q3
2014 Q2	2015 Q3
2019 Q3	2021 Q1

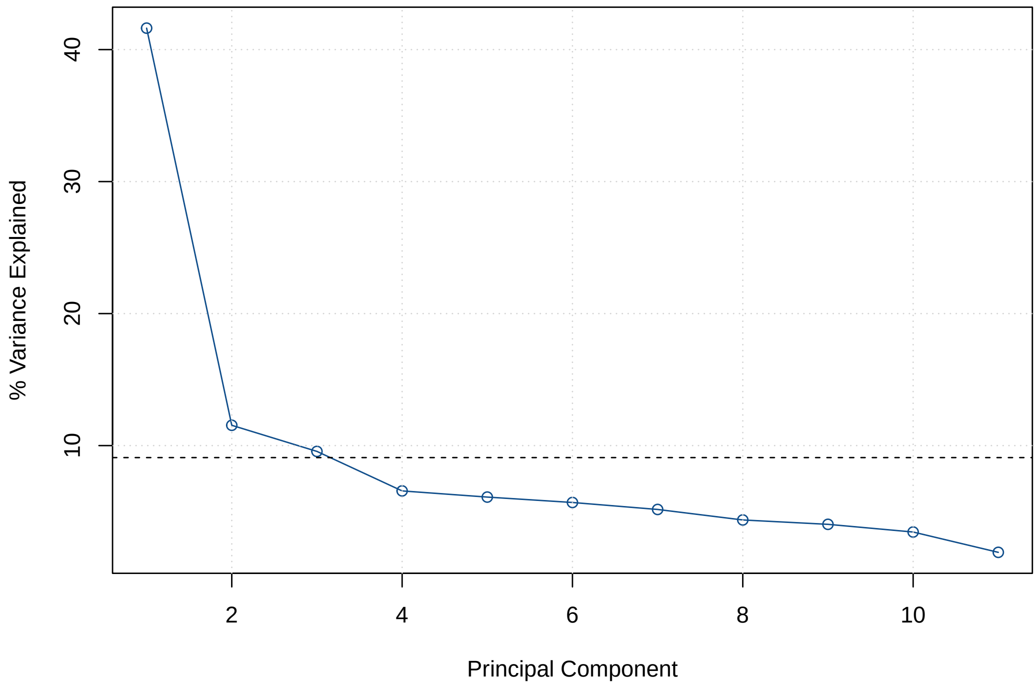
Notes: Quarterly truning points of Swiss business cycles obtained from Hamilton's (1989) Markov-switching autoregression model with censoring (i.e. identified recessions where the underlying indicator is positive over the whole time span are removed).

Not censored

Peak	Trough
1821 Q1	1823 Q3
1825 Q1	1826 Q3
1842 Q2	1843 Q2
1848 Q1	1850 Q2
1853 Q4	1855 Q4
1857 Q3	1859 Q1
1859 Q3	1859 Q4
1861 Q1	1862 Q1
1870 Q2	1871 Q3
1874 Q1	1876 Q2
1877 Q2	1878 Q1
1879 Q3	1880 Q1
1885 Q1	1886 Q1
1891 Q3	1893 Q3
1900 Q1	1902 Q1
1904 Q2	1905 Q1
1907 Q1	1908 Q3
1913 Q4	1915 Q3
1929 Q4	1932 Q1
1940 Q1	1940 Q3
1948 Q3	1950 Q1
1957 Q2	1958 Q3
1965 Q2	1965 Q3
1974 Q1	1976 Q1
1980 Q3	1983 Q1
1990 Q3	1993 Q3
1995 Q4	1996 Q2
2001 Q2	2003 Q3
2008 Q1	2009 Q4
2014 Q3	2015 Q3
2019 Q1	2019 Q2
2020 Q1	2021 Q1

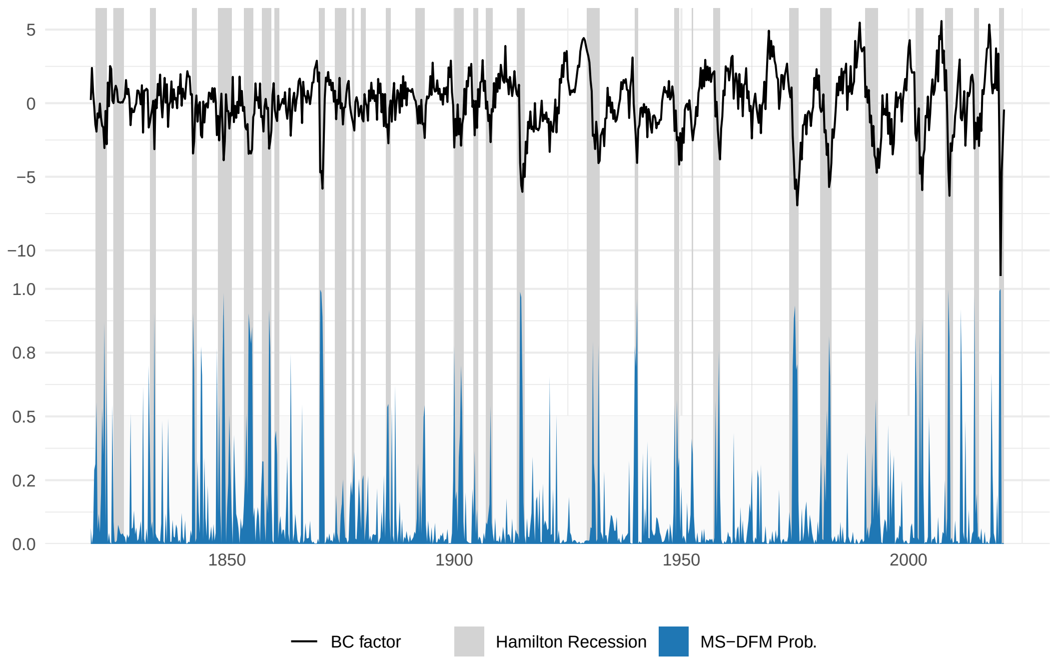
Notes: Quarterly truning points of Swiss business cycles obtained from Hamilton's (1989) Markov-switching autoregression model.

Figure 3.A.10 — Scree plot for MS-DFM used for robustness



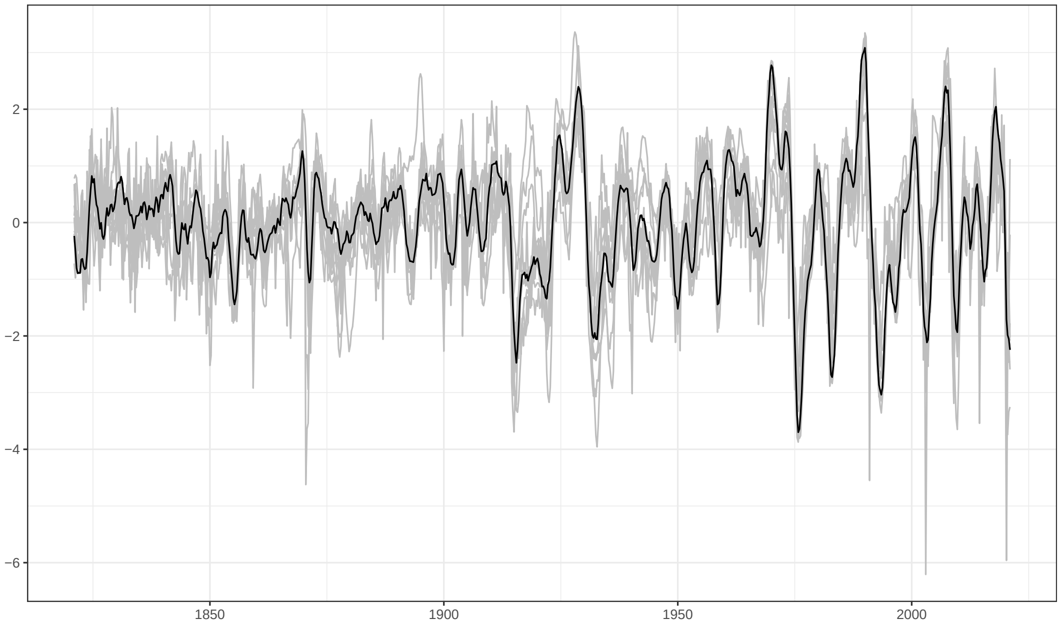
Notes: This graph shows the marginal contribution of each additional factor to R^2.

Figure 3.A.11 — Recession probabilities derived from MS-DFM



Notes: This graph shows the business cycle factor (black solid line) together with smoothed recession probabilities (blue) obtained with the MS-DFM. The gray shaded areas indicate recessions obtained with Hamilton's (1989) Markow-Switching autoregression model with the baseline indicator.

Figure 3.A.12 — Comparison to indicators based on other aggregation techniques, adjustment methods, and subsets of data.



Notes: This graph shows the baseline indicator (black solid line) together with several indicators based on other aggregation techniques, adjustment methods, and subsets of data (gray lines). See section 3.5 for more details.

Appendix 3.B Data collection

This Appendix first discusses how I processed the scanned images, put them in machine-readable format, and parsed their layout. Moreover, I present the data to validate the new business cycle indicator.

3.B.1 Textual data

The indicators for measuring business cycle fluctuations are based on textual analysis of historical documents.[25] Therefore, I collected and digitized a large body of historical documents that potentially comprise useful information on business cycle fluctuations. Annual reports of companies and reports of national and regional business associations stem mainly from the Wirtschaftsarchiv Basel, whereas newspapers primarily stem from e-newspaperarchives.ch and scriptorium.bcu-lausanne.ch. Most of the newspapers are already in machine-readable format. Therefore, I only had to apply optical character recognition (OCR) to the documents that I scanned myself and those where the OCR is of poor quality.[26] In total, I collected 105 sources[27] in German and French language.

Scanning documents

I only selected documents to scan that likely comprise information about the business cycle. These either concern a textual description of a company's situation (predominantly annual reports of companies) or a description of the state of the economy in different sectors (primarily reports of business associations). Figures 3.A.2-3.A.3 show examples of scanned pages from the annual report of the "Chambre vaudoise du commerce et de l'industrie" (1923) and the insurance company Helvetia (1860). From a substantive point of view, the paragraphs comprise a discussion of the weak industry sector and an assessment of the business situation conditional on the difficult political situation of the insurance company. Thus, firms and business associations regularly judge their economic situation and the overall situation of the economy in their writing.

[25] See table 3.A.1 in the Appendix for a comprehensive list of all data sources. Figure 3.1 provides information on how many sources were available for each period.

[26] Already digitized data are in PDF or METS/ALTO format. METS/ALTO files are digital files used to store information about documents, such as books or articles, in a structured way. METS describes the document's structure, like its chapters or pages. At the same time, ALTO contains the actual text of the document along with details about its layout, formatting, and coordinates of each element on the page. Especially for larger documents with many pages, these files can be resource-intensive to process with statistical software. As a result, loading METS/ALTO files can take a long time. I developed R routines to read METS/ALTO files in parallel, which might be helpful to researchers. Find more information here: https://marcburri.github.io/posts/2023/09/11/mets-alto-r/.

[27] A source is a publication over a more extended period. This can be a newspaper (e.g., NZZ, 1820 - 2020) or a business report (e.g., Credit Suisse, 1895 - 2016)

Image pre-processing

After scanning these reports, the files must be prepared for text recognition. I, therefore, set up an image pre-processing procedure that facilitates the quality of the scans and makes it easier to convert the documents into machine-readable format. This step includes cropping the images, turning them into black and white (eventually erasing shimmering text from the back side), removing the curvature of the text (stemming from bent book pages), removing speckles, and sharpening the contrast of the documents.[28] Figures 3.A.2-3.A.3 show that the quality of the pre-processed image is improved, which in turn improves the following OCR.

Optical text recognition (OCR)

Converting the images into machine-readable text format is a crucial step. Therefore, I use the Abbyy FineReader software based on machine-learning techniques, widely used for larger-scale digitizing projects. Figures 3.A.2 - 3.A.3 show the recognized text with that software. From a technical point of view, they demonstrate that it is possible to recognize text in high quality from different fonts and languages. We can readily convert regular font and gothic type into machine-readable text. However, the examples are not perfect. In particular, the software has trouble recognizing the punctuation marks correctly. But this is not a severe problem because they will be deleted in a further step anyway. Overall, these examples demonstrate, however, that it is possible to accurately convert scanned documents' images into text.

Layout parsing

We have different types of publications. Therefore, the length of texts differs. Annual reports might potentially be several hundred pages long. On the other hand, newspaper articles seldom exceed one page. Moreover, some advertisements and tables do not contain valuable information. Therefore, I also use the Abbyy software to parse the layout of publications, which has not been done before. Figure 3.A.4 shows an example of a parsed layout. Titles, paragraphs, and tables can be easily identified. This has several advantages. First, long texts from annual reports can be split into shorter paragraphs. Many of these do not contain any useful information. Second, tables and figures can be identified and filtered out.

[28] I mostly used the CamScanner application for mobile devices to conduct these steps. In a few cases, I also used ScanTailor.

3.B.2 Validation data

Validating the indicator is difficult, especially for the 19$^{\text{th}}$ century when data on real economic activity is scarce and inaccurately measured. Nevertheless, I collect as much real activity data as possible covering the 19$^{\text{th}}$ century to calculate correlations and validate the indicator's accuracy. In this section, I briefly describe the collected data for validation and explain why it should be correlated with the business cycle indicator. An extensive overview and data sources are listed in Table 3.A.2 in the Appendix.

Data on real activity

The indicator developed in this chapter is a measure of the business cycle. Therefore, it should be correlated with real economic activity. Real GDP is one of the most widely used measures of real economic activity. However, GDP growth rates only show how much the economy has grown from one period to the next, not whether that growth is above or below the economy's potential. Therefore, I mainly use the output gap as a validation measure. The output gap is the difference between actual and potential GDP. It measures the extent to which the economy operates above or below its potential. The output gap is a useful indicator of the health of the economy and the degree of inflationary pressures.

The output gap is calculated as suggested by Hamilton (2018). That is, I estimate an OLS regression of GDP in logs, y_{t+h}, on a constant and the $p = 2$ most recent values of y as of date t,

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + v_{t+h} \tag{3.13}$$

where the residuals

$$\hat{v}_{t+h} = y_{t+h} - \hat{\beta}_0 - \hat{\beta}_1 y_t - \hat{\beta}_2 y_{t-1} \tag{3.14}$$

offer a reasonable way to construct the cyclical component, the output gap. As suggested by Hamilton (2018), $h = 2$ for annual data.

For Switzerland, annual GDP data in real and nominal terms for the 19$^{\text{th}}$ century is presented in Historische Statistik der Schweiz HSSO (2012k). The authors state that some sector estimates of value added have "the characteristics of a rough estimate at best". For instance, in relation to estimates of the wholesale and retail value-added, they "did not have sufficient data to reliably estimate the value added of this important branch". Based on this dataset, Stohr (2016) estimates an improved GDP series by refining the estimation of certain branches.[29] This estimate is a crucial reference point for the business cycle indicator. As the reference series, I use Stohr's (2016) single-deflated estimate from 1851-1890 and the double-deflated estimate from 1890 to 1947.[30] For the period from 1948 to 2022, I use the official GDP figures from the Swiss Federal Statistical Office (FSO).

Because Swiss GDP is not accurately measured for the 19th century and business cycles tend to be international, I also compare the business cycle indicator to estimates for real GDP growth and the output gap of various countries. In particular, to Switzerland's neighboring countries, the USA and England. The data is taken from the Maddison database (Bolt & van Zanden, 2020). Finally, I also use world GDP, Swiss GDP per capita, trade data, and Swiss labor market data as validation measures.

GPT-3.5 meets Business Annals: A new take on 19th century sentiments

The validation indicators discussed above are hard data, likely measured with error in the early periods. My business cycle indicator, in contrast, is a qualitative indicator based on sentiments from textual analysis. There is, however, no existing qualitative indicator for the 19th century that I can use for validation. In his book, Thorp (1926) provides a detailed narrative account of the business cycle in selected countries for every single year up to 1925. He describes the business cycle in terms of the state of the economy, the financial situation, and the state of the labor market. I, therefore, use his account to create a qualitative sentiment indicator for Switzerland's neighboring countries, England and the USA, using large language models (LLM) as validation measures.[31] In particular, I use OpenAI's (2023b) GPT-3.5 model and feed it repeatedly with ten years of Thorp's (1926) descriptions and ask it to rate the state of the economy in a given year from -5 to 5. The mean overall ratings serve as the sentiment indicator. Because the model does not always produce the same rating, the standard deviation over all estimates can be interpreted as confidence bands. The following section provides more details on creating these sentiment indicators.

Qualitative business cycle indicators based on consumer or business surveys became

[29]This series is available in the Maddison project database (Bolt & van Zanden, 2020).
[30]See Stohr (2016) and Stohr (2017) for more information about single- and double-deflated GDP
[31]Unfortunately Thorp (1926) does not provide a narrative account for Switzerland.

popular in the mid-20th century. Before, however, there were no systematic attempts to measure the business cycle using surveys. To overcome this, I propose to use Thorp's (1926) description of the state of the economy together with OpenAI's (2023b) GPT-3.5 language model to construct economic sentiment indicators for several countries in the 19th and early 20th centuries. In this section, I show step-by-step how this can be done.

Thorp's (1926) book Business Annals is a seminal work in the field of financial history and economic analysis. In this book, Thorp (1926) meticulously compiles and analyzes historical data related to business cycles, financial situations, and economic fluctuations. He aims to record business conditions and their impact on financial markets comprehensively. He provides a detailed narrative account of the business cycle in selected countries for every year from around 1830 to 1925. Thorp's work laid the foundation for systematically studying economic business cycles (Burns & Mitchell, 1946). Therefore, I use his account to create a qualitative sentiment indicator for Switzerland's neighboring countries, England and the USA.

GPT (Generative Pre-trained Transformer) models are a type of artificial intelligence model used for natural language understanding and generation tasks. GPT models work by pre-training on a large corpus of text to learn the patterns, grammar, and semantics of language. During pre-training, they learn to predict the next word in a sentence, which helps them capture contextual information. They generate text by probabilistically predicting the most likely next word based on the input context and the knowledge they have learned during pre-training (Brown et al., 2020).

> **Prompt**:
> Given the short economic descriptions from the year {start_year} to {end_year}, rate the state of the {country} economy from the text below on a scale of -5 (worst) to 5 (best). Take into account the surrounding years for relative judgment. Return a JSON with the year and your rating only. Text: {texts}

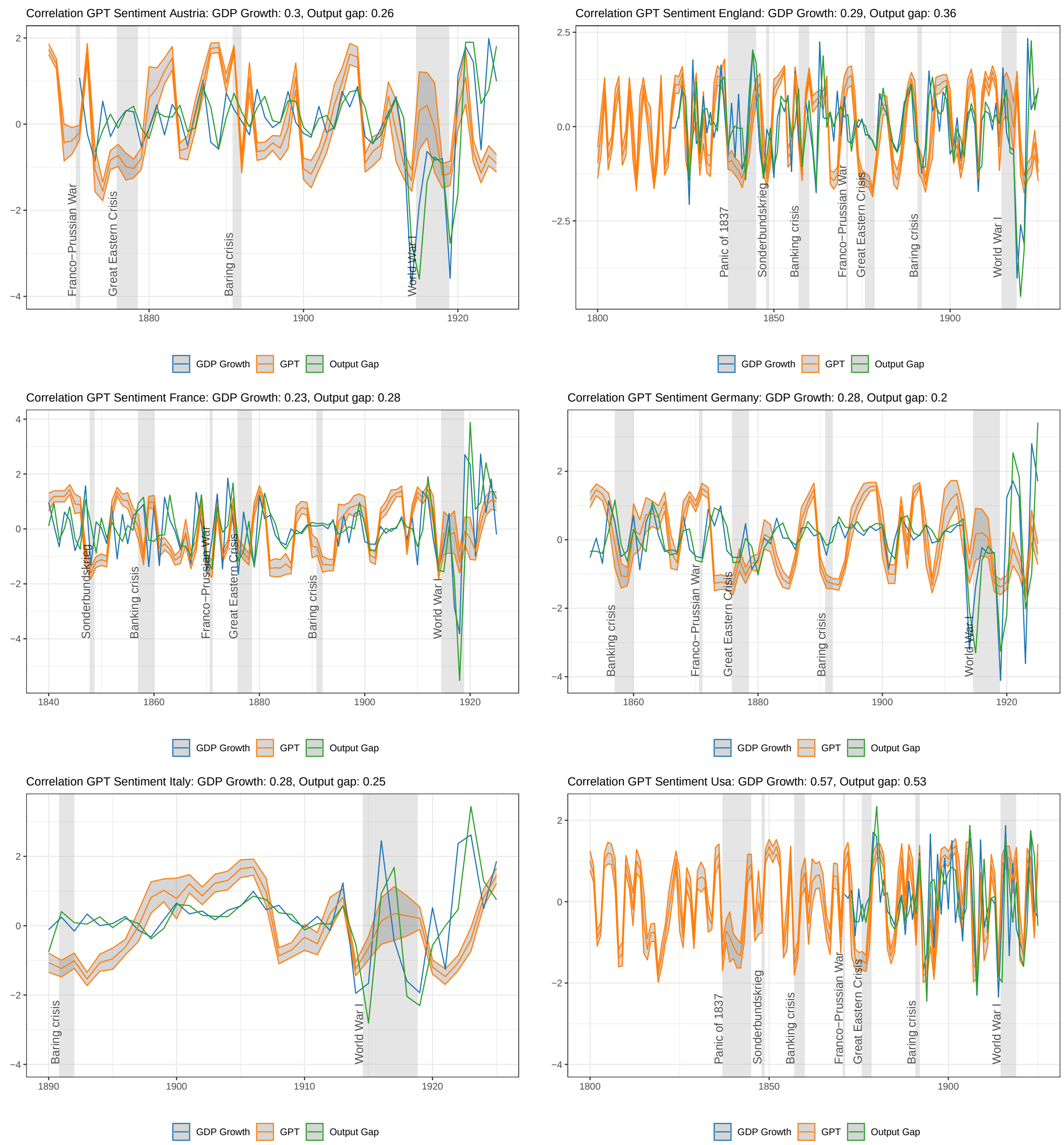
To create the sentiment indicators, I ask the model to rate the state of the economy on a scale of -5 (worst) to 5 (best) for every year until 1925. For relative judgment, the surrounding years should be taken into account. The exact prompt is depicted above. However, GPT-3.5 has a limited context window of 4'096 tokens. This means the model can roughly consider the last 3'000 words when generating the next word. Therefore, using the model in one batch is impossible because the text of Thorp's (1926) book is too long. Instead, I split the text into chunks of ten years and move forward by five years.

Hence, I get two ratings for every year. The average of the two ratings is the final rating for the year.

Providing the model with the same prompt multiple times does not necessarily produce the same output. This behavior is governed by the parameter 'temperature', which controls the randomness of the output. With a value of zero, the model becomes deterministic. I use the default value of one, which leads to a more diverse output. To get a more robust rating, I prompt the model 20 times and take the average of the 20 ratings. Figure 3.B.1 shows the normalized sentiment indicator (in orange) and one standard deviation confidence bands. Using the same degree of randomness every time allows for the interpretation of confidence bands as a measure of uncertainty. Most often, the model is quite confident about the state of the economy. However, during World War I, confidence bands become wider.

Moreover, I compare the GPT sentiment indicators with real GDP growth (blue line) and the output gap (green line) calculated from the series provided by Bolt and van Zanden (2020). The correlation between the sentiment indicator and real GDP growth ranges from 0.23 to 0.57 for the countries under consideration. With the output gap, the correlation ranges from 0.2 to 0.53. The correlation with neighboring countries is somewhat lower than with USA or England. This might be because these two countries have a lot of good quality data available for the 19^{th} century. The sentiment indicator is less erratic than real GDP growth and the output gap and, therefore, serves as an excellent alternative measure to validate the historical business cycle indicator.

Figure 3.B.1 — Thorp GPT sentiment compared to GDP growth and output gap



Notes: These graphs show the normalized sentiment indicator (in orange) based on Thorp's (1926) texts and OpenAI's (2023b) GPT-3.5 model together with one standard deviation confidence bands. The blue line is real GDP growth, and the green line is the output gap, calculated from the series provided by Bolt and van Zanden (2020). Gray shaded areas represent crises.

Appendix 3.C Source-level indicator anomaly correction

In this section, I explain the six steps of the procedure to correct for anomalies in the source-level indicators in more detail. Before proceeding with the six steps, texts where less than 20% of the words are identified in a German-French lexicon are removed. The remaining texts are then aggregated to either quarterly or annual frequency. Texts from sources published more frequently than once per quarter are aggregated quarterly, while texts from sources with a publication frequency between quarterly and annually are aggregated annually. Count-based indicators are aggregated by summing the number of identified keywords, and sentiment-based indicators are aggregated by averaging the sentiment scores. I use examples of the indicator based on the *Credit Suisse Bulletin* (CS_BUL) published annually and the *Neue Zürcher Zeitung* (NZZ) published daily. For illustration, I show the sentiment-based real activity and the count-based recession topics. The same procedure is applied to all other source-level indicators. Figures 3.C.1 (unadjusted indicators) and 3.C.2 (adjusted indicators) show that the correlation with a measure for the business cycle increases substantially after applying the procedure. The following are the six steps:

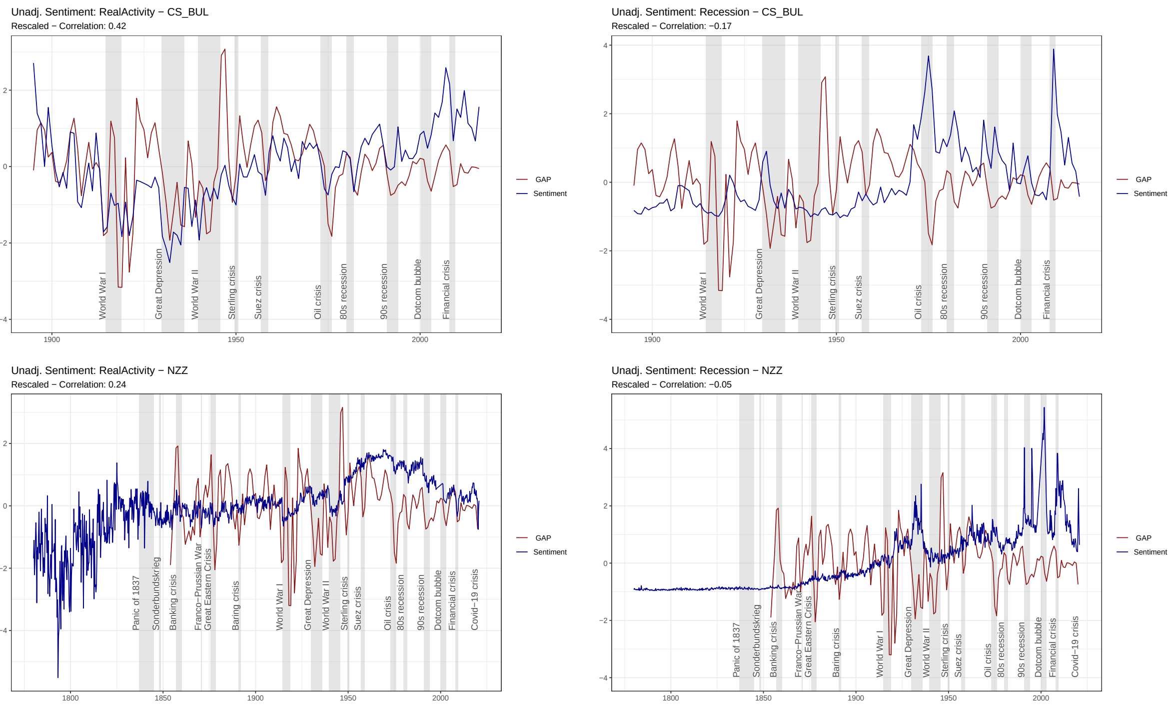
1. If the frequency is annual, interpolate missing observations using Stinemann interpolation (Stineman, 1980). This step is needed to ensure it is a regular time series before temporally disaggregating annual data to quarterly frequency.

2. Detrend the indicator using Locally Estimated Scatterplot Smoothing (LOESS) with bandwidth (or span) of 0.7 (Cleveland, 1979). LOESS (Locally Estimated Scatterplot Smoothing) is a statistical technique to create a smooth line through a scatterplot. This method selects a subset of data points and fits a local model, such as a linear or nonlinear function, to these points. This procedure is iteratively applied to each subset of the data, with the fitted models adapting to the specific characteristics of each section. The bandwidth, typically between 0 and 1, represents the proportion of the total data points in each local fit. A smaller bandwidth produces a more flexible, wiggly line, while a larger one produces a smoother line. A bandwidth of 0.7 is generally considered an effective compromise, balancing smoothness and adherence to the data points.

3. Detect structural breaks in mean and/or variance using a binary segmentation algorithm.[32] This algorithm originates from the work of Edwards and

[32] I use the implementation of Killick and Eckley (2014).

Figure 3.C.1 — Unadjusted source-level indicators



Notes: These graphs show the unadjusted source-level indicator (in blue) together with the output gap (in red). Gray shaded areas represent crises. Credit Suisse Bulletin at the top, NZZ at the bottom. Left the real activity topic, right the recession topic.

Cavalli-Sforza (1965), Scott and Knott (1974) and Sen and Srivastava (1975). The multiple parameter changepoint problem has been discussed by Horvath (1993) or Picard et al. (2005). The process initiates with a single changepoint test applied across the entire dataset to detect significant mean and/or variance changes. Upon identifying a changepoint, the data is split into two at the changepoint location. This test is then recursively conducted on each segment, continually splitting them at newly identified change points. The procedure repeats until no further significant changepoints are detected, adhering to criteria such as a predefined maximum number of changepoints or a minimum length for the segments. The outcome of this method is a division of the dataset into segments. The objective function the algorithm minimizes is given by

$$\sum_{i=1}^{m+1} \left[\mathcal{C} \left(y_{(\tau_{i-1}+1):\tau_i} \right) \right] + \beta m \tag{3.15}$$

where m is the number of breakpoints, $\mathcal{C}$ is a cost function (here, the negative log-likelihood), and βm is a linear penalty function. τ_i is the position of the i-th changepoint in the data. I assume the segments are normally distributed. Moreover, I allow a maximum of one breakpoint every ten years, with an overall cap of five breakpoints. Additionally, each segment is constrained to a minimum length of five years. This approach is based on the premise that the OCR quality of texts changes infrequently. Incorporating such prior information into the algorithm enhances its suitability for this particular application.

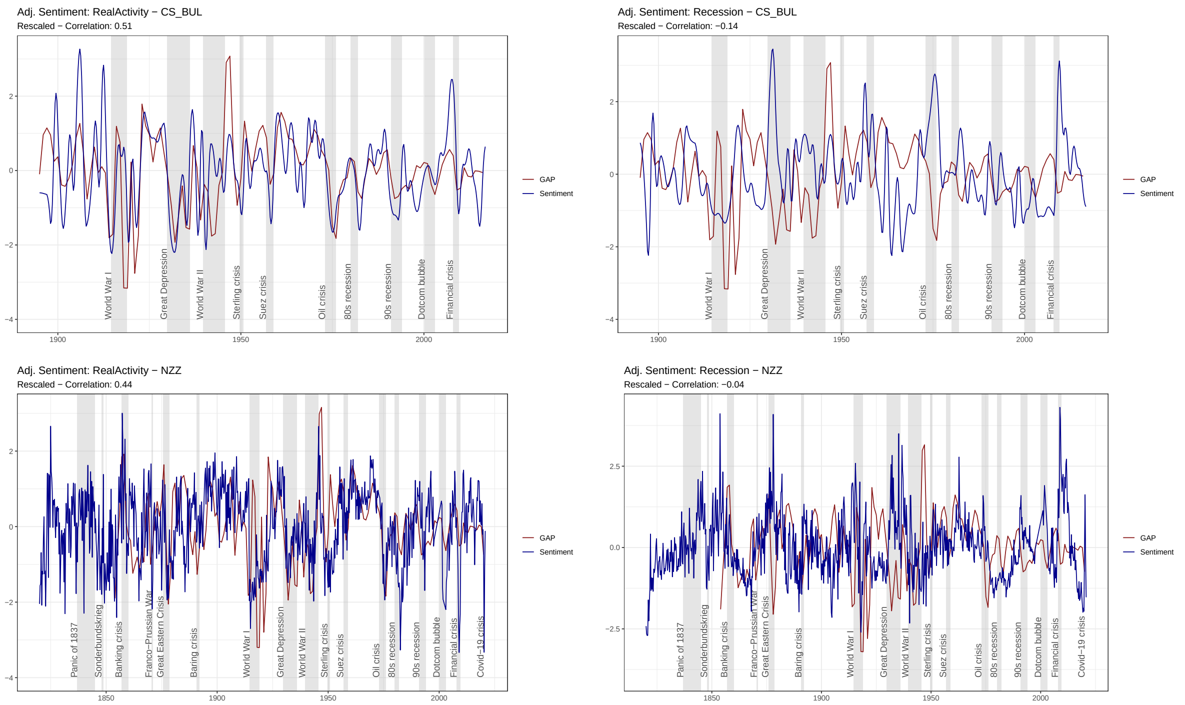
4. Normalize each identified segment.[33] The normalization is done by subtracting the mean and dividing by the standard deviation of the segment.

5. If the frequency is annual, temporally disaggregate the indicator to quarterly frequency using the method developed by Dagum and Cholette (2006), which is based on work by Denton (1971). I employ a straightforward method that does not require an indicator series. This method executes basic interpolation, adhering to the temporal additivity constraint as outlined by Sax and Steiner (2013). The temporal additivity constraint applied in this study mandates that the sum of the series at a lower frequency for count-based indicators must equal the corresponding series at a higher frequency. For sentiment-based indicators, it is the mean.

6. Remove outliers (observations more than three standard deviations away from the mean).

[33] I show in Appendix 3.D that with time-varying measurement error, this is preferable to not normalizing.

Figure 3.C.2 — Adjusted source-level indicators



Notes: These graphs show the adjusted source-level indicator (in blue) together with the output gap (in red). Gray shaded areas represent crises. Credit Suisse Bulletin at the top, NZZ at the bottom. Left the real activity topic, right the recession topic.

Appendix 3.D Normalization of time series with breaks

In this section, I show that normalizing data around identified breaks within the context of the dataset at hand is beneficial, particularly when dealing with a small number of available sources at the beginning of the sample.

The availability of data sources varies over time, with earlier periods having fewer digitized sources than later ones. Moreover, the measurement error associated with these source-level indicators is not static - it changes over time and tends to be larger in earlier periods. In most cases, the changes in the measurement error are abrupt and lead to structural breaks in the series. For instance, due to a sudden change in the optical character recognition (OCR) quality or because the frequency of the publication changes - increased publication frequency means a higher information density and, therefore, lower measurement error. This makes it essential to normalize the source-level indicators around these structural breaks.

The following section explains the challenges of normalization in the presence of time-varying measurement errors through a simplified illustration.

3.D.1 Simplified illustration

Suppose that a stochastic process determines the true but unobservable sentiment. For simplicity, I consider this process to be independent and identically distributed (iid), although it could also follow a different pattern, such as an AR(1):

$$s_t \sim iid(\mu, \sigma_s^2). \tag{3.16}$$

The sentiment observed from source i at time $t = 1, \ldots, T$, denoted $\hat{s}_{t,i}$, comprises the true sentiment plus some measurement error, which is also assumed to be iid:

$$\hat{s}_{t,i} = s_t + v_{t,i} \quad v_{t,i} \sim iid(0, \sigma_t^2). \tag{3.17}$$

I assume a constant magnitude for measurement error across all sources for simplicity. However, let us consider that from a certain time point, T_b, onwards, the magnitude of measurement error diminishes due to improvements in OCR quality or changes in publication frequency, for instance:

$$\sigma_t^2 = \begin{cases} \sigma_h^2 & \text{if } t \leq T_b \\ \sigma_l^2 & \text{if } t > T_b. \end{cases} \tag{3.18}$$

Here, σ_h^2 represents the higher measurement error variance before time T_b, and σ_l^2 is the lower variance after that. Overlooking this breakpoint and simply averaging out the indicators from all sources would lead to a higher variance in the high measurement error regime:

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} \hat{s}_{t,i|t\leq T_b}\right) = \sigma_s^2 + \frac{1}{n}\sigma_h^2 > Var\left(\frac{1}{m}\sum_{i=1}^{m} \hat{s}_{t,i|t>T_b}\right) = \sigma_s^2 + \frac{1}{m}\sigma_l^2 \tag{3.19}$$

In this expression, n signifies the number of sources before T_b, and m is the number post T_b, typically with $m > n$. If both m and n are large enough, the difference in variance would be negligible. However, in the context of this project, the number of sources is limited, especially for the early sample, and the difference in variance is substantial. It follows that the level of the aggregated indicator is not comparable pre and post-break.

If this is the case, a nice indicator property would be to have the same variance across time for an application of business cycle dating utilizing a Markov-switching model. One potential strategy is to normalize the indicators from the pre-break and post-break periods separately before combining them.

$$\begin{gathered}
Var\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{s}_{t,i|t\leq T_b}-\mu}{\sqrt{\sigma_s^2+\sigma_h^2}}\right) < Var\left(\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{s}_{t,i|t>T_b}-\mu}{\sqrt{\sigma_s^2+\sigma_l^2}}\right) \\
\frac{n^2\sigma_s^2+n\sigma_h^2}{n^2(\sigma_s^2+\sigma_h^2)} < \frac{m^2\sigma_s^2+m\sigma_l^2}{m^2(\sigma_s^2+\sigma_l^2)} \\
\frac{\sigma_s^2}{\sigma_s^2+\sigma_h^2}+\frac{\sigma_h^2}{n(\sigma_s^2+\sigma_h^2)} < \frac{\sigma_s^2}{\sigma_s^2+\sigma_l^2}+\frac{\sigma_l^2}{m(\sigma_s^2+\sigma_l^2)}.
\end{gathered} \tag{3.20}$$

However, depending on the specific parameters, the combined indicator's resulting pre- and post-break variance could differ substantially. A significant disparity between m and n, coupled with a large $\frac{\sigma_h^2}{\sigma_l^2}$ ratio, results in a higher post-breakpoint variance. The intuition behind this result is that a large pre-break noise-to-signal ratio causes a more substantial downscaling of the signal during normalization, leading to a muted signal in the averaged indicator where some noise is canceled out.

3.D.2 Correlation with true sentiment

The ultimate goal is to end up with an indicator correlated as much as possible with the true sentiment over the entire sample. In the setup of the simple illustration above, the true sentiment is known. Therefore, I can calculate the correlation between the true sentiment and the aggregated indicator for different normalization strategies. If no normalization is applied before aggregation, the correlation is given by

$$\begin{aligned}
\rho_{\text{NN}} &= Corr\left(s_t, \frac{T_b}{T}\frac{1}{n}\sum_{i=1}^{n}\hat{s}_{t,i|t\leq T_b}+\frac{T-T_b}{T}\frac{1}{m}\sum_{i=1}^{m}\hat{s}_{t,i|t>T_b}\right) \\
&= \frac{\sigma_s^2}{\sigma_s\sqrt{\sigma_s^2+\frac{T_b}{T}\frac{1}{n}\sigma_h^2+\frac{T-T_b}{T}\frac{1}{m}\sigma_l^2}}
\end{aligned} \tag{3.21}$$

where $\frac{T_b}{T}$ and $\frac{T-T_b}{T}$ are the weights of the pre- and post-break period, respectively.

If normalization is applied to both segments individually before aggregation, the correlation is given by

$$\begin{aligned}\rho_{\text{SN}} &= Corr\left(s_t, \frac{T_b}{T}\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{s}_{t,i|t\leq T_b}}{\sqrt{\sigma_s^2+\sigma_h^2}} + \frac{T-T_b}{T}\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{s}_{t,i|t>T_b}}{\sqrt{\sigma_s^2+\sigma_l^2}}\right) \\ &= \frac{\sigma_s^2/\sqrt{\sigma_s^2+\frac{T_b}{T}\frac{1}{n}\sigma_h^2+\frac{T-T_b}{T}\frac{1}{m}\sigma_l^2}}{\sigma_s\sqrt{\frac{T_b}{T}\left(\frac{\sigma_s^2}{\sigma_s^2+\sigma_h^2}+\frac{\sigma_h^2}{n(\sigma_s^2+\sigma_h^2)}\right)+\frac{T-T_b}{T}\left(\frac{\sigma_s^2}{\sigma_s^2+\sigma_l^2}+\frac{\sigma_l^2}{m(\sigma_s^2+\sigma_l^2)}\right)}}.\end{aligned} \tag{3.22}$$

The question is whether normalization pays off, that is, whether $\rho_{\text{SN}} > \rho_{\text{NN}}$. This condition can be simplified to

$$\begin{aligned}\rho_{\text{SN}} &> \rho_{\text{NN}} \\ 1 &> \frac{T_b}{T}\left(\frac{\sigma_s^2}{\sigma_s^2+\sigma_h^2}+\frac{\sigma_h^2}{n(\sigma_s^2+\sigma_h^2)}\right)+\frac{T-T_b}{T}\left(\frac{\sigma_s^2}{\sigma_s^2+\sigma_l^2}+\frac{\sigma_l^2}{m(\sigma_s^2+\sigma_l^2)}\right).\end{aligned} \tag{3.23}$$

The right-hand side of this condition is a weighted average. Therefore, the weights sum up to one. Normalizing the segments is beneficial if the terms in brackets are lower than one.

$$\begin{aligned}1 &> \frac{\sigma_s^2}{\sigma_s^2+\sigma_h^2}+\frac{\sigma_h^2}{n(\sigma_s^2+\sigma_h^2)} \\ n(\sigma_s^2+\sigma_h^2)^2 &> (n\sigma_s^2+\sigma_h^2)(\sigma_s^2+\sigma_h^2) \\ \sigma_s^2+\sigma_h^2 &> \sigma_s^2+\frac{1}{n}\sigma_h^2.\end{aligned} \tag{3.24}$$

Hence, if $n > 1$, the left bracket is also lower than one. The same holds for the second bracket if $m > 1$. At least five sources are available in the early sample of the dataset used for this chapter. These findings suggest that normalizing the segments before aggregation is beneficial.

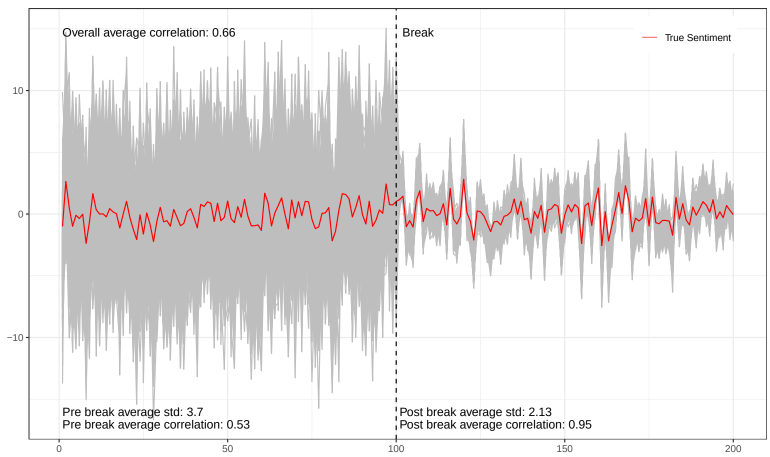
3.D.3 Simulation

To substantiate these findings, I carried out a simulation exercise. In this exercise, I compare the two aggregation methods discussed above. To mirror what is observed in the actual dataset, I employed the following parameter values:

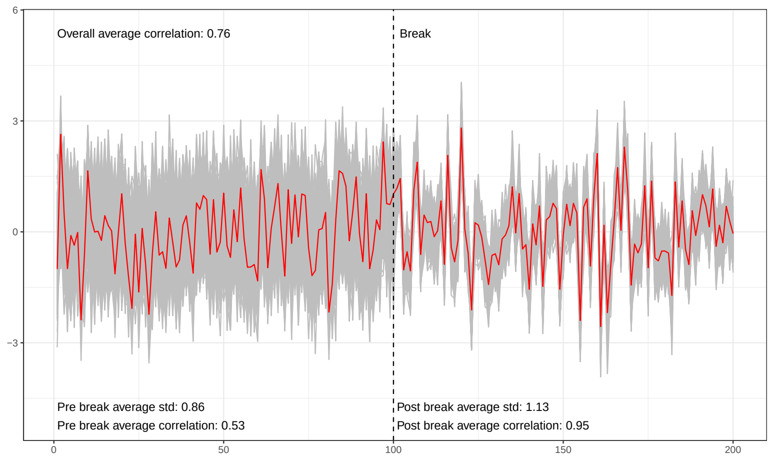
$$
\begin{aligned}
&\sigma_s = 1, \qquad && \sigma_h = 10, \qquad && \sigma_l = 2, \qquad && n = 15, \\
&m = 95, && t = 1, \ldots, 200, && T_b = 100.
\end{aligned}
$$

Figure 3.D.1 — Simulation exercise for different normalization strategies

(a) Without normalization



(b) With normalization



I ran 1000 iterations to compute the average indicator for a set of 110 time series, conducting the procedure once without any normalization and once with normalization.

The outcomes are graphically represented in Figure 3.D.1: Panel (a) illustrates the scenario without normalization, where a gray line represents each of the 1000 simulations, and the true underlying series is traced in red. As anticipated, on average, the variance of the composite indicator is notably higher in the pre-break phase compared to the post-break phase.

Panel (b) of Figure 3.D.1 displays the results when the normalization is employed. The simulations reveal a compelling aspect: while the average correlation with the true series is essentially the same for both pre-break and post-break periods regardless of normalization, it is considerably higher across the entire timeline when the data is normalized. This enhancement in correlation serves as an extra argument to normalize the data.

Appendix 3.E Markov-Switching Dynamic Factor Model

In this section, I describe the MS-DFM used for robustness.[34] MS-DFMs are pioneered by Kim (1994), Diebold and Rudebusch (1996), Kim and Yoo (1995) and Chauvet (1998). The model here closely follows Chauvet (1998). Estimating DFMs operates on the principle that an observed time series vector X_t can be split into two separate and orthogonal elements. First, the common components, often referred to as latent factors (f_t), encapsulate the joint movements among the observed variables in X_t. Second, the idiosyncratic component (e_t). These idiosyncratic components emerge from measurement errors and unique characteristics inherent in the data. The MS-DFM reads as follows:

$$X_t = \Lambda f_t + e_t \tag{3.25}$$

$$f_t = \mu_{s_t} + \sum_{p=1}^{P} A_p f_{t-p} + \eta_t \quad \eta_t \sim N(0, I) \tag{3.26}$$

$$e_t = \sum_{q=1}^{Q} C_q e_{t-q} + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma) \tag{3.27}$$

where P is the lag number of the factor and Q is the lag order of the idiosyncratic component. s_t is the realization of a two-state Markov chain with

[34] See e.g. Mariano and Murasawa (2010) and Stock and Watson (1989, 2016) for prominent examples and further information on DFMs.

$$\Pr(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots, y_{t-1}, y_{t-2}, \ldots) = \Pr(s_t = j | s_{t-1} = i) = p_{ij} \tag{3.28}$$

where $i, j = 0, 1$. Within this framework, one can label s_t = 0 and s_t = 1 as the expansion and recession states at time t.

The specific MS-DFM in this application has one unobserved factor (See also scree plot 3.A.10), which is assumed to follow an AR(5) process (i.e. $P = 5$, based on Bayesian Information Criterion). The innovations are assumed to be independent (so that Σ is a diagonal matrix), and the error term associated with each equation is assumed to follow an independent AR(2) process (i.e. $Q = 2$). Therefore, the specification considered here is:

$$x_{i,t} = \lambda_i f_t + e_{i,t} \tag{3.29}$$

$$f_t = \mu_{s_t} + a_1 f_{t-1} + a_2 f_{t-2} + a_3 f_{t-3} + a_4 f_{t-4} + a_5 f_{t-5} + \eta_t \quad \eta_t \sim N(0, \sigma_f^2) \tag{3.30}$$

$$e_{i,t} = c_{i,1} e_{i,t-1} + c_{i,2} e_{i,t-2} + \varepsilon_{i,t} \quad \varepsilon_{i,t} \sim N(0, \sigma_i^2) \tag{3.31}$$

where i denotes one of the $n = 11$ indicators ($i = 1, \ldots, n$). For identification reasons, the variance of η_t, σ_f^2, is given by unity. For estimation, the model is cast into state-space representation and estimated using the Kalman filter.

$$y_t = H\xi_t + w_t \qquad w_t \sim NID(0, R) \tag{3.32}$$

$$\xi_t = \mu(s_t) + F\xi_{t-1} + v_t \qquad v_t \sim NID(0, Q). \tag{3.33}$$

I use the following definitions:

$$y_t = [x_{1,t}, \dots, x_{n,t}]' \tag{3.34}$$

$$w_t = 0_{(n\times 1)} \tag{3.35}$$

$$R = 0_{(n\times n)} \tag{3.36}$$

$$\xi_t = [f_t, \dots, f_{t-P}, u_{1,t}, \dots, u_{n,t}, u_{1,t-1}, \dots, u_{n,t-1}]' \tag{3.37}$$

$$H = \begin{pmatrix} \lambda_1 & 0_{(1\times P-1)} & 1 & 0 & \dots & 0 & 0_{(1\times n)} \\ \lambda_2 & 0_{(1\times P-1)} & 0 & 1 & \dots & 0 & 0_{(1\times n)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_n & 0_{(1\times P-1)} & 0 & 0 & \dots & 1 & 0_{(1\times n)} \end{pmatrix} \tag{3.38}$$

$$v_t = [\eta_t, 0_{(1\times P-1)}, \varepsilon_{i,t}, \dots, \varepsilon_{n,t}, 0_{(1\times n)},]' \tag{3.39}$$

$$diag(Q) = [\sigma_f^2, 0_{(1\times P-1)}, \sigma_1^2, \dots, \sigma_n^2, 0_{(1\times n)}]' \tag{3.40}$$

$$F = \begin{pmatrix} a_1 & \dots & a_P & 0_{(1\times n)} & 0_{(1\times n)} \\ I_{P-1} & \dots & 0_{(P-1\times 1)} & 0_{(P-1\times n)} & 0_{(P-1\times n)} \\ 0_{(n\times P)} & \dots & 0_{(n\times P)} & diag(c_{i,1}) & diag(c_{i,2}) \\ 0_{(n\times P)} & \dots & 0_{(n\times P)} & I_n & 0_{(n\times n)} \end{pmatrix} \tag{3.41}$$

$$\mu(s_t) = [\mu_{s_t}, 0_{(1\times P-1)}, 0_{(1\times 2n)}]'. \tag{3.42}$$

The estimation of the dynamic factor model with regime switching is carried out by maximizing its likelihood function. For estimating this model, I utilize the techniques developed by Kim (1994) and Chauvet (1998).

Concluding remarks

This thesis explores the use of textual analysis tools in economics, focusing on developing innovative indicators and methodologies for measuring economic activity, assessing the impact of monetary policy, and understanding historical business cycles. The research presented in the three chapters provides insights for both academics and policymakers.

In the first chapter, a daily economic indicator for Switzerland was developed using financial market data and news sentiment analysis. This "fever curve" accurately tracks Swiss GDP growth and correlates with other business cycle indicators. The indicator's timely updates (with a one-day lag) and its strong performance during economic turning points demonstrate its effectiveness in nowcasting GDP growth. It outperforms traditional business cycle indicators once a month of data is available within a quarter.

Chapter 2 proposes a novel methodology to identify and measure the effects of different types of US monetary policy shocks (interest rate target, path, and term premium shocks) on the exchange rate. An advantage of this method is that does not rely on high-frequency data or precise knowledge of event timing. The findings indicate no evidence of delayed exchange rate overshooting, challenging some existing empirical results theories. The results also show that different monetary policy actions have varying degrees of impact on the exchange rate, primarily in terms of persistence rather than direction.

Using textual analysis on historical documents, a comprehensive business cycle indicator for Switzerland from 1821 to 2021 is developed in the third chapter. This indicator effectively captures major economic downturns and aligns with historical narratives of economic crises and wars. The chapter provides a first-of-its-kind business cycle chronology for Switzerland for the 19^{th} and early 20^{th} centuries, revealing that

Swiss recessions have become less frequent but not necessarily shorter over time.

There are several implications for policymakers. The daily economic indicator allows policymakers to monitor the economy's health with minimal lag, enabling more timely and effective responses to economic fluctuations. This is particularly valuable during economic crises when rapid intervention is critical. The methodology for identifying monetary policy shocks offers a more nuanced understanding of how different policy actions affect the exchange rate. This may be particularly important for policy makers in small open economies, where the impact of the exchange rate on the economy is often widely accepted, but pundits sometimes question the effectiveness of monetary policy to affect the exchange rate (see e.g. Yeşin, 2017). Additionally, the historical business cycle indicator provides a long-term perspective on economic fluctuations, helping policymakers understand the historical context of current economic trends. This can inform more robust policy decisions by learning from past economic events and their outcomes.

References

Yeşin, P. (2017). Capital flows and the swiss franc. *Swiss Journal of Economics and Statistics*, *153*(4), 403–436. https://doi.org/10.1007/BF03399513