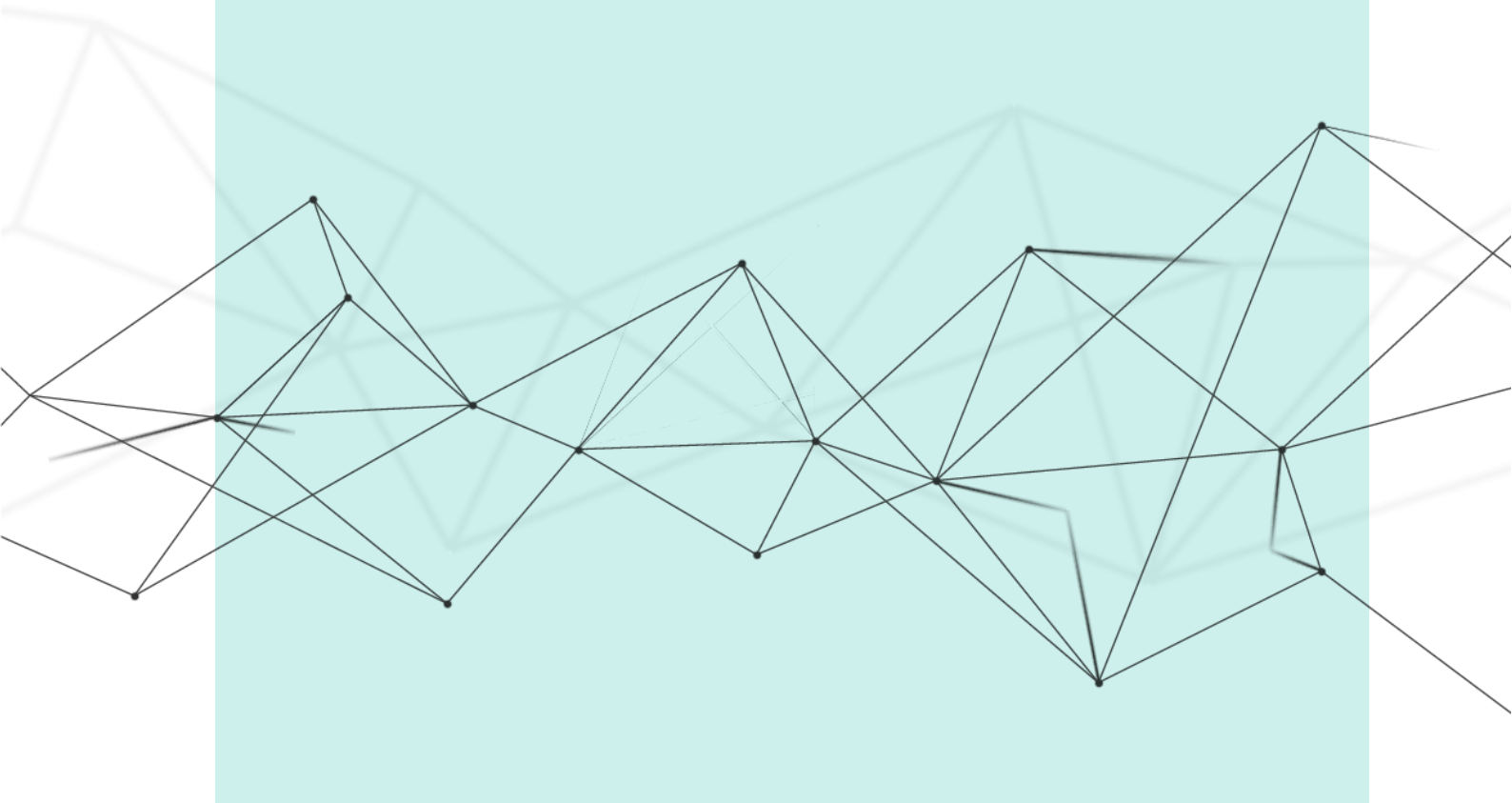Studie │ Februar 2024

# Study to evaluate the future data compilation for the Swiss Consumer Sentiment Index

# Grundlagen für die Wirtschaftspolitik

**Autoren**
Prof. Georg Lutz, Prof. Boris Wernli, Dr. Erika Antal, PD Dr. Oliver Lipps, Victor Legler
FORS, Universität Lausanne
Bâtiment Géopolis, 1015 Lausanne

Prof. Daniel Kaufmann, Marc Burri
Institut de recherches économiques (IRENE),
Rue A.-L. Breguet 2, 2000 Neuchâtel

# Table of Contents

# Executive Summary

*Summary English*

This evaluation on the future data compilation for the Swiss Consumer Sentiment Index, conducted by FORS (Swiss Centre of Expertise in the Social Sciences) in collaboration with IRENE (Institut de recherches économiques de l'université de Neuchâtel) looked at a number of different possible design features.

Overall, the way how consumer sentiment is calculated in Switzerland is well in line with similar measures in other parts of the world, with a methodology in data collection showing a high standard of quality compared to other countries' practices. But **Switzerland is the only country under review with only quarterly data collection and index calculation**. This severely limits the information content available for business cycle monitoring, a monthly index, as is the case in the other countries reviewed, allowing a much closer monitoring of the evaluation of consumer confidence. Under the assumption that *at least* a monthly index shall be calculated, we recommend the following general design for the future:

- The consumer index shall be done using a **weekly rolling cross section design (RCS)**. In this design new addresses are released every week of the year.
- We recommend having **500 weekly target interviews**. Our analysis of the current Consumer sentiment data as well as data from other studies shows, that 400 interviews are sufficient to calculate a stable mean index, with a RCS design there will be fluctuation in the number of interviews between the weeks, so there will always be sufficient interviews per week with a slightly higher average number of interviews per week. 500 interviews per week means that there will be about 2167 interviews per month.
- We recommend moving to **a single mode, web-only survey**, with offline recruiting through an invitation letter with at least one reminder letter. We are convinced that a well-executed web-only survey allows to get similar if not better data quality compared to the current design. A single mode is easier to administer – and therefore more cost efficient – and data is available much quicker.
- Shall SECO want to calculate a monthly **regional index** for the seven NUTS II regions (Grossregionen), the sample needs to be stratified. If SECO wants to have at least 500 interviews per region/month, this means that 3500 interviews (7*500) have to be conducted per month instead of 2167 interviews per month that you would get with only a national sample. For the weekly and monthly national indicator, design weights would then have to be applied.
- Overall, we **don't see a strong need to change the question wording nor applying a weighting scheme for the aggregate consumer sentiment index**.
- We also see **no strong reason to apply non-response weights**, given that the representativeness is not a key objective, and we can assume that the non-response bias is relatively stable between the different waves.
- **We see, however potential to add additional questions.** The questionnaire is very short compared to the recruitment costs, some additional questionnaire time should not have any negative effect on response rates and response quality and should not create substantial costs if the survey is done in web only mode.
- If SECO **revises the methodology**, these changes should be systematically documented in an easily accessible way. To link indexes constructed with the old and new design one should adjust the mean of the old series to match the one of the new. This requires running the old and new design of data collection for 6-12 month in parallel.
- In normal times, the release of **Flash estimates** is of limited use for a monthly consumer confidence index. They may be more useful during crises with rapid and large changes in economic activity. If Flash estimates are released, they should focus on the aggregate index only.
- **To retropolate and create a monthly indicator back in time,** one may consider multivariate model-based retropolation methods. Unemployment data, KOF surveys in manufacturing, newspaper-based text sentiment, and export data are the variables with the largest potential to construct such an indicator. Simple retropolation schemes, such as a linear retropolation, should be considered too, as they are easy to apply and do not suffer from estimation uncertainty.

## Zusammenfassung Deutsch

Diese von FORS (Schweizer Kompetenzzentrum Sozialwissenschaften) in Zusammenarbeit mit IRENE (Institut de recherches économiques de l'université de Neuchâtel) durchgeführte Evaluation der zukünftigen Datenerhebung für den Schweizer Konsumentenstimmungsindex untersuchte eine Reihe von möglichen Designelementen.

Insgesamt entspricht die Art und Weise, wie die Konsumentenstimmung in der Schweiz berechnet wird, durchaus ähnlichen Messungen in anderen Teilen der Welt, wobei die Methodik der Datenerhebung im Vergleich zu den Praktiken anderer Länder einen hohen Qualitätsstandard aufweist. Die **Schweiz ist jedoch das einzige untersuchte Land, in dem die Datenerhebung und Indexberechnung nur vierteljährlich erfolgt**. Dies schränkt den Informationsgehalt, der für die Konjunkturbeobachtung zur Verfügung steht, stark ein. Ein monatlicher Index, wie er in den anderen untersuchten Ländern verwendet wird, ermöglicht eine viel genauere Überwachung der Bewertung des Verbrauchervertrauens. Unter der Annahme, dass *zumindest* ein monatlicher Index berechnet werden soll, empfehlen wir das folgende allgemeine Design für die Erhebung:

- Der Verbraucherindex soll mit einem wöchentlich rotierendem Querschnittsdesign (**rolling cross section design, RCS**) erhoben werden. Bei diesem Design werden jede Woche des Jahres neue Adressen veröffentlicht.
- Wir empfehlen, ein **Soll von 500 Antworten pro Woche** anzupeilen. Unsere Analyse der aktuellen Konsumentenstimmungsdaten sowie Daten aus anderen Studien zeigt, dass 400 Interviews ausreichen, um einen stabilen Durchschnittsindex zu berechnen. Bei einem RCS-Design wird die Anzahl der Interviews zwischen den Wochen schwanken, so dass es stets genügend Interviews pro Woche geben wird, wobei die durchschnittliche Anzahl der Interviews pro Woche etwas höher ist. 500 Interviews pro Woche bedeuten, dass es etwa 2167 Interviews pro Monat geben wird.
- Wir empfehlen die Umstellung auf einen **einzigen Erhebungsmodus, d. h. eine ausschliesslich internetbasierte Befragung**, mit Offline-Rekrutierung durch ein Einladungsschreiben mit mindestens einem Erinnerungsschreiben. Wir sind davon überzeugt, dass eine gut durchgeführte reine Internetbefragung eine ähnliche, wenn nicht sogar bessere Datenqualität als das derzeitige Design ermöglicht. Ein einziger Modus ist einfacher zu verwalten und somit kostengünstiger und die Daten sind viel schneller verfügbar.
- Möchte das SECO einen monatlichen **Regionalindex** für die sieben NUTS-II-Regionen (Grossregionen) berechnen, muss die Stichprobe stratifiziert werden. Wenn das SECO mindestens 500 Interviews pro Region/Monat durchführen möchte, bedeutet dies, dass 3500 Interviews (7*500) pro Monat durchgeführt werden müssen, anstatt 2167 Interviews pro Monat, die man mit einer nationalen Stichprobe erhalten würde. Für den wöchentlichen und monatlichen nationalen Indikator müssten dann Gewichtungen vorgenommen werden.
- Insgesamt sehen wir **keine grosse Notwendigkeit, die Frageformulierung anzupassen oder ein Gewichtungsschema für den aggregierten Index der Verbraucherstimmung anzuwenden**.
- Wir sehen auch **keinen triftigen Grund, Gewichtungen für Nichtantwortende anzuwenden**, da die Repräsentativität kein Kernziel darstellt und wir davon ausgehen können, dass die Verzerrung aufgrund von Nichtantworten zwischen den verschiedenen Wellen relativ stabil ist.
- Wir sehen jedoch **Potenzial für die Aufnahme zusätzlicher Fragen**. Der Fragebogen ist im Vergleich zu den Rekrutierungskosten sehr kurz, etwas zusätzliche Zeit für die Beantwortung des Fragebogens sollte keine negativen Auswirkungen auf die Antwortquoten und die Antwortqualität haben und sollte keine wesentlichen Kosten verursachen, wenn die Umfrage nur im Internet-Modus durchgeführt wird.
- Falls das SECO **die Methodik überarbeitet**, sollten diese Änderungen systematisch und leicht zugänglich dokumentiert werden. Um die mit dem alten und dem neuen Design erstellten Indizes zu verknüpfen, sollte der Mittelwert der alten Erhebungsreihe an denjenigen der neuen angepasst werden. Dazu müssen das alte und das neue Datenerhebungsdesign 6-12 Monate lang parallellaufend durchgeführt werden.
- Unter üblichen Gegebenheiten ist die Veröffentlichung von **Flash-Schätzungen** für einen monatlichen Konsumentenstimmungsindex von begrenztem Nutzen. Sie können in Krisenzeiten mit schnellen und grossen Veränderungen in der Wirtschaftstätigkeit nützlicher sein. Wenn Flash-Schätzungen veröffentlicht werden, sollten sie sich nur auf den Gesamtindex konzentrieren.
- Für die **Retropolation und die Erstellung eines monatlichen Indikators in der Vergangenheit** können multivariate modellbasierte Retropolationsmethoden in Betracht gezogen werden. Arbeitslosenzahlen, KOF-Umfragen im verarbeitenden Gewerbe, zeitungsbasierte Stimmungsdaten und Exportdaten sind die Variablen mit dem grössten Potenzial für die Erstellung eines solchen Indikators. Einfache Retropolationsverfahren, wie z.B. eine lineare Retropolation, sollten ebenfalls in Betracht gezogen werden, da sie einfach anzuwenden sind und nicht von Schätzungsunsicherheiten betroffen sind.

## Résumé français

Cette évaluation de la future collecte de données pour l'indice suisse du climat de consommation, menée par FORS (Centre de compétences suisse en sciences sociales) en collaboration avec IRENE (Institut de recherches économiques de l'université de Neuchâtel), a examiné un certain nombre de caractéristiques de conception possibles.

Dans l'ensemble, le mode de calcul du climat de consommation en Suisse s'inscrit dans la lignée de mesures similaires dans d'autres parties du monde, la méthodologie de collecte des données montrant un niveau de qualité élevé par rapport aux pratiques des autres pays. **Mais la Suisse est le seul pays examiné dont la collecte de données et le calcul de l'indice ne sont que trimestriels.** Cela limite fortement le contenu informatif disponible pour le suivi du cycle économique, un indice mensuel, comme c'est le cas dans les autres pays réexaminés, permettant un suivi beaucoup plus étroit de l'évaluation du climat de consommation. En partant du principe qu'un indice *au moins* mensuel sera calculé, nous recommandons le design général suivant pour l'enquête :

- L'indice du climat de consommation doit être réalisé en utilisant un plan hebdomadaire rotatif de sondage en coupe transversale (**rolling cross section design, RCS**). Dans ce plan, de nouvelles adresses sont publiées chaque semaine de l'année.
- Nous recommandons de **cibler 500 réponses par semaine**. Notre analyse des données actuelles de l'indice du climat de consommation ainsi que des données d'autres études montrent que 400 réponses sont suffisantes pour calculer un indice moyen stable. Avec un design RCS il y aura une fluctuation dans le nombre de réponses d'une semaine à l'autre, c'est-à-dire qu'il y aura toujours suffisamment de réponses par semaine avec un nombre moyen de réponses par semaine légèrement plus élevé. 500 réponses par semaine signifie qu'il y aura environ 2167 réponses par mois.
- Nous recommandons de passer à un **mode d'enquête uniquement en ligne**, avec un recrutement hors ligne par le biais d'une lettre d'invitation avec au moins une lettre de rappel. Nous sommes convaincus qu'une enquête en ligne bien menée permet d'obtenir une qualité de données similaire, voire supérieure, à celle de la méthode actuelle. Un mode unique est plus facile à administrer - et donc moins coûteux - et les données sont disponibles beaucoup plus rapidement.
- Si le SECO veut calculer un **indice régional** mensuel pour les sept régions NUTS II (Grossregionen), l'échantillon doit être stratifié. Si le SECO veut avoir au moins 500 interviews par région/mois, cela signifie que 3500 interviews (7*500) doivent être menées par mois au lieu des 2167 interviews par mois que l'on obtiendrait avec un seul échantillon national. Pour l'indicateur national hebdomadaire et mensuel, il faudrait alors appliquer des pondérations.
- Dans l'ensemble, nous ne **voyons pas de forte nécessité de modifier la formulation des questions ni d'appliquer un schéma de pondération pour l'indice agrégé du climat de consommation**.
- Nous ne voyons **pas non plus de raison importante d'appliquer des poids de non-réponse**, étant donné que la représentativité n'est pas un objectif clé et que nous pouvons supposer que le biais de non-réponse est relativement stable entre les différentes vagues.
- **Nous voyons cependant un potentiel pour ajouter des questions supplémentaires**. Le questionnaire est très court par rapport aux coûts de recrutement, un peu de temps supplémentaire ne devrait pas avoir d'effet négatif sur les taux de réponse et la qualité des réponses et ne devrait pas créer de coûts substantiels si l'enquête est réalisée en mode web uniquement.
- Si le SECO **révise la méthodologie**, ces changements devraient être systématiquement documentés de manière à être facilement accessibles. Pour relier les indices construits avec l'ancien et le nouveau design, il faut ajuster la moyenne de l'ancienne série à celle de la nouvelle. Pour ce faire, il faut appliquer en parallèle l'ancienne et la nouvelle conception de la collecte de données pendant 6 à 12 mois.
- Dans des circonstances normales, la publication d'**estimations rapides** est d'une utilité limitée pour un indice mensuel du climat de consommation. Elles peuvent être plus utiles lors de crises avec des changements rapides et importants dans l'activité économique. Si des estimations rapides sont publiées, elles doivent porter uniquement sur l'indice agrégé.
- **Pour rétropoler et créer un indicateur mensuel en remontant dans le temps**, on peut envisager des méthodes de rétropolation basées sur des modèles multivariés. Les données sur le chômage, les enquêtes du KOF dans l'industrie manufacturière, des extractions de textes de confiance dans des journaux et les données sur les exportations sont les variables qui présentent le plus grand potentiel pour construire un tel indicateur. Des schémas de rétropolation simples, tels qu'une rétropolation linéaire, devraient également être envisagés, car ils sont faciles à appliquer et ne souffrent pas d'incertitude d'estimation.

Questa valutazione sulla futura compilazione dei dati per l'indice del clima di fiducia dei consumatori, condotta da FORS (Centro svizzero di competenze nelle scienze sociali) in collaborazione con IRENE (Institut de recherches économiques de l'université de Neuchâtel), ha esaminato una serie di possibili caratteristiche di disegno.

Nel complesso, il modo in cui viene calcolato clima di fiducia dei consumatori in Svizzera è ben in linea con misure simili in altre parti del mondo, con una metodologia nella raccolta dei dati che mostra un alto standard di qualità rispetto alle pratiche di altri paesi. Ma la Svizzera è l'unico paese in esame con una raccolta di dati e un calcolo dell'indice solo trimestrale. Questo limita fortemente il contenuto informativo disponibile per il monitoraggio nel ritmo del ciclo economico, un indice mensile, come avviene negli altri paesi riesaminati, permette un monitoraggio molto più stretto della valutazione della fiducia dei consumatori. Partendo dal presupposto che *almeno* un indice mensile sarà calcolato, raccomandiamo il seguente disegno generale per il futuro:

- L'indice del clima di fiducia dei consumatori deve essere fatto usando un disegno di sezione trasversale a rotazione settimanale (**rolling cross section design, RCS**). In questo disegno vengono rilasciati nuovi indirizzi ogni settimana dell'anno.
- Raccomandiamo di puntare a **500 risposte al sondaggio per settimana**. La nostra analisi degli attuali dati del clima di fiducia dei consumatori così come i dati di altri studi mostrano che 400 risposte sono sufficienti per calcolare un indice medio stabile. Con un disegno RCS ci sarà una fluttuazione nel numero di risposte tra le settimane; quindi, ci saranno sempre risposte sufficienti per settimana con un numero medio di risposte per settimana leggermente superiore. 500 risposte a settimana significa che ci saranno circa 2167 risposte al mese.
- Raccomandiamo di passare ad **una modalità unica con un sondaggio solo su Internet**, con reclutamento offline attraverso una lettera d'invito con almeno una lettera di ricordo. Siamo convinti che un sondaggio solo su Internet ben eseguito permetta di ottenere una qualità dei dati simile se non migliore rispetto al design attuale. Una modalità singola è più facile da amministrare - e di conseguenza più efficiente in termini di costi - e i dati sono disponibili molto più rapidamente.
- Se la SECO vuole calcolare un **indice regionale** mensile per le sette regioni NUTS II (grandi regioni della Svizzera), il campione deve essere stratificato. Supponendo che la SECO voglia avere almeno 500 risposte per regione/mese, questo significa che devono essere condotte 3500 interviste (7*500) al mese invece delle 2167 interviste al mese che si otterrebbero con un campione solo nazionale. Per l'indicatore nazionale settimanale e mensile, dovrebbero essere applicati dei pesi del disegno di campionamento.
- Nel complesso, non vediamo una forte necessità di cambiare la formulazione delle domande né di applicare uno schema di ponderazione per l'indice aggregato del clima di fiducia dei consumatori.
- **Non vediamo nemmeno una forte ragione per applicare pesi di non risposta**, dato che la rappresentatività non è un obiettivo chiave e possiamo assumere che il bias di non risposta sia relativamente stabile tra le diverse ondate.
- **Vediamo, tuttavia, il potenziale per aggiungere ulteriori domande.** Il questionario è molto breve rispetto ai costi di reclutamento, un po' di tempo in più per il questionario non dovrebbe avere alcun effetto negativo sui tassi di risposta e sulla qualità delle risposte e non dovrebbe creare costi sostanziali se l'indagine è fatta solo in modalità Internet.
- Se la SECO **rivede la metodologia**, questi cambiamenti dovrebbero essere sistematicamente documentati in modo facilmente accessibile. Per collegare gli indici costruiti con il vecchio e il nuovo disegno si dovrebbe aggiustare la media della vecchia serie per farla coincidere con quella del nuovo. Questo richiede l'esecuzione del vecchio e del nuovo disegno di raccolta dati per 6-12 mesi in parallelo.
- In tempi normali, il rilascio di **stime rapide** è di utilità limitata per un indice mensile del clima di fiducia dei consumatori. Possono essere più utili durante le crisi con rapidi e grandi cambiamenti nell'attività economica. Se le stime rapide venissero rilasciate, dovrebbero concentrarsi solo sull'indice aggregato.
- **Per retropolare e creare un indicatore mensile a ritroso nel tempo**, si possono considerare metodi di retropolazione basati su modelli multivariati. I dati sulla disoccupazione, le indagini KOF nel settore manifatturiero, il 'sentiment' testuale basato sui giornali e i dati sulle esportazioni sono le variabili con il maggior potenziale per costruire un tale indicatore. Anche schemi di retropolazione semplici, come una retropolazione lineare, dovrebbero essere considerati, poiché sono facili da applicare e non soffrono di incertezza di stima.

# 1   Introduction

The state secretariat of economic affairs (SECO) mandated FORS in collaboration with IRENE to evaluate various aspects of the future design for the data compilation for the Swiss Consumer Sentiment Index in 2021. More specifically this includes the following aspects with respect to sampling, frequency and regionalization of the data collection:

- How many people have to be interviewed in which intervals and in different regions if a regional indicator were to be calculated?
- How can the time series be continued, especially with a move to monthly surveys, are there to be the same number of interviews to be conducted over the month or should there be a weighting of the index and are there examples for such a transition in Switzerland or abroad?
- How could the quarterly index be retropolated to monthly frequency?
- How would a sample size look like to allow to have regional results?
- What would the sample size be in order to have a weekly index?

Further methodological questions include whether the index should be weighted and if the mode or other control variable should be included in the weighting. The study should also include an international comparison on how the index is calculated in Switzerland compared to selected other countries.

This report addressed the questions in a slightly different order. Chapter 2 outlines how consumer sentiment is measured in selected countries both in terms of questionnaire and various aspects of the field work. In chapter 3 various study design issues for the future data collection are discussed as well as how the past quarterly data could be retropolated to monthly frequency. In this chapter we also discuss the possible weighting procedures on nonresponse and design weights. The relevant code is then provided in an Annex.

# 2   International comparison

The following countries are included in the international comparison: Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands and the USA. They allow an overview of the situation that is not exhaustive but sufficiently representative of the diversity of international practices, including all of Switzerland's neighbouring countries, as well as the largest and longest-standing international players in this type of data collection.

To draw up this comparative overview, we have based ourselves on the documentation available on the Internet. In the text, we provide the link to access them in each case.[1] It is very structured and quite complete as regards the European countries, which have a partially harmonised survey. We have supplemented this information with evaluation documents, including the evaluation conducted by Bierbaumer-Polly et al. (2019) and the document mandated by the European Commission (2018).

For the US surveys, the situation is different. We have retained the two main indicators, always basing ourselves on the documentation available on the Internet, but also on scientific articles (Merkle, Langer and Sussman 2004; Curtin undated) and/or information available on economic news sites, to complete the fragmentary official information.

As a preliminary remark, it should be noted that all these international surveys considered in the international comparison are conducted on a monthly basis, with monthly calculation of the indices, which is quite different from the Swiss consumer sentiment index, which is only available every quarter.

---

[1] In a few cases, we contacted the institutes by e-mail to obtain further information.

## The Consumer Confidence Indicator for the member states of the European Union (EU)

We mainly used the methodological documents available on the European Commission's website, relating to the harmonised surveys[2]. In addition to the methodological documents relating to the guidelines for conducting the European surveys (Surveys User Guide 2021 and List of best practice for the conduct of business and consumer surveys 2014), we used the methodological sheets of the countries and their questionnaires for this review.

The Joint Harmonised EU Programme of Business and Consumer Surveys (BCS) was set up in 1961. The first survey was the harmonised business survey in the manufacturing industry launched in 1962. Since then, the sector coverage of the programme has broadened considerably. The BCS programme was extended to the construction sector and to investment plans in the manufacturing sector in 1966, to consumers in 1972, to the retail trade in 1984, and to the services sector in 1996.

In the meantime, the geographical coverage of the programme has been extended to include new Member States as well as new candidate countries. The programme currently covers all 27 EU Member States and all five EU candidate countries (i.e. Montenegro, North Macedonia, Albania, Serbia and Turkey). The early integration of the candidate countries into the programme was considered as necessary to provide reliable and comparable data to follow their economic situation, and to guarantee the production of accurate EU aggregates once these countries become members of the EU.

The European Commission has published Consumer Confidence Indicators (CCI) since the 1970s. The current CCI, has been adapted several times and was last amended in 2019.

The harmonised surveys are carried out at national level by partner institutes such as ministries, statistical offices, central banks, research institutes, business associations or private companies. The surveys are conducted monthly according to a common methodology, which consists essentially of harmonized questionnaires and a common timetable. The core instrument combines survey questions about consumers' personal finances with their expectations in respect of macro-economic developments. Harmonisation does not mean uniformity. The national questionnaires are likely to include additional questions. As for the other methodological aspects, they differ widely from one country to another, with each national partner having a large margin of freedom to carry out the survey in the conditions that best suit the local context.

## The Michigan Consumer Sentiment Index (MCSI)

The Michigan Consumer Sentiment Index (MCSI)[3] is a monthly national survey conducted on US territory by the Institute for Social Research of the University of Michigan since 1946. The survey aims at collecting information about consumer expectations regarding the overall economy, as it has long stressed the important influence of consumer spending and saving decisions in determining the course of the national economy. The survey asks consumers about their views on their own personal finances, as well as the short-term and long-term state of the U.S. economy. The Index of Consumer Expectations, produced by the Surveys of Consumers, is included in the Leading Indicator Composite Index published by the U.S. Department of Commerce, Bureau of Economic Analysis.

## The Consumer Confidence Index (CCI), published by the Conference Board

The Consumer Confidence Index (CCI)[4] is published monthly by the Conference Board, a not-for-profit research organization for businesses. Even if led by partners from the business world rather than academics, the Consumer Confidence Index and its related series are among the earliest sets of economic indicators available each month and are closely watched as leading indicators for the U.S. economy.

---

[2] https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/business-and-consumer-surveys/methodology-business-and-consumer-surveys_en
[3] https://data.sca.isr.umich.edu/survey-info.php

[4] https://conference-board.org/data/consumerconfidence.cfm

The index is based on consumers' perceptions of current business and employment conditions, as well as their expectations for six months hence regarding business conditions, employment, and income.

## 2.1   Questionnaire/question wording

This chapter analyse the different questions used in the different surveys. A similar structure is found in the Swiss[5] and European[6] questionnaire's core. We first focus on the four questions that compose the Swiss and European indexes, in order to analyse their wording. The question number in the first column of table 1 corresponds to that of the Swiss questionnaire in the methodological annex available online, the last column referring to the European numbering.

Small differences can already be seen *between the language versions of the Swiss questionnaire*, with the Italian version of Q1.2, changes in the general economic situation, specifying 'of Switzerland', as in the European survey, whereas this is not specified in French and German. In Italian, the question is therefore posed more specifically regarding the national context.

Another point is that in Q4.1 and 4.2 the German and Italian versions refer to the financial situation of the household, but in the French version to the household budget. In the latter case, the difference in wording is subtle, but it refers to a particular dimension of the financial situation, not to its totality. It is an open question whether adaptation is recommended after so many years of investigation, but there are slight semantic differences in these questions that can be raised.

In the comparison between the *Swiss and the European index questions*, in addition to the discrepancy already noted regarding question Q1.2 for French and German, the main difference lies in the question Q5.2, the favourable time for large acquisitions.

- The EU index retains a question (Q9) asking about intentions to make major purchases over the next 12 months compared to the last 12 months, whereas the question in the Swiss survey refers to the present.
- In Switzerland, the question is asked in a general way "do you think this is the right time", in the European countries, the question is asked about the real household's intentions "do you expect to spend".
- The response categories are different, with 3 categories for CH and 5 for the EU.

Note that the European questionnaire also includes a question (Q8) identical to the Swiss Q5.2, but that is not part of their aggregate index.

Concerning the different European versions of the questions retained in the aggregated index, at least in the most common languages, it can be said that they are very strongly harmonised, no semantic difference emerges from the comparison.

For the rest of the questionnaires, i.e. the questions that are not taken into account in the aggregated indices, several points can be noted.

---

[5] https://www.seco.admin.ch/dam/seco/fr/dokumente/Wirtschaft/Wirtschaftslage/Konsumentenstim-mung/Methodik.pdf.download.pdf/Methodik.pdf
[6] https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/business-and-consumer-surveys/methodology-business-and-consumer-surveys/national-questionnaires_en

*Table 1 – Wording of the Swiss and European index questions*

| n. CH | Switzerland, english version | CH, german version | CH, italian version | CH, french version | UE, english version, from the bcs user guide | n.UE |
|---|---|---|---|---|---|---|
| Q1.2 | How do you think the general economic situation will develop over the next 12 months? Will it… | Wie wird sich Ihrer Ansicht nach die «allgemeine Wirtschaftslage» in den kommenden zwölf Monaten entwickeln. Wird sie sich… | A Suo giudizio, nei prossimi 12 mesi, la situazione economica generale della Svizzera dovrebbe… | À votre avis, comment se développera la situation économique générale au cours des 12 prochains mois ? | How do you expect the general economic situation in this country to develop over the next 12 months? It will… | Q4 |
| | Significantly improve ++ | Wesentlich verbessern ++ | Migliorare nettamente ++ | S'améliorera considérablement ++ | Get a lot better ++ | |
| | Marginally improve + | Etwas verbessern + | Migliorare leggermente + | S'améliorera légèrement + | Get a little better + | |
| | Remain virtually unchanged = | Wird sie in etwa gleich bleiben = | Rimanere quasi invariata = | Restera à peu près inchangé = | Stay the same = | |
| | Marginally deteriorate - | Etwas verschlechtern - | Peggiorare leggermente - | Se détériorera légèrement - | Get a little worse - | |
| | Significantly deteriorate -- | Wesentlich verschlechtern -- | Peggiorare nettamente -- | Se détériorera considérablement -- | Get a lot worse -- | |
| | Don't know 0 | Weiss nicht 0 | Non so 0 | Ne sait pas 0 | Don't know 0 | |
| Q4.1 | How, in your view, has your household's financial situation changed over the last 12 months? Has it… | Wie hat sich Ihrer Ansicht nach die finanzielle Lage von Ihrem Haushalt in den letzten zwölf Monaten entwickelt? Hat sie sich… | A Suo giudizio, nel corso degli ultimi 12 mesi, come è evoluta la situazione finanziaria del Suo nucleo familiare? È… | Selon vous, comment s'est développé le budget de votre ménage au cours des 12 derniers mois ? | How has the financial situation of your household changed over the last 12 months? It has… | Q1 |
| | Significantly improved ++ | Wesentlich verbessert ++ | Migliorata nettamente ++ | S'est considérablement amélioré ++ | Got a lot better ++ | |
| | Marginally improved + | Etwas verbessert + | Migliorata leggermente + | S'est légèrement amélioré + | Got a little better + | |
| | Remained virtually unchanged = | Ist in etwa gleich geblieben = | Rimasta quasi invariata = | Est resté à peu près inchangé = | Stayed the same = | |
| | Marginally deteriorated - | etwas verschlechtert - | Peggiorata leggermente - | S'est un peu détérioré - | Got a little worse - | |
| | significantly deteriorated -- | wesentlich verschlechtert -- | Peggiorata nettamente -- | S'est considérablement détérioré -- | Got a lot worse -- | |
| | Don't know 0 | weiss nicht 0 | Non so 0 | Ne sait pas 0 | Don't know 0 | |
| Q4.2 | How do you expect your household's financial situation to change over the next 12 months? Will it… | Wie wird sich Ihrer Ansicht nach die finanzielle Lage von Ihrem Haushalt in den kommenden zwölf Monaten entwickeln? Wird sie sich… | A Suo giudizio, nei prossimi 12 mesi, la situazione finanziaria del Suo nucleo familiare dovrebbe… | À votre avis, comment se développera le budget de votre ménage au cours des 12 prochains mois ? | How do you expect the financial position of your household to change over the next 12 months? It will… | Q2 |
| | Significantly improve ++ | Wesentlich verbessern ++ | Migliorare nettamente ++ | S'améliorera considérablement ++ | Get a lot better ++ | |
| | Marginally improve + | Etwas verbessern + | Migliorare leggermente + | S'améliorera légèrement + | Get a little better + | |
| | Remain virtually unchanged = | Wird sie in etwa gleich bleiben = | Rimanere quasi invariata = | Restera à peu près inchangé = | Stay the same = | |
| | Marginally deteriorate - | Etwas verschlechtern - | Peggiorare leggermente - | Se détériorera légèrement - | Get a little worse - | |
| | Significantly deteriorate -- | Wesentlich verschlechtern -- | Peggiorare nettamente -- | Se détériorera considérablement -- | Get a lot worse -- | |
| | Don't know 0 | Weiss nicht 0 | Non so 0 | Ne sait pas 0 | Don't know 0 | |

| Q5.2 | Do you think that now is a good time to make major purchases (major household appliances, furniture, car, consumer electronics, etc.)? | Glauben Sie, dass es jetzt eine gute Zeit ist, grössere Anschaffungen (z.B. grössere Haushaltgeräte, Möbel, Unterhaltungselektronik, usw.) zu tätigen? | Ritiene che attualmente sia un momento opportuno per effettuare acquisti importanti (elettrodomestici costosi, mobili, automobile, ecc.)? | Croyez-vous le moment favorable pour faire de grandes acquisitions (par exemple : des appareils élec- troménagers d'importance, meubles, autos, etc.) ? | Compared to the past 12 months, do you expect to spend more or less money on major purchases (furniture, electrical/electronic devices, etc.) over the next 12 months? I will spend… | Q9 |
|---|---|---|---|---|---|---|
| | | | | | Much more ++ | |
| | Yes, now is a good time + | Ja, der Augenblick ist günstig + | Sì, il momento è piuttosto favorevole + | Oui, le moment est plutôt favorable + | A little more + | |
| | Yes and no: there is no such thing as a right or a wrong time = | Halb, halb: der Augenblick ist weder günstig noch ungünstig = | Sì e no, non è né favorevole né sfavorevole = | Moitié-moitié, ni particulièrement favorable ni défavorable = | About the same = | |
| | No, now is not the right time to make major purchases; it is better to put it off until a later date - | Nein, der Augenblick ist eher ungünstig, es ist besser, den Kauf aufzuschieben - | No, il momento è piuttosto sfavorevole, è meglio rimandare gli acquisti importanti - | Non, le moment est défavorable, il serait préférable de reporter les achats - | A little less - | |
| | | | | | Much less -- | |
| | Don't know 0 | Weiss nicht 0 | Non so 0 | Ne sait pas 0 | Don't know 0 | |

Some of the Swiss questions have their equivalent in the European questionnaire. We do not analyse them in detail here, as they generally have the same semantic content. The following table (Table 2) gives the equivalent numbers of these questions.

*Table 2 – Equivalence between Swiss and European questions not retained in the index*

| number CH | dimension | number EU |
|-----------|-----------|-----------|
| Q1.1 | general economic development of economic situation in Switzerland, last 12 months | Q3 |
| Q2.1 | prices change direction, last 12 months | Q5 |
| Q2.2 | prices change direction, next 12 months | Q6 |
| Q3.2 | number of people unemployed in Switzerland, next 12 months | Q7 |
| Q5.1 | Household's ability to make ends meet | Q12 |
| Q5.3 | Household's ability to save money, next 12 months | Q11 |

Only one question in the Swiss questionnaire, on the general development of job security (Q3.1), does not appear in the European questionnaire.

Table 3 shows the questions in the European questionnaire that are not included in the Swiss questionnaire, with their number according to the European user guide and their frequency (M=monthly, Q=quarterly, in January, April, July and October).

There are monthly questions on estimated past (Q5.1) and future (Q6.1) inflation (reference period of 12 months), the previously mentioned question (Q9) on the timing of major purchases included in the European aggregate index but not in the Swiss one, and another on the current opportunity/adequacy to save (Q10). Three other questions are asked each quarter, and concern purchases and expenses envisaged in the next 12 months, such as the acquisition of a car (Q13), a house (Q14), and major spending on home improvements or renovations (Q15).

*Table 3 – European core questions not included in the Swiss questionnaire*

| number EU | frequency | UE, english version, from the bcs user guide |
|-----------|-----------|----------------------------------------------|
| **Q5.1** | **M** | **By how many per cent do you think that consumer prices have gone up/down over the past 12 months? (Please give a single figure estimate).** |
| | | consumer prices have increased by…% |
| | | consumer prices have decreased by …% |
| **Q6.1** | **M** | **By how many per cent do you expect consumer prices to go up/down change in the next 12 months? (Please give a single figure estimate).** |
| | | consumer prices will increase by…% |
| | | consumer prices have decrease by …% |
| **Q9** | **M** | **Compared to the past 12 months, do you expect to spend more or less money on major purchases (furniture, electrical/electronic devices, etc.) over the next 12 months? I will spend…** |
| | | much more ++ |

| | | |
|---|---|---|
| | | a little more + |
| | | about the same = |
| | | a little less - |
| | | much less -- |
| | | don't know 0 |
| **Q10** | **M** | **In view of the general economic situation, do you think that now is...?** |
| | | a very good moment to save ++ |
| | | a fairly good moment to save + |
| | | not a good moment to save - |
| | | a very bad moment to save - |
| | | Don't know 0 |
| **Q13** | **Q** | **How likely are you to buy a car over the next 12 months?** |
| | | very likely ++ |
| | | fairly likely + |
| | | not likely - |
| | | not at all likely -- |
| | | don't know 0 |
| **Q14** | **Q** | **Are you planning to buy or build a home over the next 12 months (to live in yourself, for a member of your family, as a holiday home, to let etc.)?** |
| | | yes, definitely ++ |
| | | possibly + |
| | | probably not - |
| | | no -- |
| | | don't know 0 |
| **Q15** | **Q** | **How likely are you to spend any large sums of money on home improvements or renovations over the next 12 months?** |
| | | very likely ++ |
| | | fairly likely + |
| | | not likely - |
| | | not at all likely -- |
| | | don't know 0 |

In America, the Michigan Consumer Sentiment Index[7] is based on 5 questions, 4 of which are relatively similar to the Swiss questionnaire (table 4).

*Table 4 - Wording of the Swiss and Michigan index questions*

| number | Switzerland, english version | USA, Michigan |
|---|---|---|
| **Q1.2** | **How do you think the general economic situation will develop over the next 12 months? Will it** | **Now turning to business conditions in the country as a whole--do you think that during the next 12 months we'll have <u>good</u> times financially, or <u>bad</u> times, or what?** |
| | Significantly improve ++ | Good times ++ |
| | Marginally improve + | Good with qualifications + |
| | Remain virtually unchanged = | Pro-con = |
| | Marginally deteriorate - | Bad with qualifications - |
| | Significantly deteriorate -- | Bad times -- |
| | Don't know 0 | Don't know 0 |
| **Q4.1** | **How, in your view, has your household's financial situation changed over the last 12 months? Has it…** | **We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are <u>better off</u> or <u>worse off</u> financially than you were <u>a year ago</u>?** |
| | Significantly improved ++ | |
| | Marginally improved + | Better now + |
| | Remained virtually unchanged = | Same = |
| | Marginally deteriorated - | Worse off - |
| | significantly deteriorated -- | |
| | Don't know 0 | Don't know 0 |
| **Q4.2** | **How do you expect your household's financial situation to change over the next 12 months? Will it…** | **Now looking ahead--do you think that <u>a year from now</u> you (and your family living there) will be <u>better off</u> financially, or <u>worse off</u>, or just about the same as now?** |
| | Significantly improve ++ | |
| | Marginally improve + | Will be better off + |
| | Remain virtually unchanged = | Same = |
| | Marginally deteriorate - | Worse off - |
| | Significantly deteriorate -- | |
| | Don't know 0 | Don't know 0 |
| **Q5.2** | **Do you think that now is a good time to make major purchases (major household appliances, furniture, car, consumer electronics, etc.)?** | **About the big things people buy for their homes--such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or a bad time for people to buy major household items?** |
| | Yes, now is a good time + | Good + |
| | Yes and no: there is no such thing as a right or a wrong time = | Pro-con = |
| | No, now is not the right time to make major purchases; it is better to put it off until a later date - | Bad - |
| | Don't know 0 | Don't know 0 |

| | | Looking ahead, which would you say is more likely - that in the country as a whole we'll have continuous good times during the next 5 years or so, or that we will have periods of widespread unemployment or depression, or what? IF R ANSWERS IN COMPARATIVE TERMS, I.E., "BETTER," "WORSE," OR "SAME," PROBE: "Would that be good times or bad times?" |
|---|---|---|
| | | Better + |
| | | Same = |
| | | Worse - |
| | | Don't know 0 |

In the Michigan questionnaire, the question 5.2 about major purchases is general ("do you think now is a good or a bad time for people to buy...."), and relating to the present moment, as in the Swiss questionnaire and contrary to the European practice, as previously stated.

For Q1.2 and Q4.1 and Q4.2, the wording is slightly different, probably adapted to local contexts (e.g. "business conditions" in the US, and "general economic situation" in Switzerland). There are also three categories of responses in the US for the household's financial situation, compared to five in the Swiss questionnaire. Here again, a swiss trade-off was probably made between comparability with the European and American questionnaires.

The 5th question is about the development of the US economy over the next 5 years. A question that has never been asked in Switzerland in any of the versions (1972, 2009, 2019).

As for the rest of the American Michigan questionnaire, it includes several questions that are relatively specific to the American context, such as the evolution of the price of petrol over 12 months and 5 years respectively, and the evolution of house prices or interest rates.

More interestingly, it also includes several questions that are part of the European survey and not included in Switzerland (table 3), such as those concerning the evaluation of future inflation (12-month and 5-10 year time horizon), or the purchase of a house and a car.

Concerning the Conference Board Consumer Confidence Index (CCI), the questions are completely different. The CCI is based on five questions (Merkle, Langer and Sussman 2004), which will be described in the chapter dedicated to the index's construction. The rest of the questionnaire is also relatively adapted to the specificities of the American context, with a large focus on economy and economic policies. We do not analyse them in detail here, because of their completely different structure and wording from the other surveys under review.

*Recommendations*

The Swiss questionnaire is very much aligned with the European questionnaire and has a high degree of semantic comparability with the Michigan survey, at least on the core items. Given the limited size of the Swiss questionnaire and the low costs of such a change, we would advise to include also the not yet adopted 7 items of the European questionnaire (Table 3). The inclusion of the European question Q9 about major purchases seems to us to be a real priority, to be able to replicate exactly the European index. These additions would make it possible to study the evolution of these other items, but also to adapt the index quickly in case of change, at low cost.

> **Adding additional questions?**
>
> The current questionnaire is relatively short. When moving to a CAWI only survey, which is what we recommend later on, it would in our view be easily possible to add some more questions to the survey that are used in other studies on consumer sentiment or that might be of general interest for the SECO beyond consumer behaviour without having a substantive impact on the response rates or on the cost. SECO could then also chose for how long such questions should be in the field in a very flexible way.

## 2.2 Methodology - sample Size, sample design and sample base, survey mode

At the European level, the project is presented as a standardised survey, but in fact this harmonisation is mainly about the questionnaire, the timing (monthly 12 questions and quarterly 3 questions), the construction of the aggregate index, the compilation and the transmission of the results. Methodological differences and the diversity of actors carrying out these projects are wide, as noted in their respective methodology sheets. We summarize here the information contained in the methodological sheets on the official website[8] of the European Union.

Starting with the survey organizers, sometimes they are the national statistical office (France, Italy, Netherlands), but also polling institutes (Germany, Austria) or the central bank (Belgium).

The sample size is at least 1000 monthly observations according to the European guidelines (List of best practice) but ranges between 1500 and 3000 in the countries under review.

The sampling framework is, however, very different: surveys are sometimes based on an exhaustive population register (Denmark, Germany, Netherlands), but also on a fixed telephone register of unblinded numbers (Belgium, France, Italy), a mix of fixed-mobile telephone register and an online panel in the case of Austria.

The European countries reviewed all have stratified sampling, usually with regional and demographic quotas.

The method of data collection differs greatly from country to country, CATI (Computer Assisted Telephone Interview) for Belgium, Italy, France and the Netherlands, a CATI-CAWI (Computer Assisted Web Interview) mix in Austria and Denmark and CAWI alone in Germany.

Efforts to reduce non-response are also very different, with 5 to 15 telephone contacts and an email reminder in Austria, whereas in the other countries the main objective is to reach a given number of cases surveyed, with simple replacement of non-respondents. The notion of response rate itself is also very variable; it is indicated in Denmark, France and the Netherlands (between 60% and 70%), although the method of calculation is not at all clear, whereas in other contexts no attention is paid to it at all, the declared objective being the number of interviews.

Two practices are then noted in the calculation of the indices. Austria, Belgium, Denmark, France and Germany apply a weighting of the results, according to socio-demographic and geographical parameters. Italy and the Netherlands do not apply a weighting scheme.

With regard to the US confidence consumer indexes under review there are also major differences in methodology.

The Michigan Consumer Sentiment Index (MCSI)[9] is based on a monthly sample of 500 telephone respondents using random digital dialing (RDD) of all possible mobile phones. Any single monthly sample is composed of two parts, an RDD sample of cell telephone subscribers selected in that month and a sample of RDD sample cell telephone subscribers who were surveyed six months previously. The latter

---

[8] https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/business-and-consumer-surveys/methodology-business-and-consumer-surveys/metadata-partner-institutes_en

[9] https://data.sca.isr.umich.edu/survey-info.php

is referred to as the re-contact sample, and the former the RDD sample. The total sample for this rotating panel is normally made up of 60% new respondents, and 40% being interviewed for the second time.

The Conference Board's Consumer Confidence Index (CCI)[10] is based on a monthly sample of 3,000 respondents from an online commercial panel (Toluna) in four weekly waves. As is usually the case for this type of online panel, there are no reference to the construction of the sample, neither to the coverage of the sampling frame, nor to the response rate of the surveys. The experiences at FORS have shown great heterogeneity in the quality of these methodological setups. This is not necessarily a sign of poor sample quality, but it is impossible to comment on it without further information.

In both US cases, weights are used to correct the representativeness of the surveys, based on socio-demographic and regional criteria, but also telephone ownership in the case of the Michigan survey.

All in all and concerning the European surveys and the Michigan consumer confidence index, we have in all cases surveys based on random samples. These are drawn from very different sampling frameworks, with a variable coverage of the resident population. For some countries exhaustive population registers are used, for others the survey is based only on households with a fixed telephone listed in the telephone directory, whereas in the US they are focusing on all possible mobile phone numbers. Of course, these sampling frames also illustrate the variety of data collection methods, often a telephone survey conducted on landlines, but also mobile phones as in the US and in Austria, or web surveys, or a mix of these methods, as in Austria.

The survey design also varies, with a cross-sectional survey repeated monthly in the European countries, while the Michigan survey is more like a two-wave rotating panel, allowing to reduce the sample size while keeping the same estimation variance. This design comes at the cost of greater complexity in the management of the surveys, in the weighting, but also in the calculation of variances, steps that must be done separately for the two sub-samples.

Most countries make corrections based on socio-demographic and geographic criteria to achieve representativeness at least on these basic criteria, but this is not the case for all.

From this perspective, and in the short to medium term, if these methodological practices are stable over time, there is strong evidence that the bias towards the 'true' values of consumer confidence remains constant in each context. This makes them, a priori, rather equally valid instruments for measuring variations over time in each context. It is in the longer term that these methodological devices are likely to prove relatively unequal in the study of variations over time, due to a bias that is likely to evolve. One thinks of the coverage of populations with fixed-line telephones, which is undergoing a long erosion as it is being replaced by mobile telephones, particularly among young people or in urban areas. The fact that almost no effort is made to convert non-respondents in some countries may also be problematic in the long term, with a tendency towards less acceptance of the surveys and non-randomly distributed non-response. On the contrary, it is highly likely that Internet surveys, which were biased towards young, well-educated, and rather urban populations in the past, will tend to be less and less selective in the future, due to the spread and increasingly widespread use of the Internet among all strata of the general population.

*Recommendations*

Current Swiss practices are among those that generally guarantee a limited bias towards the 'true value' of consumer confidence, with a random sample based on an exhaustive population register, recall attempts in the event of non-contact, and a mixed-mode strategy that allows the greatest number to respond. As outlined above, this is already a good guarantee of quality, which is appreciable. On the other hand, with regard to the frequency of data collection and publication of the index, it should be noted that the calculation of the Swiss index on a quarterly basis is below the standard of the other

---

[10] https://conference-board.org/pdf_free/press/TCB_CCS_TechNote_May2021.pdf

countries under consideration, all of which offer a monthly index, allowing a much closer monitoring of the evaluation of consumer confidence. From this perspective, we would recommend that the SECO align itself with the practices of the other countries considered in the international comparison.

In Chapter 3, we will return to several alternatives in terms of methodological design that would make it possible to maintain this excellent quality, at a reasonable price, while adapting to future conditions for such surveys in the medium and longer term.

## 2.3　Construction of indexes and aggregate sentiment indicators

All consumer sentiment surveys we studied use the balance statistic to compute indexes for each question. For constructing aggregate sentiment indicators, most surveys use a simple average of the balance statistics of a small number of individual questions in the questionnaire. This selection and procedure largely corresponds to the method used for the Swiss consumer sentiment index. In what follows, we discuss the computation of the individual and aggregate indexes in detail, highlight some exceptions, and derive some recommendations for Switzerland.

### *Computing indexes for each question*

Most questions limit possible answers to three, four, or five categories. Based on these categorical answers, all institutes we investigated calculate some form of a balance statistic for a particular question. If a question has four (without neutral) or five possible categories, institutes use the following formula:

Balance = (PP + 1⁄2 P) - (1⁄2 M + MM),

where the variables on the right side of the equation give the share of people with a very positive (PP), positive (P), negative (M), or very negative (MM) response.[11] Therefore, very positive or very negative responses receive a larger weight than positive or negative responses.

In Europe virtually all questionnaires offer more than three possible answers. In the U.S. three categories are more common. In this case, the balance statistic is based on the formula:

Balance = P - M.

Although most questions ask about the twelve months change of the situation, the indexes may comprise a seasonality. Therefore, many institutes compute seasonally adjusted indexes. Only the indexes for Denmark, the UK (PwC) and the US (Michigan Survey) are not seasonally adjusted. There exists no uniform method to seasonally adjust the data. To the best of our knowledge, most institutes seasonally adjust the indexes for the individual questions before computing an aggregate index.

### *Computing aggregate sentiment indicators*

When computing an aggregate index, the question arises how much weight to give to the individual indexes. Most countries select a subset of questions, implicitly attributing a weight of zero to the excluded questions. The selection is usually based on an in-sample statistical analysis or common sense. The number of questions included in the aggregate indexes is often quite small compared to the overall number of questions (see Table 5). Most institutes compute an unweighted average of the selected indexes. Only a few institutes determine the weights based on statistical criteria.

---

[11] As indicated in the previous section, these shares may be based on weighted responses accounting for the stratified nature of the sampling schemes.

Table 5 - Construction of sentiment indicators

| | Questions | US (U Mich.) | US (Conf. B.) | UK (GfK) | UK (PwC) | EC | Germany | Italy | France | Denmark | Netherlands | Belgium | Switzerland |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Personal financial situation | Financial situation over past twelve months | 3 | | 5 | | 5 | | 5 | 5 | 5 | 5 | | 5 |
| | Current financial situation | | | | | | | | 5 | | | | |
| | Expected change in financial situation over next six/twelve months | 3 | 3 | 5 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| Business conditions | Change of general economic situation over past twelve months | | | 5 | | | | 5 | 5 | 5 | 5 | | |
| | Current general economic situation | | 3 | | | | | | | | | | |
| | Expected change in general economic situation over next six/twelve months | 5 | 3 | 5 | | 5 | | 5 | 5 | 5 | 5 | 5 | 5 |
| | Expected change in general economic conditions over next five years | 3 | | | | | | | | | | | |
| Employment conditions | Current number of available jobs | | 3 | | | | | | | | | | |
| | Expected change in number of available jobs over next six months | | 3 | | | | | | | | | | |
| | Expected change in unemployment over next twelve months | | | | | | | 5 | 5 | | | 5 | |
| Large purchases | Current buying conditions for consumer durables | 3 | | 3 | | | 3 | 3 | 3 | 3 | 3 | | 3 |
| | Expected spendings for consumer durables over next twelve months | | | | | 5 | | | | | | | |
| Savings conditions | Current condition for savings | | | | | | 4 | 4 | 4 | | | | |
| | Expected savings over next twelve months | | | | | | | 4 | 4 | | | 4 | |
| | Number of questions in index / Total number of questions in questionnaire | 5 / ~ 40 | 5 / 5 | 5 / 14 | 1 / 1 | 4 / 15 - 23 | 3 / 17 | 9 / 15 | 8 / 23 | 5 / 15 | 5 / ? | 4 / ~ 18 | 4 / 11 |
| | Aggregation | Average | Average | ? | | Average / country weights | Regression Analysis | Average | Factor Analysis | Average | Average | Average | Average |
| | Seasonal Adjustment | No | Yes (X-12) | ? | No | Yes (Dainties) | Yes | Yes | Yes | No | Yes | Yes (TRAMO-SEATS) | Yes (X-13 ARIMA SEATS) |
| | Frequency | Monthly | Monthly | Monthly | Irregular | Monthly | Monthly | Monthly | Monthly | Monthly | Monthly | Monthly | Quarterly |

Notes: The numbers indicate how many response categories are coded. ? means unclear. Only the Conference Board considers a time horizon of six months. Total number of questions does not include quarterly and personal questions. The countries depicted deviate slightly from the selected countries in sections 3.1 and 3.2 as we focus on the available non-harmonized indizes to provide additional information on how different countries calculate the aggregated index.

Most aggregate indexes are based on up to five questions relating to the financial situation, business conditions, and purchases of large durable goods. Often, the indexes include questions about the change over the past and future expected situation. Both major consumer sentiment indexes, the Michigan Survey for the United States and the harmonized indexes by the European Commission (EC), follow this general rule.

There are a few exceptions. First, one index for the UK (PwC) is based on a single question about the expected change in the personal financial situation. Second, the non-harmonized index for France includes additional questions about labour market and savings conditions. Third, the non-harmonized index for Italy includes the same questions as the French index and an additional question on the assessment of the current financial situation. Fourth, the non-harmonized index for Belgium is calculated based on four questions about the change of the future expected situation.[12] Finally, one index for the US (Conference Board) includes questions on employment conditions rather than on purchases for large durable goods.

We find various justifications for the selection of questions. The EC selects the individual questions so that the aggregate tracks real private consumption growth (at various levels of geographical aggregation) and has "a solid theoretical foundation" (see EC, 2018). Italy selects the questions that are "best suited to assess consumer optimism/pessimism" (see Istat, 2021). The Michigan Survey selects the questions based on the following criteria: economic significance, statistical adequacy, consistency of timing at business cycle peaks and troughs, conformity to business expansions and contractions, smoothness, and prompt availability (see Curtin, undated). This corresponds largely to the desirable properties of leading indicators (see Marcellino, 2006).[13] For most countries, however, no clear justification is given.

Before calculating the aggregate index, some countries, Germany for example, standardize the individual indexes to make them more comparable in terms of their mean and variance (see Bürkl, undated). The Michigan Survey is a special case. They normalize the aggregate index rather than the disaggregate questions.

Most countries then compute a simple average of the selected questions. However, there are two exceptions that determine the weights of the various questions using statistical procedures. First, the weights for Germany's non-harmonized index are determined in a regression analysis with real private consumption growth as the dependent variable and selected balance statistics as independent variables. However, the potential regressors only include a subset of all variables in the survey. The aggregate index therefore reflects the explanatory power of the information comprised in the selected individual indexes for year-on-year consumption growth. The change in consumption growth corresponds to the index value divided by 10. For example, an index of 5 points suggests that consumption will increase by 0.5% year-on-year (see Bürkl, undated). We suspect that this somewhat cumbersome interpretation is caused by either a normalization of the weights or the index.

Second, the non-harmonized index for France is calculated using factor analysis of 8 indexes. The method extracts a common component that explains most of the fluctuations in the underlying series (see, e.g., Stock and Watson, 2002). In addition, it cleans the data from uncorrelated noise. The method does not ensure that the resulting indicator is highly correlated with a target variable, similar to computing a simple average of the indexes.[14]

---

[12] By contrast, the Michigan survey publishes a separate index based only on questions about the future situation.

[13] Marcellino (2006) highlights the following desirable properties for leading indicators: consistent timing (anticipate peaks and troughs); conformity to the general business cycle; economic significance; statistical reliability of data collection; prompt availability without major later revisions; smooth month to month changes

[14] It is possible, however, to relate the factors to a target variable in a second step (see Kaufmann and Scheufele, 2017).

*Geographical aggregation*

One reason why the consumer opinion questionnaires are harmonized across Europe is that the EC calculates geographical aggregates for the EU and the euro area. The EC computes a weighted average of the country-level results. The weights correspond to the share of the member states in EU-wide real private consumption, smoothed with a 2-year moving average (see EC, 2021).

*Recommendations*

The SECO's methodology to compute indexes for individual questions and the aggregate consumer sentiment index are broadly in line with the methodology by the European Commission (EC) and the Michigan Survey. In addition, Bierbaumer-Polly et al. (2019) evaluated the correlation alternative aggregate Swiss indexes with macroeconomic data. They find that the current selection of questions performs at least as well as the previously used definitions. Further harmonization with the EC or the Michigan Survey is therefore unlikely to make a big difference. In addition, there may be a trade-off between harmonization and using the best method. Therefore, in our view, there is no pressing need to adjust the methodology in order to make the Swiss consumer sentiment index more comparable to other measures.

In what follows we still provide two interesting avenues for future research:

- The current index includes only a small selection of the series collected in the survey, consistent with the EC's methodology. It may be interesting to investigate whether exploiting all series of the survey, for example with a factor model or a weighted average, will provide an aggregate index that better tracks the desired target variable.[15]
- Other countries distinguish between an index about the current and future situation. In future research, it may be interesting to evaluate whether an index based only on forward-looking questions exhibits a stronger lead on macroeconomic data than an index based on backward-looking questions.

## 2.4 Flash estimates

This section discusses the procedure for Flash estimates, that is, preliminary consumer sentiment index releases based on incomplete data, by the EC and the Michigan Survey. We also evaluate the accuracy of the Michigan Survey Flash estimates. So far, Switzerland does not systematically publish Flash estimates, except during severe crises (see SECO, 2020). Against the backdrop of the evaluation of the Michigan Survey, we therefore provide recommendations for a possible systematic Flash estimate for the Swiss consumer confidence index.

*Procedures*

Preliminary results of the consumer sentiment surveys are published in the US (Michigan Survey) and the harmonized program by the EC. The EC publishes EU- and euro area-wide Flash estimates one week ahead of the detailed consumer survey results (i.e. around the 20th of each month; see EC, 2021). The Flash release only reports the aggregate consumer confidence index. The results are based on the countries where the final results are available. In September 2021, for example, the Flash estimate was based on 26 EU countries and all 19 euro area member states (see EC, 2021a). The field work is typically conducted during the first two to three weeks of the month. The final results are released at the end of the month. The Business and Consumer Survey team of the EC confirmed in e-mail correspondence that the only reason why the Flash estimates are revised is that additional countries report their results

---

[15] One has to keep in mind, however, that every procedure that aims to determine the weights based on statistical analysis will be subject to revisions because of estimation error and structural breaks. In addition, these procedures typically maximize the in-sample fit of the index with a target variable, which may not work well out-of-sample.

to the EC. However, the Flash estimate does not include countries where the field work has not yet been finished.[16]

By contrast, the Michigan Survey publishes preliminary results in the middle of the month. These Flash estimates differ from the EC because they are based on the subset of responses rather than regions. That is, the field work has not yet been finished when the preliminary results are released. The main reason for revisions is therefore sampling error and incomplete coverage of the month, although the latter may be relevant only during sharp economic downturns. The final results are published at the end of the month. Recall that the number of interviews in the final results is quite small (500). Therefore, we suspect that the preliminary releases are subject to substantial sampling error. Curtin (2002) suggests, however, that these preliminary results are quite accurate predictors of the final results.

### Accuracy of the Michigan Survey

We conduct an analysis of the preliminary releases of the Michigan Survey from 2000-2021.[17] The study by Curtin (2002) uses data from 1992-2002. We provide more recent results for three reasons. First, the period after 2000 was characterized by more economic volatility, which may affect the accuracy of preliminary estimates. Second, we provide results for the aggregate consumer sentiment index as well as for individual questions. Third, we computed additional statistics to judge whether the preliminary estimates detect turning points.

We transcribed preliminary and final data for the aggregate consumer sentiment index, as well as, for the individual questions underlying the index from the preliminary and final reports (Michigan Survey, 2021a). We then computed the difference between the final and preliminary release. Thereby, we assume that the final release is indeed an accurate measure of the true value. Although this is standard in the forecasting literature, this is of course not necessarily the case, because even the final release is subject to sampling error (see section 3.5).

*Table 6 - Accuracy of preliminary releases Michigan Survey*

| Questions | Bias | RMSE | MAE | Share correct change | ACF(1) | ACF(2) | R2 |
|---|---|---|---|---|---|---|---|
| Consumer sentiment | 0.35 | 1.56 | 1.18 | 0.88 | 0.06 | 0.13 | 0.08 |
| Personal financial situation | 0.77 | 3.20 | 2.41 | 0.79 | 0.12 | 0.04 | 0.03 |
| Expected personal financial situation | 1.51 | 2.68 | 2.03 | 0.80 | 0.03 | 0.03 | 0.07 |
| Business conditions in 1Y | 0.52 | 3.98 | 3.06 | 0.88 | 0.11 | 0.14 | 0.06 |
| Business conditions in 5Y | 0.04 | 3.56 | 2.67 | 0.82 | -0.03 | 0.09 | 0.01 |
| Buying conditions durables | -0.41 | 3.12 | 2.39 | 0.80 | -0.03 | -0.02 | 0.01 |

Notes: Statistics based on the revision computed as the difference of the final and preliminary release over the period 2002-2020. Bias: average revision; RMSE: square-root of the average squared revision; MAE: average absolute revision; Share correct: share of periods where the preliminary release indicated the correct sign of the monthly change; ACF(s): autocorrelation of the revision compared to the revision s periods in the past; R2: explanatory power for the revision of information available at the time the preliminary results were released.

Table 6 shows descriptive statistics to evaluate the accuracy of the preliminary releases. We compute whether the error is on average zero (bias), how widely the error is dispersed on average (root-mean-

---

[16] To the best of our knowledge, individual countries do not release Flash estimates.

[17] We do not evaluate the EC flash estimates because incomplete regional coverage does not seem to be relevant in the Swiss context.

squared error, mean absolute error), how often the preliminary release moves in the correct direction compared to the final value in the previous month (share correct change), the autocorrelation function of the errors at various lags (ACF(1), ACF(2)), and how much of the revision we can predict based on observed data (R2). For all statistics, a value closer to 0 is better. The only exception is the share correct change, where a value of 1 would imply that the preliminary release correctly predicts the change in the index 100% of the time.

Focusing on the aggregate index, we see that the preliminary release underestimates the true value on average by 0.35 points. Given that the index ranges from 60 to 140, and the RMSE amounts to 1.6, this bias seems irrelevant. The RMSE and MAE indicate that the revisions are substantial, however. On average the absolute revision amounts to 1.2 points. In addition, the preliminary releases move in the wrong direction quite regularly (12% of the time or more than once per year).[18]

The revisions comprise at least two sources of errors. First, there is sampling error because the number of observations is smaller in the preliminary than in the final release. This sampling error is likely unpredictable noise. Second, there is a timing error because the preliminary release is based on the first half of the month. In a prolonged downturn (upswing) this implies that the index will be revised down (up) repeatedly.

If most of the revision is caused by sampling error, we would expect the aggregate preliminary release to be more accurate than the individual preliminary releases. Cross-sectional averaging over multiple questions may cancel some uncorrelated sampling error from the preliminary releases. Indeed, the RMSE and MAE are between 40% and 60% lower for the aggregate index than for the individual questions. In addition, the preliminary release predicts the correct change in the index 88% of the time, which is higher or equal than for all individual questions.

In addition, sampling error should be uncorrelated over time and with any other information available. By contrast, the timing error should be autocorrelated because business cycles are quite persistent. In addition, it may be predictable based on other information available at the time of the Flash estimate. Focusing again on the aggregate index, we see that the autocorrelations at lag 1 and 2 are economically quite small and, as we show in the Appendix, mostly not statistically significantly different from zero (see Figure A.1 in the Appendix).[19] In addition, we estimated a simple forecasting model to predict the revision error based on observed past data (lagged survey results as well as initial unemployment insurance claims). Although the coefficients in these forecasting models are sometimes statistically significant, the table shows that the R2 is very low. The systematic part of the revision error we can explain with observed data amounts only to 8%. Of course, there may be other variables that may be able to predict the revision. However, these data are usually released much later than the Michigan survey. Therefore, we believe that the revisions mostly stem from unsystematic sampling error rather than systematic revisions due to the timing.

During the Covid-19 crisis, statistical offices and government agencies started to publish preliminary results because there was a strong need for timely information on the state of the economy. But how should a decision be taken on whether to publish a preliminary estimate? For this, one may construct confidence intervals for the change of the index and examine whether it is statistically significant. We focus on the change because a leading indicator should be useful to detect turning points and signal a change in the state of the economy. Therefore, we may be mainly interested in whether the indicator declined or increased compared to the previous period.

---

[18] This paints a more nuanced picture than Curtin (2002) who reports that the correlation between the preliminary and final releases is close to 1. In the Appendix, we compare our results with Curtin (2002) and find a similarly high correlation. The two statistics answer to different questions, however. The percentage of correct changes is the relevant statistic if we are interested in whether the situation has improved or deteriorated compared to the previous month (turning points). The correlation of the final and preliminary releases is relevant, if we are interested in whether the consumer confidence index is above or below its mean (level).

[19] The corresponding figures for the individual questions can be computed with the R codes we provide.

*Figure 1 - Preliminary estimates of change with confidence intervals*



Having observed the accuracy of the past preliminary releases, that is the RMSE, one can construct confidence intervals for the change of the preliminary index change assuming that the RMSE remains constant over time.[20] One can then judge whether a change signalled by the preliminary release is indeed statistically significantly different from zero. Presumably, one would like to publish a preliminary release only if one is quite certain that the change is indeed statistically significant. Figure 1 shows such an exercise for the Michigan Survey. Here we show 99% confidence intervals; if the interval excludes the zero line, the corresponding change is statistically significant. For the aggregate index, various intervals exclude zero, that is, most changes are statistically significant. By contrast, the change for the individual question, we show the personal financial situation as an example, is rarely significant. This highlights that the preliminary releases are more accurate for the aggregate index than for the individual question.

*Recommendations*

Switzerland has published preliminary results of the consumer confidence index only during periods of severe economic distress. Given the uncertainty associated with Flash estimates, and the wrong signals that can be given as a result, we believe that such a procedure makes sense. Flash estimates have benefits and costs. They may give timely information on the state of the economy during severe crises. However, they are subject to revisions and may therefore give wrong signals during calmer periods. In addition, Flash estimate publications must be commented on which increases the workload.

To what extent publication of Flash estimates makes sense depends also on how quickly the final results are available. How field work is conducted and how data cleaning, preparation and analysis is organised and standardised has a big influence when and how the results are available.

In what follows, we provide some recommendations for future work to formalize the decision to publish Flash releases if SECO considers doing so:

- Based on the past micro data it is possible to evaluate the revision errors for hypothetical Flash releases at various states of information along the lines of the framework outlined in this section.

---

[20] We compute the confidence interval by adding and substracting 2.58xRMSE from the preliminary change of the index. We thus assume that the errors are normally distributed. In addition, we assume that the final release is measured without error. We provide histograms in the Appendix to show that this assumption is reasonably close to reality (see Figure A.1). In addition, we can compute similar confidence intervals for the level of the preliminary release. To preserve space, we provide these graphs in the Appendix.

- Because the past revision errors may be different than the ones based on a new sampling scheme, we recommend to first calculate Flash estimates internally for some time and then evaluate their accuracy ex-post. This would then serve as a basis to decide whether to publish regular Flash estimates or not.
- Even if no regular Flash estimates are released, exceptional Flash estimates may be justified during severe economic crises. During such periods, the signal likely dominates the noise in the data. To easily detect such periods, one could regularly monitor Flash estimates and the corresponding confidence intervals internally to make a more informed decision whether to publish a Flash estimate.
- It is advisable to report Flash estimates for the aggregate index rather than for all individual indexes, because sampling error is smaller due to cross-sectional averaging.
- It may be useful to publish confidence intervals for the Flash estimates based on past revision errors to highlight that they are subject to revisions.

## 2.5   Revisions

This section compares the procedures to deal with structural breaks and revisions due to methodological changes by the European Commission (EC) and the Michigan Survey. We also provide some general recommendations for Switzerland.

*Procedures*

Since 2007 the EC has published information about the timing of past methodological changes, the likely impact on the aggregate index, as well as whether the data have been adjusted for structural breaks.

The revision notes are published in a separate sheet accompanying the data. These notes cover the consumer, business tendency, and investment surveys.[21] Four major types of changes occur: Addition of new countries; changes in the questionnaire; changes in the survey scheme; changes of the survey frequency; changes in the definition of the aggregate index.

The country sample includes the EU member states. This implies that if countries enter or leave the EU, the country weights and the aggregate index are adjusted accordingly. According to the revision notes, adding or removing countries leads to revisions of the series over the entire period. Although, no information is provided how important these revisions were.

The questionnaire of individual countries changes mostly because of harmonizing them with the EC guidelines. Such a change occurred, for example, for Denmark for several questions in 2020 in the business tendency survey. Changes in questions may lead to structural breaks in the series. To make the past series comparable with the future one, they mention that "the series before these dates were revised based on an ARIMA-estimation of the breaks created by the changes." This implies that they performed a model-based retropolation of the unobserved series in the past. They also mention that this changed the long-term average of the past series. Because the revision is based on a model with multiple parameters to be estimated, an accurate retropolation requires a sufficiently long sample where both, the old and new methodology are observed. Thus, the retropolation occurred sometime after the introduction of the new question. No information is provided about how long the corresponding samples were.[22]

Several changes in the survey methodology occurred since 2007. Although the exact changes are not known, they mention how they handled potential structural breaks in the series. The notes mention "backcasting" or "level-shift" procedures. Usually, the old and new methodologies were conducted at

---

[21] Here, we summarize the information from all of those because they may be informative about how to handle structural breaks caused by methodological differences.

[22] Alternatively, one could envisage to estimate the relationship between old and new series by running the two methods in parallel for some time.

the same time over a period of three months up to one year allowing to estimate the average difference and correct the old series.[23] Some methodological changes were introduced without the overlapping surveys. One, but not the only, example was the switch from face-to-face interviews to CAWI and CATI during Covid-19. In these cases, the revision notes warn of potential structural breaks in the series.

As mentioned in section 3.3 the EC changed the definition of the aggregate consumer sentiment index after evaluating how well different definitions match macroeconomic data. This change was introduced in January 2019 and led to a revision of all past values of the aggregate index.

There are multiple breaks due to changes in survey frequency and missing data. First, several country series show a break from quarterly to monthly frequency since 1985 (see Table A.1 in the Appendix). The EC applies a naïve retropolation scheme by repeating the quarterly value for every month of the quarter. Of the 33 countries participating in the program, we identified 11 country series showing such a break. To the best of our knowledge, no retropolations other than the naïve one has been applied in the harmonized consumer sentiment index.[24] Second, for some countries the surveys have been suspended for some months during the past. For example, Portugal did not collect data on one question during several months in 1997. In these cases, the value for the country aggregate is missing, but the EU aggregate is still computed. This confirms that the EC does not perform systematic interpolation of missing data. From a statistical point of view, this procedure is certainly not optimal, as we will discuss in Section 3.6.

In contrast to the harmonized EC survey, the Michigan Survey is not revised ex-post, except for the preliminary results discussed in section 3.4. We checked whether the figures published in some of the original reports published in the past correspond to the time series published today. For example, the June 1978 report states that the index of consumer sentiment amounted to 80, down from 82.9 in May (see Michigan Survey, 1978). These are exactly the same values as we found in the long time series published today (see Michigan Survey, 2021b).

However, this does not imply that the Michigan survey did not change the methodology over time. Instead, "to correct for sample design changes" they add a constant to the current index rather than revising the past series (see Michigan Survey, undated). This constant changed in 1972 and 1981. We do not know for how long the old and new methodologies have been used in parallel to determine this constant. In contrast to the EC, this level adjustment is applied to the current aggregate index rather than to the past individual indexes.

In sum, we observe more regular revisions for the EC survey because of the heterogeneous sample of countries and the aim to harmonize the surveys. A list of these changes is reported in the excel file with the most recent data. Changes in methodology that lead to level shifts are often corrected by adjusting the past series at the question level, which leads to revisions of the past series at the aggregate level. By contrast, the Michigan Survey is rarely revised. They make old and new aggregate series comparable by adding a constant to the current aggregate index. Therefore, past data is not revised.

*Recommendations*

In what follows, we list several recommendations. They are based on our reading of the literature and accounting for the fact that methodological changes and adjusted series must be published with a short delay.

- A list of methodological changes (timing, adjustments, and likely impact on past data) could be provided along with the data. This may help data users to identify the timing of structural breaks and incorporate them in their modelling strategies. In our view, the most important categories are:

---

[23] The overlapping data are available upon request.

[24] We performed a visual inspection of past monthly series to detect changes in the volatility in the series that may stem from ex-post retropolations in the country indexes. We did not find evidence for such breaks

changes in the questionnaire; changes in the survey scheme; changes of the survey frequency and retropolated values; changes in the definition of the aggregate index.

- We recommend adjusting the past index rather than the current one. Methodological changes should usually lead to a more accurate index. Therefore, it makes more sense to correct the past series. The disadvantage is that this leads to a revision of all past values of the index. If these revisions occur rarely, and only concern the mean of the series, this will not matter much for most data users.
- It is preferable to adjust individual indexes rather than the aggregate index. This ensures that the aggregate is consistent with the individual indexes and the aggregation method can be replicated.
- It is advisable to compute the results based on old and new methodologies for some time in parallel. This allows to detect and quantify structural breaks in the mean. One can then link the old and new series by adjusting the old series with the average difference . The EC sometimes bases such adjustments only on three months of data. This seems very short to estimate the average difference between the old and new series; a longer period seems preferrable. This implies, however, that the index according to the new methodology can only be published with a longer delay.
- We are sceptical about accounting for a break in the variance. Recall that the measured series comprises a signal (the true index) and noise (e.g. from sampling error). A more accurate method will therefore reduce the variance of the series because it reduces noise. If we would normalize the old series to the same variance as the new one, we would change the variance of the signal as well as of the noise component.[25]
- Alternatively, one could disentangle the signal and the noise with an econometric model and then backcast the series. However, we are generally skeptical about using model-based backcasts of the unobserved series under the new methodology. Model-based approaches imply more parameters to estimate and therefore require a larger sample to perform accurate backcasts. In addition, such backcasts are based on models estimated on modern data and there is no way to evaluate the accuracy of the model for the out-of-sample backcast several decades in the past. We are more optimistic about model-based retropolation of quarterly series because we actually observe a monthly observation every three periods, which allows us to examine the model fit for actually observed past data. We will discuss various retropolation methods in chapter 3.6.

# 3  Study design

This chapter discusses various design issues in a systematic way. When reflecting on design two key aspects have to be discussed:

- Is a measure valid? Validity involves the extent to which the research instrument measures what it is intended to measure.
- Is a measure reliable? Reliability refers to the consistency of the results, when repeated measurements are made, so would you come to the same conclusion if you repeated a study in the same way.

While both aspects are relevant, the main objective to keep in mind for this indicator is reliability, so the indicator/the mean and variation in the mean over time need to be correct in order to detect relevant turning points in consumer sentiment. What is less relevant is that to minimize error/variation in any given study as long as the error and the variation is constant.

A further aspect to consider is that the design of such a study is happening in a cost-quantity-quality triangle. While there may be good arguments that especially a high quality and reliable data collection is costly, budget constraints are always a factor in a study design. Collecting more data in more regular intervals means also higher cost. These costs are not in a simple linear relationship to the number of interviews, since there are also fixed costs irrespective of the number of interviews. Quality measures

---

[25] This is not the case for adjusting level shifts because the estimate of the mean over several periods will average out uncorrelated noise.

usually aim at increasing response rates and decreasing response bias. Here again, there is usually not a linear relationship between the cost of a measure and its effect on response rate. Therefore, a discussion needs to take place what quality measure may have which impact on response rates and response biases in relation to the cost of a measure.

*Figure 2 - The cost-quiality-quantitiy triangle for study design*

The attempt of any study is to find an optimal balance between those three aspects. While a cost estimate is not part of this report, costs considerations are nevertheless an important aspect of the different recommendations.

## 3.1   General Design and Sampling

Currently the sample is drawn from the population register (SRPH) from the Federal Statistical office with a population 15-85 years of age according to the technical report provided by the survey agency. There is no reason to change this as it is the best available sampling base and allows for a stratified sample if the SECO wishes to have regional indicators. We see also no reason to change the age range, with the exception that it is not clear why there should be a limit at 85.

Currently the consumer sentiment index is using a repeated cross-section design. Every first month of a quarter (January, April, July, October) a new sample is released with a three-week fieldwork time. All selected sample members receive an announcement letter and a reminder is sent out to those that have not responded yet. The mode is mixed CATI/web, while web is open for those without a telephone number (ALTEL in the terminology of the federal statistical office). Should SECO not want to change frequency of the data collection, a change in survey mode may be up for discussion nevertheless. This will be discussed in a later chapter.

Change propositions in order to have a more frequent measure of consumer sentiment available that includes different options:

- For a *monthly survey*: For such a survey, it would be possible to keep the current cross-section design and simply launch a new survey wave every month. It would be possible to extend the fieldwork and always then consider the interviews conducted in each month. Alternatively, also for the monthly design, a weekly rolling-cross section could be used since a monthly survey .

- For a *weekly survey*: Having a weekly indicator does need an adaption of the typical cross-section design, unless the fieldwork would be reduced to one week only which is not recommended for a high-quality study. As an alternative, a rolling-cross section design can be envisaged with weekly launches of new samples that are then treated in the exact same way (see box).

- Moving to *a panel design or a rotating panel design*: With such a design, the same respondents are interviewed multiple times. Given that a weekly or even a monthly interview would be too high of burden on respondents, separate panels would need to be set up so that respondents participate in the study in longer intervals, for example four times a year. Such a design would be interesting to study individual changes in consumer sentiment over time and for academic research, however

the design is very complicated and panel attrition (people dropping out of the panel) needs to be controlled and new respondents would be recruited constantly. Such a design makes the fieldwork very complicated and as a consequence also costly and is therefore likely not of interest for the SECO.

*The rolling cross-section (RCS) design*

The RCS design has been used initially to study campaign dynamics in elections (Johnson and Brady 2002). In an RCS design new samples are launched in a regular interval. This can be done daily or weekly. Each sub-sample then must be treated in the exact same way to make sure that each address has the same contact pattern and contact probability. So for example, in a telephone survey the contact scheme has to be the same for each address, or for a web or paper survey with offline recruiting, the reminders have to be sent always in the same intervals for each sub-sample. If this is respected, the conducted interviews in each interval (may it be daily or weekly) are in itself a random sub-sample.

*Figure 3 - The RCS field work design*



*Days/time*
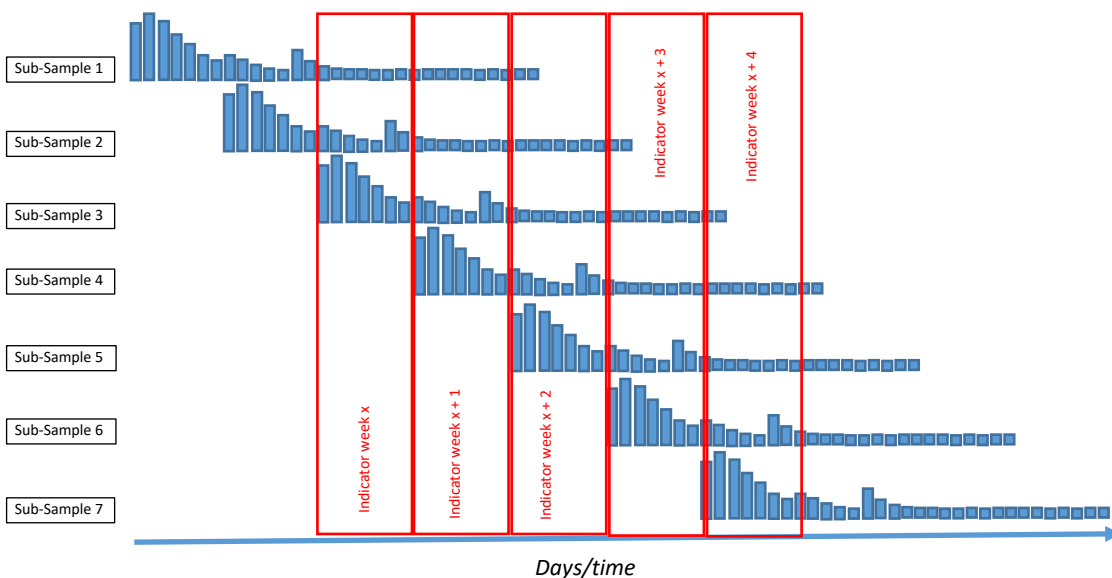
In theory all modes and also mixed mode as well as all intervals can be applied to a RCS design. In practice

- RCS studies typically use daily or weekly intervals, as any other interval is too complicated to handle. Handling daily launches of sub-samples is nevertheless quite demanding for a fieldwork agency.
- The data collection needs to start ahead of when the first measure is to be calculated to a maximum of the normal fieldwork of any sub-sample.

Sub-sample of an RCS are usually smaller than a normal cross-section. And the number of respondents in an interval can also vary substantially, because of public holidays or seasonal effects. This has to be accepted, since measures to increase response rates for some but not all sub-samples should not be done. To be able to deal with this, it is recommended that a slightly larger sample than the absolute minimum is taken.RCS design usually use a single mode. Mixed-modes studies are more complicated and therefore more costly to handle in an RCS design. Face-to-face is almost impossible and also CATI is complicated. With CATI interviews the number of contacts have to be limited, this is part of the self-regulation of survey companies. Because the number of contacts cannot be increased, sophisticated contact schemes have to be put in place in order to keep the contact probability stable. Easiest for such a design are therefore web interviews.

Under the assumption, that SECO wants to move to at least a monthly indicator and that the fieldwork should not be shortened, we suggest moving to a RCS design with weekly sub-samples. Even with a monthly survey you are in the field almost constantly anyway as you are with an RCS. Such a design allows you to aggregate the weeks in a very flexible way for the monthly or quarterly indicator. Under the assumption, that already 400-500 interviews per week are sufficient for a stable mean, that would give you about 1600-2000 interviews per month. Taken into account that you want to have a minimum number of 400 interviews each week, it would make sense to aim for a higher average number to compensate for weekly fluctuations. If you use web-only, the preparation for the letters can be done four times a year (because the sampling frame is only updated four times a year and it does not make sense to draw samples more frequently). In such design you would need to start about 3 weeks prior to when you need the first indicator.

---

**Using incentives?**

Incentives are often used in surveys to boost response rates. Such incentives can be conditional on completion of the survey or unconditional to everybody in the sample. They can be small giveaways or cash or based on a lottery for some price. Research shows that unconditional cash incentives work best to increases response rates, however those incentives are also very costly and politically controversial. We assume it is politically not feasible in this case for a government agency to distribute cash incentives. SECO needs to decide if it wants nevertheless use incentives, e.g. a lottery of several ipads or something similar for those who have participated.

---

## 3.2   Survey mode

Currently the study used CATI interviewing, with a web option (CAWI) for those without telephone number and 25% CAWI, 75% CATI. The following table gives an overview of the different option, which can also be combined as it is currently the case.

| Mode* | Pros | Cons |
|---|---|---|
| **CATI (Telephone)** | Interviewers can moderate = higher response rate, lower item-non-response, comparability with current mode | Decreasing coverage and response willingness, high cost, interviewer effects |
| **CAWI (Online, web)** | High flexibility when to participate, lower cost, data available immediately | Older people with greater difficulty (but problem declines), lower motivation, lower control who participates |
| **Paper** | Good possibility for older people, high flexibility when to fill out interview, low cost (but cost to scan and process data) | Slow, filtering difficult (but not necessary in this case), lower control who participates |

*Theoretically also the "Face to face/CAPI" option: too expensive, not necessary for such a short survey

CATI only becomes increasingly problematic because of decreasing availability of telephone number and response willingness. In addition the persons that cannot be contacted through telephone are not a random selection of the population, especially many younger households have only mobile phones. To combine this with web is a possible solution, but it has some disadvantages, mainly mode effects. It would be possible to keep a mixed-mode design CATI/CAWI with push to CAWI and use CATI only as a follow-up option. However, to what extent this increases the quality is unclear. In addition, a mixed mode design is very complicated to handle in an RCS setting. However, an RCS, as shown in the previous chapter, is almost necessary if SECO wants a weekly indicator.

On the other hand, moving to CAWI or CAWI/paper is an option increasingly used by the Federal statistical office and also in scientific studies.

- We recommend that SECO considers abandoning CATI altogether. While this seems like a bold step to take, it may likely be necessary anyway in the future because of a further decrease in CATI interviewing. We believe with other modes you reach similar and acceptable response rates and response quality while reducing the response bias especially for younger people.
- We recommend that you move to a CAWI only survey. Paper survey would allow you to target the older population. However, in Switzerland with few exceptions also the elderly population has very high internet access and usage rates. And for the elderly over time this will not be a problem anymore and this group is less likely to participate in surveys because of health reasons.
- A CAWI only option would also be the easiest to implement in an RCS design. Announcement letter with personalized access information/QR codes could be sent every week to a random sub-sample of the population, followed by at least one reminder letter 10-14 days after the announcement letter. Possibly a second reminder could be sent as well, but there the cost/benefit ratio is already much lower than with the first reminder letter.
- A paper option would increase the response rate, especially among the older population and may be politically feasible, but it comes at a price. Adding a paper option complicates fieldwork substantially and makes the data collection also more costly. Without a paper option data is available much quicker. So overall we believe that the advantages without a paper option outweigh the addition to a paper option also given that high response rates are not the main objective of the study.
- For a transition phase it would make sense to have both designs (the current and the new) in parallel for a certain period, for example for 6-12 month. This allows you to assess possible variation between the two types of data collections.

## 3.3 Regionalisation

SECO requested the assessment what it would take to calculate the indicator at the regional level as well. There are different regions that could be used:

- Political/administrative units: this could be the three (or four) linguistic regions, the 26 cantons, the seven Eurostat NUTS 2 regions ("Grossregionen" which are the basic regions for the application of regional policies, they do not cut cantonal boundaries), border regions (that would need to be defined).
- According to economic criteria: this could be the WEMF regions (5 regions or 25 economic regions, cut cantons, detailed definition and classification slightly unclear), "BFS Arbeitsmarktgrossregionen" (=16 regions, cut cantons, mainly based on commuting regions where people live and work)

The SECO considers having the indicator also for the NUTS 2 regions as a possibility, but only on monthly or quarterly bases. For this the sample needs to be expanded and stratified, which is further discussed in chapter 3.5.

## 3.4 Comparability

It is impossible to predict theoretically how the new way of data collection affects the indicator compared to the traditional way of collecting the data. Too many different factors influence response rates and response bias and to what extend the different socio-demographic composition of the sample affect the mean can also not be assessed. We believe that the composition of the samples may vary but given that we have a cross-sectional design in both ways of data collection, there is no fundamental difference. Therefore, the only way to assess this is to run pilot studies or to plan for a certain period where both ways of collecting the data run in parallel.

Should SECO want to continue publishing a quarterly indicator to continue the current series, we recommend using the data from the first month in each quarter as it is done currently and not to combine the data from an entire quarter. The data within a quarter can vary and this would make it less comparable to the current way of data collection. Alternatively, SECO can rely on retropolation if it wishes to create a monthly time series that include data before the move to a monthly/RCS data collection.

## 3.5 Sample Size

*Theory: Power analysis to determine sample size*

To estimate necessary (and sufficient) sample size in empirical design, statistical power analysis can be used (Phillips and Jiang 2016, Tomczak et al. 2014). The sample size should be large enough to detect an effect that is substantively significant. The power of a test (e.g., if the mean values of two regions or between two points in time differ) tells us how likely we are to detect a statistically significant effect in the data, given that it exists in the population. The power of a test depends on several factors, including:

- effect size,
- variability of the measure,
- probability of rejecting a true null hypothesis (significance level; Type I error),
- probability of not rejecting a false null hypothesis (Type II error). Statistical power is defined as 1 minus the probability of a Type II error,
- directionality (directional or one-sided vs. non-directional or two-sided hypotheses),
- sample size,
- complex sampling, i.e., involving stratification, unequal probability of selection, and clustering,
- weighting and stratification

Failure to consider the power of a test may lead to errors such as reporting no statistically significant differences between groups while in fact differences exist in the population. Such distortion may result from using a sample size that is too small to reliably detect an effect.

There is a consensus that a larger sample size is better than a small sample size. However, using a sample larger than is needed to obtain reliable results brings no research benefits, but may involve unnecessary respondent burden, tie up research resources, and inflate costs.

*Example of sample size calculation using the consumer sentiment index.*

According to Kemeny and Bachmann (2019), the items q12 q41 q42 q52 form the new index from 3/2019 on. Based on the data file 200734_2018-2021-09092021.sav, Table 1 describes the new index by quarters between wave 8 and wave 19:

*Table 7 - Descriptive statistics of the new consumer sentiment index (wave 8-wave 19).*

| wave | N | mean | sd | se(mean) |
| --- | --- | --- | --- | --- |
| 8 | 1255 | 2.818659 | .4800771 | .0135516 |
| 9 | 1295 | 2.82529 | .5114235 | .0142117 |
| 10 | 1278 | 2.832877 | .4921122 | .0137657 |
| 11 | 1246 | 2.852595 | .5098945 | .0144451 |
| 12 | 1262 | 2.890848 | .5091022 | .014331 |
| 13 | 1299 | 2.819348 | .5165232 | .0143313 |
| 14 | 1895 | 3.12942 | .5565204 | .0127843 |
| 15 | 1372 | 2.890671 | .520521 | .0140527 |
| 16 | 1395 | 2.912485 | .5310599 | .0142186 |
| 17 | 1505 | 2.872093 | .5491354 | .014155 |
| 18 | 1278 | 2.807316 | .5885096 | .0164622 |
| 19 | 1267 | 2.689884 | .5193226 | .0145898 |

The mean value amounts to between 2.819 (wave 8) and 3.129 (wave 14) and the second highest value 2.912 (wave 16) such that the wave 14 value is probably an outlier. The standard deviation of the mean amounts to between .480 (wave 8) and .589 (wave 18) and the standard error to between .013 (wave 14) and .014 (all other quarters, and .016 in wave 18).

Suppose we like to calculate the sample size necessary to investigate if the index decreased significantly on the 5% level (type I error) between wave 16 and wave 17. The null hypothesis assumes no difference. To verify the null hypothesis, we measure sample means as standard error units (= sd /square root(N)). If the sample mean of the wave 16 index falls in the top 5% of the distribution of the wave 17 sample, we reject the null hypothesis. The region of the top 5% of values corresponds to a z-score of 1.645. Specifically, the mean index in wave 17 is with a probability of 95% smaller than 2.872 + 1.645*.014 = 2.895, which is smaller than the mean index in wave 16 (2.912) so we reject the null hypothesis.

The type I error is the chance that we reject the null hypothesis when in fact it is true (power of a statistical test). So, we assume that the null hypothesis of no difference is false, and we want to know the probability of rejecting the null hypothesis. The distance between the actual (true) mean of the sampling distribution (under the alternative hypothesis) and the critical z-value (z = 1.645) (for the sampling distribution under the null hypothesis) equals (2.912-2.872) – 1.645*.014 = .04. The region above z = .04 in the z-table is 0.48, the value of a Type II error. The complementary area in the sampling distribution under the alternative hypothesis is the power of a test (.52). Generally, one assumes a power of .80. However, there is a trade-off relationship between the two error types.

### *Complex sampling, weighting, and measurement error*

As Phillips and Jiang (2016) discuss, measurement error and weighting reduce power and inflate sample size requirements In addition, stratification provides a small reduction in variance of the sampling distribution and sampling weights related to the unequal probability of selection causes a small increase. The biggest inflationary impact on the sampling distribution is usually the clustering in relation to the sampling design (deff). Clustering therefore requires a higher number of interviews to keep the sampling error stable compared to a non-clustered sample. In addition, a small reliability R reduces power and increases sample size requirements beyond the inflation caused by cluster sampling. The sample size requirements due to the combined effect of complex sampling and unreliability are increased by approximately a factor deff/R. For example, we calculate a scale reliability coefficient (Cronbach's alpha) of .50 for the items q12 q41 q42 q52 between wave 8 and wave 19. This means that the required sample size more than doubles.

*Within-respondent stability of the consumer sentiment index*

In repeated cross-sectional surveys, changes of the consumer sentiment index are calculated as differences of two independent samples.[26] In a panel design with partially overlapping samples, the variance of the change can be much smaller than if based on two independent samples. The variance of the change $\hat{\theta}$ is $V(\hat{\theta}) = V(\hat{t}_1) + V(\hat{t}_0) - 2C(\hat{t}_1, \hat{t}_0)$ , where $C(\hat{t}_1, \hat{t}_0)$ is the covariance between $\hat{t}_1$ and $\hat{t}_0$ (Andersson et al. 2011). This means that if there is a high correlation between the two measurements within the same individuals, the variance of the change may be strongly reduced. A caveat is the risk for attrition bias. In practice the correlation tends to be stronger among units responding at both occasions since change tends to be associated with drop out (Voorpostel and Lipps 2011).

*Sample size and time in the field to reach stability of mean index*

In Figure 4, we illustrate the fluctuation of the mean index by quarter over time (x-axes: number of respondents; index limited to between 2.5 and 3.5, number of interviews limited to N=500):

*Figure 4 - Fluctuation of the mean index by wave: Consumer sentiment survey*
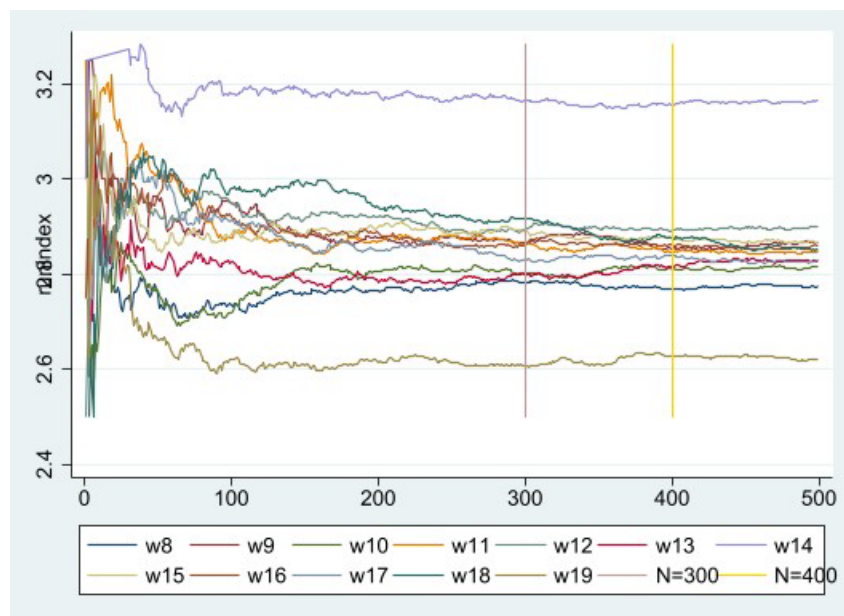


The mean index mostly stabilizes after 300 or 400 respondents, which is similar to the Michigan Index (see chapter 2.4). Instability over time is due to a mix of time effects (see the increasing wave 16 index) and different sample compositions at different times of the fieldwork.

In an RCS design, the sample composition does in principle not vary over time such that changed sample compositions at different fieldwork times are probably less of an issue. In an RCS design there is in principle an arbitrarily high number of interviews available in each time window, as long as enough samples with high enough sizes started at different times. As a tendency, the telephone mode allows for a shorter fieldwork duration. For example, in the current consumer sentiment survey the fieldwork is more or less finished after about 30 days (Figure 5).

---

[26] Further reading: Ployhart and Vandenberg (2010) discuss the decision about frequency and timing of repeated measures. To assess variances and covariances in rates of change in the context of longitudinal studies, Rast and Hofer (2014) investigate the interplay among number and spacing of occasions, total duration of studies, effect size, and error variance on power and required sample size.

*Figure 5 - Fieldwork duration: Consumer sentiment survey*
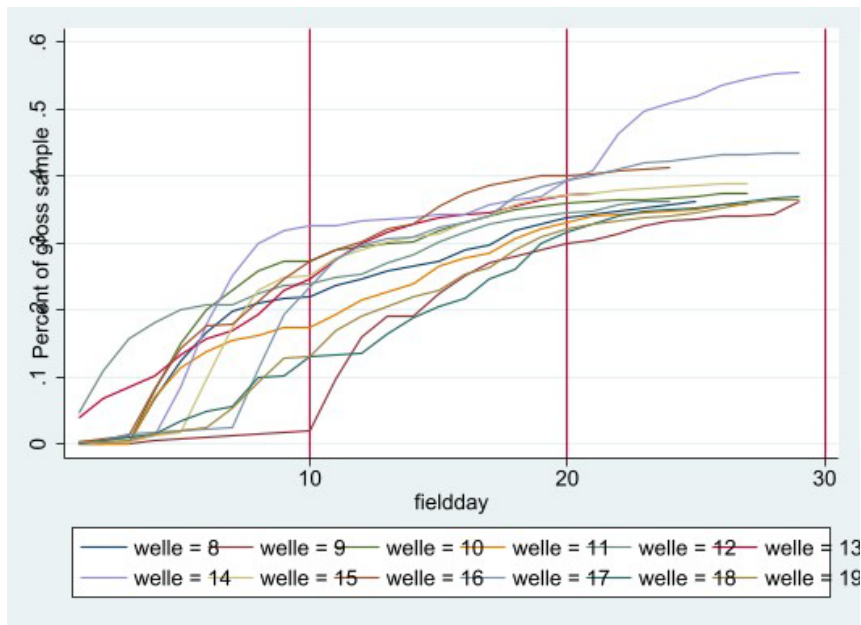


With a comparable response rate, the fieldwork for the MOSAiCH 2020 survey (here: only web respondents; the mail follow-up respondents are dropped) at FORS took a bit longer:

*Figure 6 - Fieldwork duration: MOSAiCH 2020 survey by incentive used.*
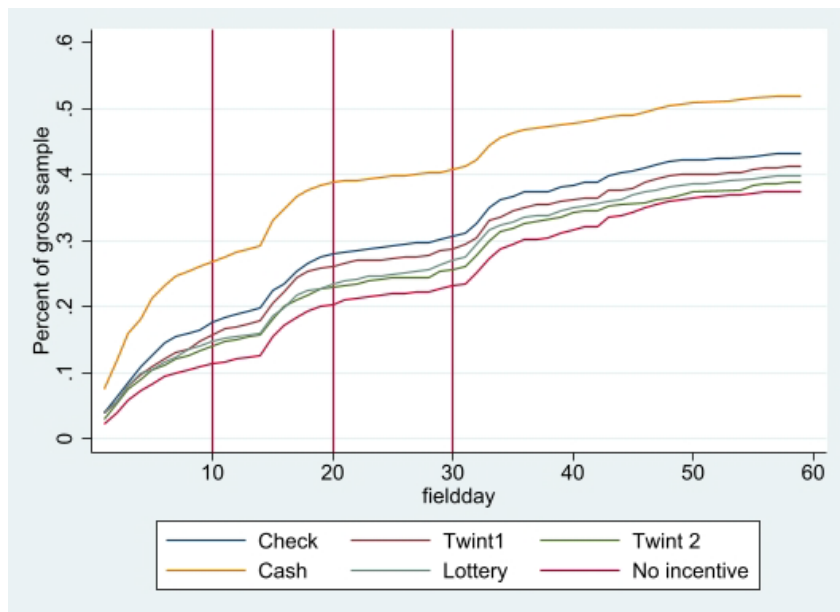


Figure 6 shows also, that in addition to increasing the response rate, using unconditional cash incentives can speed-up fieldwork considerably compared with other incentives.[27]

---

[27] See for more detial Ernst Stähli, Michèle, Sapin, Marlène, Pollien, Alexandre, Ochsner, Michael, and Nisple, Karin. (2021). *MOSAiCH 2020 on Environment and related topics. Survey Documentation*. Lausanne: FORS – Swiss Centre of Expertise in the Social Sciences

As for a regional analysis, we see that the standard deviations are similar across the different Swiss large regions.

*Table 7- Descriptive statistics of the new consumer sentiment index (wave 8-wave 19) by region.*

| Big Regions | N | mean | sd | se(mean) |
|---|---|---|---|---|
| Genferseeregion | 1739 | 2.939141 | .549052 | .0131663 |
| Espace Mittelland| | 3892 | 2.842262 | .5408581 | .0086696 |
| Nordwestschweiz | 1487 | 2.876261 | .5360726 | .0139017 |
| Zürich | 1333 | 2.834146 | .5139998 | .0140782 |
| Ostschweiz | 3086 | 2.870571 | .5431183 | .0097768 |
| Zentralschweiz | 3472 | 2.87512 | .5279111 | .0089592 |
| Tessin | 1338 | 2.9072 | .5311631 | .0145211 |

This means that irrespective of the region, the same number of respondents are necessary to reach the same stability of the mean.

Next, we check the number of observations to reach stable mean values for selected variables using data from the Selects 2015 panel survey. The second wave was conducted as a RCS, with around 120 individuals interviewed daily during 61 days. This means that apart from the first week (the short lines in the figures), about 800 individuals were interviewed in each week. In the absence of an index such as the consumer sentiment index, we check stability across the weeks reported for the variables political interest (Figure 7), voted in the last election (Figure 8), having tertiary education (Figure 9), working fulltime (Figure 10), gender male (Figure 11), and household size (Figure 12). Note that variables are rescaled to range between 0 and 1 and that extreme values (<.2 and >.8) are dropped:

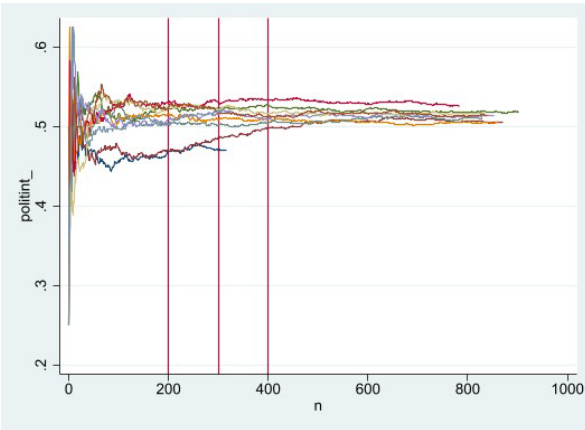*Figure 7 - Political interest in Selects RCS*     *Figure 8- Voted in the last election in Selects RCS*
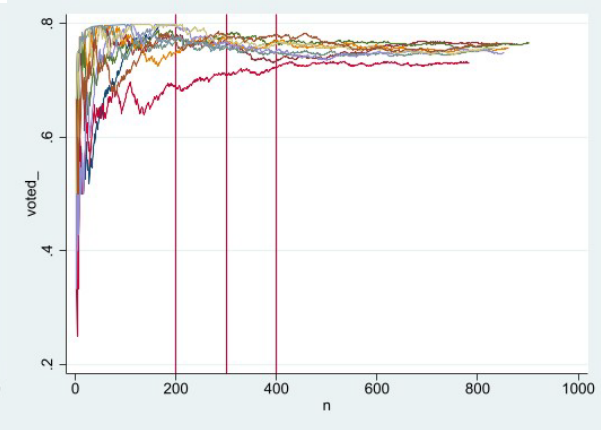
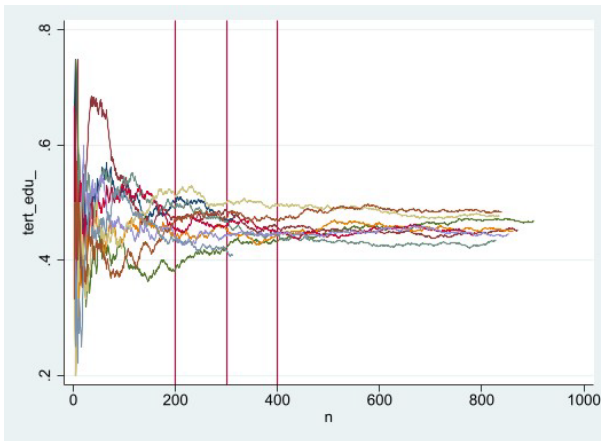*Figure 9 - Tertiary education in Selects RCS*          *Figure 10 - Working full time in Selects RCS*
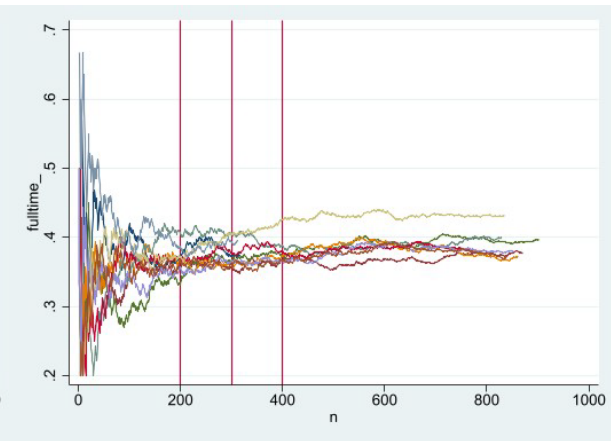


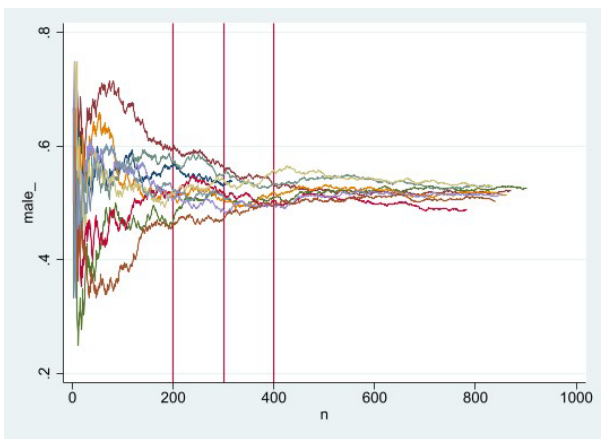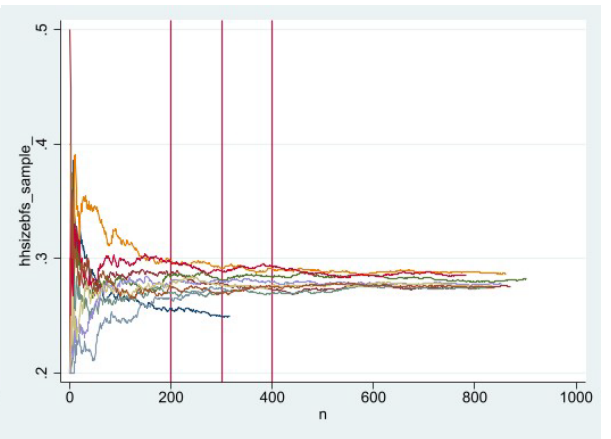*Figure 11- Gender male in Selects RCS*          *Figure 12 - Household size in Selects RCS*

We find that although there are still some fluctuations, most variables show stable means after about 400 observations for most of the weeks.

*Recommendations:*

Empirical considerations show that to reach a stable mean value, about 400 respondents seem to be enough. This finding is based on data from the Seco consumer sentiment index itself, but also from general population web surveys at FORS based on a cross-sectional (MOSAiCH) and a RCS design (Selects RCS). This figure is independent of the temporal (week, month, quarter) and regional (whole of Switzerland, large region) level. This means that the number of respondents should not be below 400 on the smallest aggregation level, such as Ticino (the smallest large region) or on a weekly base (the smallest temporal level). Theoretically,  there would be a need for a much higher sample size than the current sample size given the lower than generally recommended power of .8 and a relatively low reliability of the index.

- Given, that with an RCS design, the weekly number of interviews can fluctuate, it might be feasible to conduct about 500 interviews per week or any unit where an indicator shall be calculated. This would amount to about 2150 interviews per month, which is slightly above the current number of monthly interviews.
- Should SECO wants to have a monthly indicator for each of the 7 Nuts-2 regions and applying the same logic as above, the monthly number of interviews will have to go up to 3500 interview to have

37

500 interviews per region and month, or 875 per week for a weekly RCS. In this case the sampling is stratified, so for the overall indicator design weights have to be applied.

- When SECO wants to have a quarterly indicator for each of the 7 Nuts-2 regions and a monthly representative indicator the same logic as above in principle, at least 500 interviews per region and 3500 in total are needed over three month. If the basis is a weekly indicator with 500 interviews, this target number is reached in each region with the exception of Ticino, where an oversampling is required (see table below). That would then give a total number of 6725 interviews per quarter. Should the basis for the regional indicator be a monthly national indicator, 167 interviews per month and region (so 1166 per month in total) are necessary to get to the 3'500 required interview per quarter.
- the quarterly number of interviews will have to go up to 3500 interview to have 500 interviews per region and quarter as well, or 269 per week (meaning: 39 per region and week) for a weekly RCS. In this case the sampling is stratified, so for the overall indicator design weights have to be applied.

*Table 8- Number of target interviews*

| Nr. | Region | Cantons | Population 31.12.2015 (STATPOP) | Interviews withouth cluserting for 500 interviews per week or month | Interviews per month without clustering with 500 interview weekly | Annual number of interviews | Number of interviews per month or quarter with clustering to allow for regional indicator | Quarterly indicator with weekly samples of 500 | | | Quarterly indicator with monthly sample of 500 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Number of interviews per quarter without clustering with weekly 500 interviews | Need for oversampling in TI | Total Number of interviews | Interview/month | Total |
| 1 | Genfersee-region | GE, VD, VS | 1'593'839 | 96 | 415 | 4'976 | 500 | 1'244 | | 1'244 | 167 | 500 |
| 2 | Espace Mittelland | BE, FR, JU, NE, SO | 1'842'251 | 111 | 479 | 5'752 | 500 | 1'438 | | 1'438 | 167 | 500 |
| 3 | Nordwest-schweiz | AG, BL, BS | 1'128'723 | 68 | 294 | 3'524 | 500 | 881 | | 881 | 167 | 500 |
| 4 | Zürich | ZH | 1'466'424 | 88 | 382 | 4'579 | 500 | 1'145 | | 1'145 | 167 | 500 |
| 5 | Ostschweiz | AI, AR, GL, GR, SG, SH, TG | 1'153'485 | 69 | 300 | 3'602 | 500 | 900 | | 900 | 167 | 500 |
| 6 | Zentral-schweiz | LU, NW, OW, SZ, UR, ZG | 790'458 | 47 | 206 | 2'468 | 500 | 617 | | 617 | 167 | 500 |
| 7 | Tessin | TI | 351'946 | 21 | 92 | 1'099 | 500 | 275 | 225 | 500 | 167 | 500 |
| | Switzer-land/ Total | | 8'327'126 | **500** | **2'167** | **26'000** | **3'500** | 6'500 | 225 | **6'725** | **1'167** | **3'500** |

## 3.6 Retropolation

We have seen in chapter 2.5 that the EC retropolates quarterly series with a naïve scheme repeating the quarterly value for all months during the quarter. In what follows, we use German data to show that simple alternative schemes work better. Then, we outline a more complicated model-based interpolation scheme that could be applied to Swiss data. Further, we perform a correlation analysis to identify potential indicators to interpolate the Swiss consumer confidence index. Finally, we use these indicators to provide a preliminary monthly retropolation of the aggregate Swiss consumer confidence index for illustration.

To examine various methods to disaggregate quarterly to monthly data we provide a simulation exercise with German data since 1985. We construct artificial quarterly data for all indexes underlying the harmonized aggregate index. We assume that the surveys take place in January, April, July and October, as in Switzerland. Then, we retropoloate these series using five approaches. The first repeats the available value for the entire quarter (EC methodology). The second assumes that the missing monthly values evolve linearly between the observed ones. The third uses a cubic spline. Fourth, we use a Chow-Lin interpolation model with business tendency surveys as indicator variables.[28] Fifth, we use the approach that Huwiler and Kaufmann (2013) used to retropolate quarterly and annual CPI indices to monthly frequency, that is, a univariate unobserved-components autoregressive model. We evaluate the methods by comparing an aggregate of the retropolated series to the actual monthly harmonized consumer sentiment for Germany.

*Table 9 - Accuracy of various retropolation methods for Germany*

| Method | Bias | RMSE | MAE | Share correct change | ACF(1) | ACF(2) |
|---|---|---|---|---|---|---|
| Repeated values | -0.02 | 1.74 | 1.04 | 0.28 | 0.33 | 0.01 |
| Linear | -0.04 | 1.21 | 0.73 | 0.69 | 0.31 | 0.00 |
| Spline | -0.06 | 1.33 | 0.80 | 0.66 | 0.34 | -0.03 |
| Chow-Lin | -0.05 | 1.20 | 0.76 | 0.64 | 0.27 | 0.02 |
| Huwiler-Kaufmann | -0.04 | 1.21 | 0.74 | 0.68 | 0.30 | -0.01 |

Notes: Statistics based on the retropolation error as the difference of the actual and retropolated monthly aggregate index over the period 1985-2020. Bias: average error; RMSE: square-root of the average squared error; MAE: average absoulte error; Share correct: share of periods where the retropolation gives the correct sign of the monthly change; ACF(s): autocorrelation of the retropolation error s periods in the past.

We can use the same statistics as in section 2.4 to evaluate the retropolation error (see Table 8).[29] The repeated values approach performs worse than all other retropolation schemes. The corresponding RMSE amounts to 1.7, whereas the linear interpolation, Huwiler-Kaufmann, and Chow-Lin yield 1.2. The cubic spline lies in between. Thus, the RMSE is almost 30% lower when using a simple retropolation method compared to the EC methodology. A similar picture emerges when looking at the MAE. We can also examine in how many cases the retropolated series match the sign of the change of the actual data. Not surprisingly, perhaps, this is the case only in 30% for the repeated values scheme and more than 60% for the other methods.

The upshot of this simple analysis is that we can easily improve upon the repeated values retropolation scheme of the EC. In all cases, however, the retropolated series is quite different than the original one because it is smoother.[30] Moreover, the retropolation error displays significant autocorrelation. This is typically a sign that we can improve upon the scheme by including additional information or using a

---

[28] Note that we chose indicator variables that are significantly related to the consumer sentiment. But, most likely one would be able to find better indicators or lead-lag specifications.

[29] Graphs, as well as ACF with confidence intervals and histograms for the retropolation errors are available in Figures A.2 and A.3 in the Appendix. To preserve space, we only show results for the repeated values and linear method. The upshot is that the retropolation errors match the normal distribution more closely than the repeated values method. Graphs for the other methods can be computed with the R codes we provide.

[30] Note that a lower variance is indeed a property of an optimal forecast, however.

better model. In what follows, we therefore outline a multivariate approach and its advantages and disadvantages.

## A multivariate retropolation model

In recent years, progress has been made to estimate multivariate models with Bayesian methods. Bayesian methods allow us to use prior information in order to reduce estimation uncertainty. This is key for forecasting with multivariate models with many parameters. Otherwise, the additional uncertainty from estimating more parameters can easily offset the gains from additional information.

Schorfheide and Song (2015) provide a Bayesian framework to estimate vector-autoregressive models (VARs) with variables measured at monthly and quarterly frequency. They show that this model performs well in a pseudo-out-of-sample forecasting exercise.[31] Hauzenberger et al. (2021) use this model to retropolate annual CPI inflation to quarterly frequency for 19th century Switzerland. Their model assumes that year-on-year inflation is the average of (unobserved) quarterly year-on-year inflation rates. The key advantage is that the low frequency series depends on its lags – capturing potential persistence – as well as on the lags of other, higher frequency variables – potentially improving the prediction for the unobserved monthly series. Because the model is set up in state-space form, one can then compute the best estimate of the unobserved high frequency series using all data. The framework provides uncertainty intervals for the unobserved monthly predictions to measure its accuracy. Hauzenberger et al. (2021) then use Norwegian data, for which monthly data is available, to show that this approach yields a more than 30% reduction in the RMSE compared to a univariate approach.

## Potential indicators for Switzerland

Chow-Lin and more advanced multivariate approaches require monthly data that are correlated with unobserved monthly consumer sentiment. We therefore provide a correlation analysis for a broad set of monthly indicators.[32] In total we analyze five data categories (aggregated indexes, hard and soft data, financial market data and foreign data), covering various aspects of the economy (e.g. labor market, financial markets, production, trade and business surveys).[33] We also include a text indicator which is calculated on the basis of a large database of newspaper articles using the so-called keyword-in-context (KWIC) method.[34] The biggest advantage of text indicators is that they can be calculated retrospectively and therefore potentially be tuned to be highly correlated with the aggregate or individual consumer sentiment indexes. In total, we obtained more than 380 series.

Since only a few series exist over the entire sample, we perform the analysis on three different samples (starting in 1972, 1985, and 1995). The end of all three samples is December 2020. We use seasonally adjusted data if there is a seasonality. If no official seasonally adjusted series is available, we conduct our own seasonal adjustment.[35]

Because we perform the analysis over a large number of series, we use an automatic procedure to identify potential indicators with high correlation coefficients. In a first step, we remove quarterly data

---

[31] A possible alternative would be to use a mixed-frequency factor model along the lines of Mariano and Murasawa (2003) and Banbura et al. (2013). This would reduce the number of parameters to be estimated by assuming that the underlying monthly series are driven by one or more common factors.

[32] We thank Jérôme Hirschi and SECO for allowing us to collect data from their database.

[33] See Table A.3 in the appendix for a description of selected indicators. Additionally, we provide an Excel file with meta information on all the variables used in the analysis.

[34] First we clean the newspaper corpus from useless information. Second, for each article we count the number of positive and negative words appearing 6 words around the set of keywords: industrie, rezession, konsum, lohn, löhne, einkomm*, kaufkraft wirtschaft*, konjunktur, volkswirt* and invest*. See Burri and Kaufmann (2020) for more details about text cleaning and the sentiment dictionary used. For every appearing keyword we calculate a sentiment score as the number of positive minus the number of negative words divided by the total number of words. Then we normalize the sentiment score on the publication level to have a mean of 0 and a variance of 1. The text indicator is then calculated as the monthly mean of all sentiment scores.

[35] For this, we use the default seas() command of the seasonal package (see Sax and Eddelbuettel, 2018).

interpolated with a repeated values scheme. In a second step, we aggregate the series to quarterly frequency by taking the mean over three months. In a third step we conduct a unit root test (at a 5% confidence level).[36] If the series is stationary, no transformation is applied. Otherwise, we include year-on-year and quarter-on-quarter growth rates of the series. If a variable is measured in percent, we compute differences instead of growth rates. Fourth, we remove autocorrelation from the data to avoid spurious cross-correlations (pre-whitening; see chapter 12.1, Neusser, 2016).[37] Finally, we use the pre-whitened data for the indicators and the aggregate consumer sentiment index and calculate the coincident correlation coefficients as well as cross-correlations for four lags and leads.

*Figure 13 - Cross-correlation with selected indicators 1972-2020*
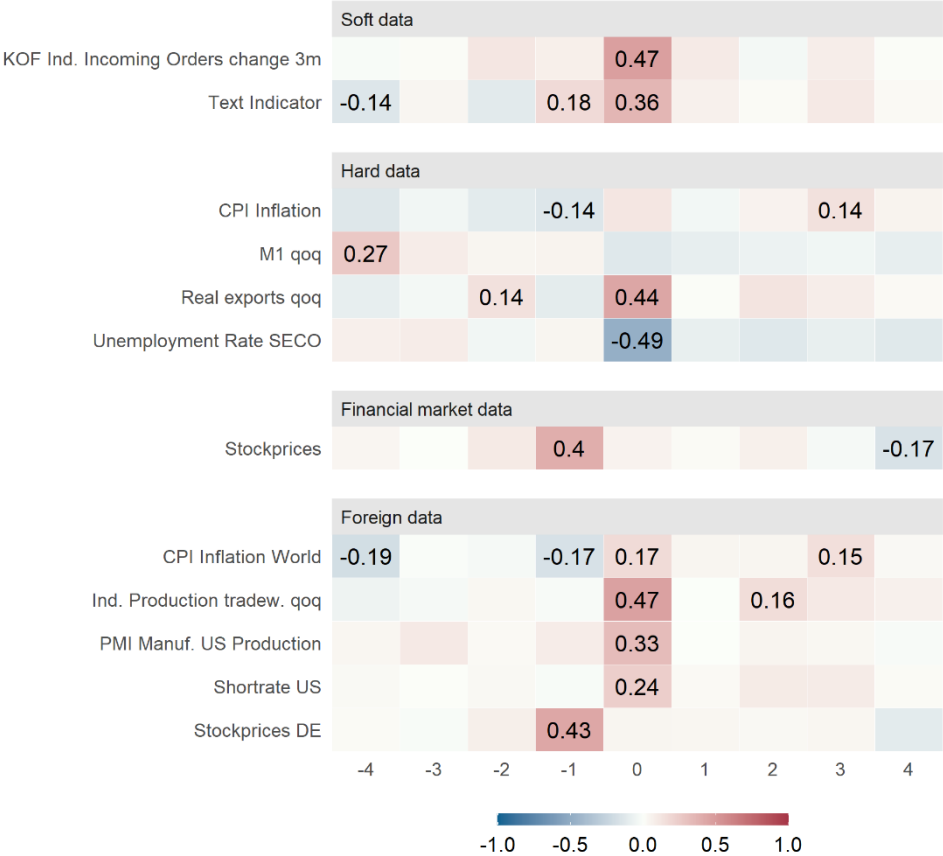


Notes: Cross-correlation between the Swiss consumer sentiment index and a selection of relevant monthly indicators. If the correlation-coefficient is labelled, it's statistically significant different from zero at the 5% level. Quarterly leads/lags are given on the x-axis. Before computing the cross-correlation, the series have been aggregated to quarterly frequency and pre-whitened with an AR(p) model (see Ch. 12.1, Neusser, 2016). The lag order has been determined using the Bayesian information criterion.

Figure 13 shows a selection of relevant indicators for four categories.[38] Because many of the indicators are highly correlated with each other (for example various questions in the KOF survey), we refrain from showing all results. Instead, we selected variables that show the potential of the four categories.

We find the highest coincident correlations (in absolute value) with the SECO unemployment rate, trade weighted industrial production (comprising the EU and US) and the KOF index for expected change of incoming orders in the manufacturing sector. Moreover, we find high correlations with real

---

[36] We use an Augmented Dickey Fuller (ADF) test (see Dickey and Fuller, 1979).

[37] The results without pre-whitening can be computed with the R codes we provide.

[38] The cross-correlations with all series can be computed with the R codes we provide.

exports, stock prices (with a lead of one quarter), the US PMI in manufacturing production and the text indicator. The maximum correlation amounts to almost 0.5. We also find significant correlations that are difficult to reconcile (for example, the wrong sign and/or at very long leads and lags). This was to be expected, however, because we perform a large number of tests.

We find high correlations for similar indicators when starting in 1985 or 1995, although the corresponding samples include more potential indicators (see Figures A.4 and A.5 in the Appendix). In what follows, we describe the most important differences. Not surprisingly perhaps, consumer sentiment in Germany, which starts only in 1985, shows a high correlation. Starting in 1995, new car registrations, arrivals in the tourism sector and SECO's Swiss Economic Confidence (SECO-SEC) are relevant. Recall, however, that SECO-SEC, similar as other aggregate business cycle indicators, includes the other indicators we investigated, as well as the components of the Swiss consumer sentiment itself. More interesting, perhaps, is that two measures of uncertainty are negatively correlated with Swiss consumer confidence: the width of the confidence interval of the SECO-SEC and stock market volatility. This is in line with recent research finding a relevant impact of uncertainty on macroeconomic fluctuations (see Baker, Bloom, and Davis, 2016).

Note that the correlations increase on the shorter sample period (for example from 0.47 to 0.61 for incoming orders). One possible reason is that the consumer sentiment survey as well as the indicator variables are more precisely estimated later in the sample, reducing a downward bias in the correlation due to measurement error. Another possibility is that the actual relationship between indicators and consumer sentiment has changed.

The correlations may be affected by the Covid-19 crisis. As a robustness check, we therefore perform the correlation analysis with a sample that excludes the year 2020 (see Figure A.6 in the Appendix).[39] The correlations are somewhat lower overall. For some variables the correlation drops only slightly (stock prices from 0.4 to 0.38; text indicator from 0.36 to 0.29). For others it drops relatively strongly. For example, the correlation for trade-weighted industrial production falls from 0.47 to 0.17 and for real export growth from 0.44 to 0.29. Nevertheless, the overall picture does not change much. The variables identified above remain those with the highest potential, with the exception of trade-weighted industrial production. A similar pattern emerges when we start the analysis in 1985 or 1995. However, the correlations for the confidence interval for the SECO-SEC, new car registrations and arrivals in the tourism sector switch from highly significant to insignificantly different from zero.


*Monthly consumer sentiment for Switzerland: an illustration*

We investigated which variables end up being statistically significant in a Chow-Lin regression model (see Table 9). We started with a large model including the preferred variables from the correlation analysis over the entire sample. Then, we subsequently removed regressors which are not statistically significant. Note that we included all variables concurrently, except for stock prices which we include with a lag of three months.

Variables from all categories (except financial market data) end up in the model. However, the R2 is still relatively low (0.35). The coefficients have the expected sign (positive for pro-cyclical indicators, negative for counter-cyclical indicators). In addition, the newly constructed text indicator ends up in the model, suggesting that it comprises additional information compared to existing data. Surprisingly, perhaps, the coefficient on world CPI inflation is negative. This may indicate that price rises abroad reduce consumer sentiment because price rises reduce the incentive to purchase large durable goods.

---

[39] The results for the sample starting in 1985 and 1995 can be computed with the R codes we provide.

*Table 10: Significant variables Chow-Lin disaggregation model 1972-2020*

| Variable | Estimate | Standard error |
|---|---|---|
| Constant | 10.07 | 5.21 |
| KOF manufacturing | 0.23 | 0.05 |
| Text | 16.96 | 5.78 |
| Real exports | 0.46 | 0.14 |
| Unemployment | -4.33 | 1.36 |
| CPI world | -1.81 | 0.60 |
| Autoregressive parameter | 0.93 | |
| R2 | 0.35 | |
| Adj. R2: | 0.34 | |

Notes: Chow-Lin regression model based on a selection of variables. Non-significant regressors have been iteratively eliminated in a general-to-specific method

The first panel of Figure 14 shows the resulting disaggregated series jointly with the actual quarterly data for illustration. The second panel compares the interpolation with and without the indicators. The series with indicators is more volatile. Whether this additional volatility comprises additional information or stems from measurement and estimation error introduced by the model is hard to judge, however, because we do not observe the true underlying monthly data.

*Figure 14  - Retropolated monthly Swiss consumer sentiment index*

**Swiss consumer sentiment index**
Chow-Lin retropolation and actual quarterly data (in grey)
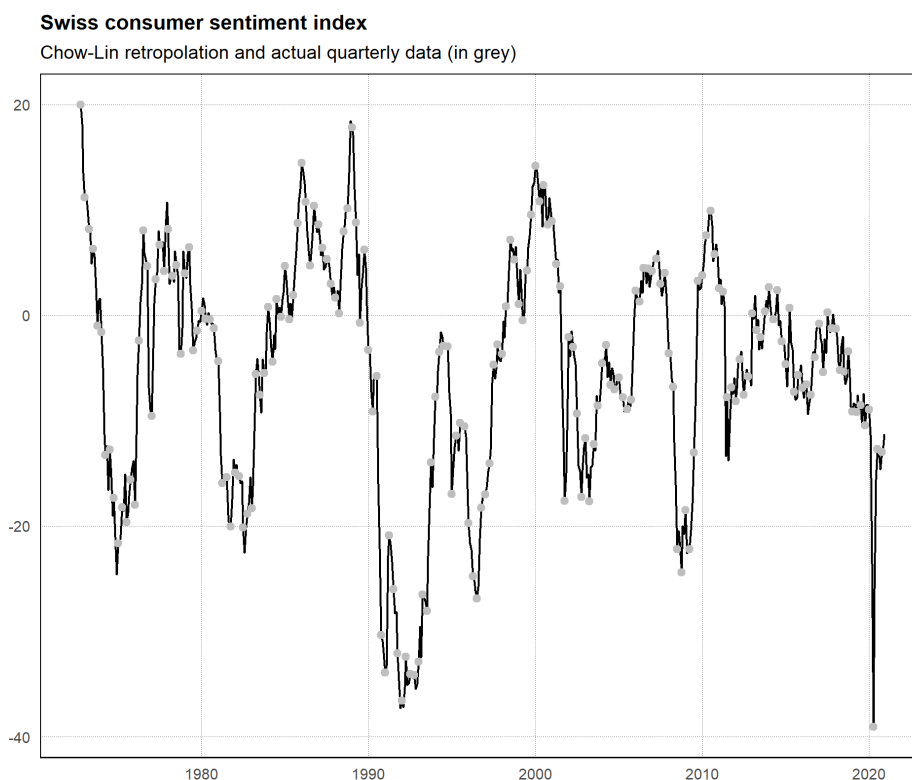
**Swiss consumer sentiment index**

Chow-Lin retropolation with and without indicators (in grey)
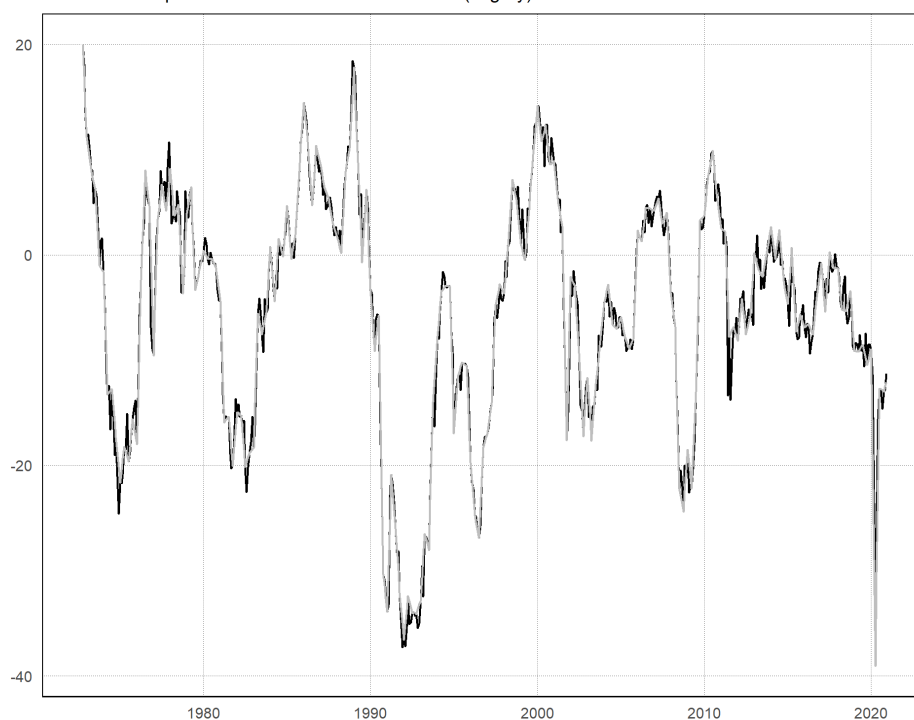


## Recommendations

If the SECO will switch to a monthly survey in the future, one may envisage to retropolate the existing quarterly data to monthly frequency. This would be welcome, as researchers may alternatively use their own methods to retropolate the series or refrain to include the data in monthly models until a sufficiently long series becomes available. In what follows, we provide some recommendations on how to proceed against the backdrop of the previous analysis.

- Simple retropolation schemes work better than the repeated values scheme used by the EC. Therefore, the repeated values scheme should be avoided.

- Retropolation may be seen as a form of forecasting. Although one can add more information by using indicator variables, this comes at the cost of higher estimation uncertainty. Therefore, additional evidence should be gathered whether Chow-Lin with indicator variables can indeed beat, for example, a linear retropolation.

- Along the same lines, one may ask whether the data should be retropolated at the individual question level and then aggregated to an overall monthly index. Doing a retropolation at the individual question level has the advantage of providing researchers with a consistent set of monthly series, including indexes not included in the aggregate. In addition, cross-sectional aggregation of the predictions may cancel some of the individual retropolation errors. But this comes at the cost of having to retropolate more series and estimate more, potentially misspecfied, retropolation models.

- Against the recent successes of multivariate forecasting methods estimated using Bayesian methods, one could consider using such methods for retropolation. This may allow to incorporate more information and retropolating all individual series of the consumer sentiment survey in the same framework. In addition, the models allow to limit estimation uncertainty and incorporate unbalanced panels of indicator variables. The latter is relevant because more and more well-measured indicators become available over time.

- The correlation analysis identifies unemployment data, KOF surveys in manufacturing, trade weighted foreign industrial production, newspaper-based text sentiment, and export data as the variables with the largest potential. With the exception of stock prices, financial market variables

do not work well on the entire sample. On the shorter samples, measures of uncertainty are the most important additional indicator type.

## 3.7 Design weights, non-response adjustment and post-stratification

When sample is selected with the aim of drawing conclusion about unknown mechanism in the whole population, weights are often needed.

Weights are used in several different cases. Typically, when the sampling design is complex, the inclusion probabilities are different, weights are applied for compensate these differences. They are also used to adjust for non-response or for post-stratification. In general, weights are intended to reconstitute the entire population in regard of the variables of interest.

If there is need to oversample certain person-groups or regions (so for example to have an indicator of the NUTS 2 regions/Grossregionen) sampling weights are definitively needed. The sampling weights are simply the inverse of the inclusion probabilities. This will need to be calculated based on the population distribution of the target population (in this case, the resident population 15 years and older) and it needs be updated regularly with changes in the population. As these latter entirely determined by the sampling design and can be delivered by the FSO for each region or unit, the sampling weights are easily calculable.

To evaluate the need of adjustment for non-response it is important to take decisions concerning interests and assumptions.

- If there is concern about representativeness of (measurable) variables, non-response adjustment is needed.
- If the interest is in a change of the index and not in its level, assumptions can be made on the non-response mechanism:
    - If we believe that the non-response mechanism is the same across quarters/region/other socio-demographic characteristics such that by calculating the difference, a systematic effect is eliminated, there is probably no need for adjustment for non-response. This assumption is rather strong and even if it holds true for the data from previous waves there is no reason to not to doubt its long-term validity.
    - If we suppose that non-response depends on variables, adjustment should be done for it. If the variables on which the non-response depends are available at the sampling frame level (thus can be known for the whole selected sample) the adjustment should be done applying these variables. If we suppose that the non-response mechanism depends rather on variables like social participation, political interest, etc., (unknown in the selected sample) and less on social-demographic variables (Lipps and Pekari 2021) decision should be taken if other variables (known at sampling frame level) correlated with these ones are predisposed to be used for simulated the non-response mechanism.

Summing up the use of weights to adjust for non-response:

- First, decision should be taken about which variables generate the mechanism.
- Second, the availability of these variables at sampling frame level should be checked.
- Third, if these variables are not available at the sampling frame level, maybe there are others that are correlated with them and could be used to simulate the mechanism and correct for the non-response.
- Finally, decision should also be taken about if it is worth to have a more precise estimate using weight for non-response correction, knowing that weights are inflate the variance of estimates such as the mean values and therefore require larger sample size.
- As the main questionnaire is relatively short and its proposed frequency is really high, specific non-response survey is not necessary/recommended.

The third typical situation when weights are used it is for post-stratification or calibration. Unlike stratification, post-stratification relies on data obtained in the survey itself that were not available before

the sample selection. The weights are adjusted so that the totals in each group are equal to the known population totals. The values of these variables need not necessary to be known for each unit of the whole population, only the totals of mutually exclusive groups of the whole population.

## *Recommendations*

Sampling-weights: Currently the inclusion probability of the Italian speaking population is higher. If this aspect of the sampling design stays the same we should account it and apply sampling weights.

Adjustment for non-response: Based on the analyse on data from 2017 the non-response mechanism seems to be influenced by the variable "nationality". Consequently, non-response adjustment on variable "nationality" is recommended.

Post-stratification or calibration: The decision about applying or not post-stratification or calibration is rather arbitrary. If a variable is considered important (in a theoretical point of view, for example based on other publication) and available at sampling frame level, post-stratification or calibration can be applied. If there is no particular variable or it is already used during previous steps, the post-stratification is not necessary.

# 4 Annex

## 4.1 Additional tables and figures

### Figure A.1: Additional results on preliminary releases Michigan Survey

**Michigan consumer confidence**

Preliminary estimate and 95% confidence interval

Table A.1: Date of change from quarterly to monthly frequency for participants in the EC har-monized consumer survey program

| Country | Financial situation past (Q1) | Financial situation future (Q2) | Economic situation future (Q4) | Major purchases future (Q9) |
|---------|-------------------------------|---------------------------------|--------------------------------|------------------------------|
| BE | Oct-1985 | Oct-1985 | Oct-1985 | Oct-1985 |
| CZ | Jan-1999 | Jan-1999 | Jan-1999 | Jan-2001 |
| DK | Apr-1986 | Apr-1986 | Apr-1986 | Apr-1986 |
| EE | Apr-2001 | Apr-2001 | Apr-2001 | Apr-2001 |
| IE | Jun-1986 | Jun-1986 | Jun-1986 | Jun-1986 |
| EL | Jan-1986 | Jan-1986 | Jan-1986 | Jan-1986 |
| FR | Apr-1986 | Apr-1986 | Apr-1986 | Apr-1986 |
| LV | Jan-2001 | Jan-2001 | Jan-2001 | Jan-2001 |
| NL | Apr-1986 | Apr-1986 | Apr-1986 | Apr-1986 |
| SK | Apr-2000 | Apr-2000 | Apr-2000 | Apr-2000 |
| FI | Oct-1995 | Oct-1995 | Oct-1995 | Oct-1995 |

Figure A.2: Additional results repeated values retropolation for German data

**German consumer confidence**

Repeated retro. and actual monthly data (in grey)
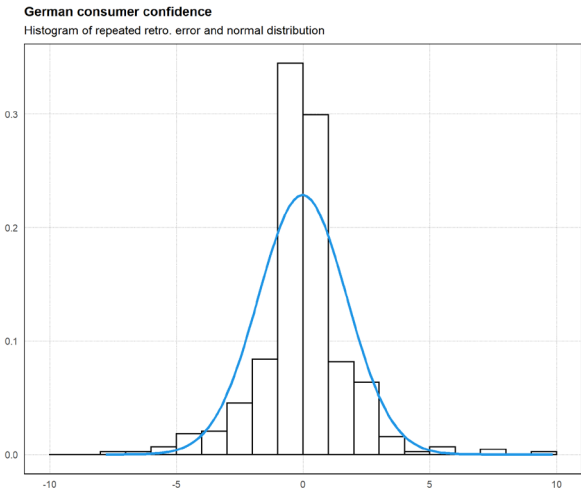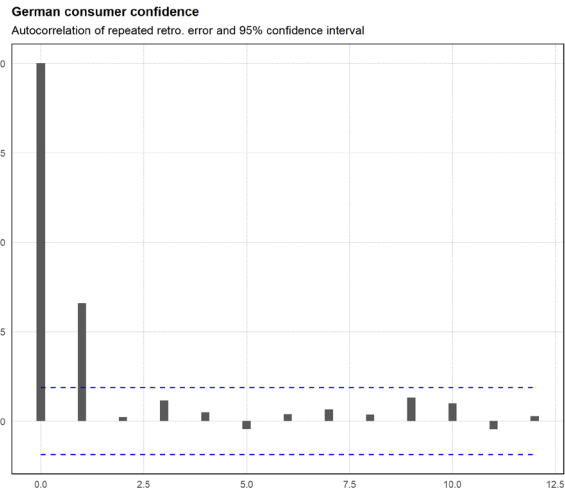
Figure A.3: Additional results linear retropolation for German data

**German consumer confidence**

Linear retro. and actual monthly data (in grey)



German consumer confidence
Autocorrelation of linear retro. error and 95% confidence interval
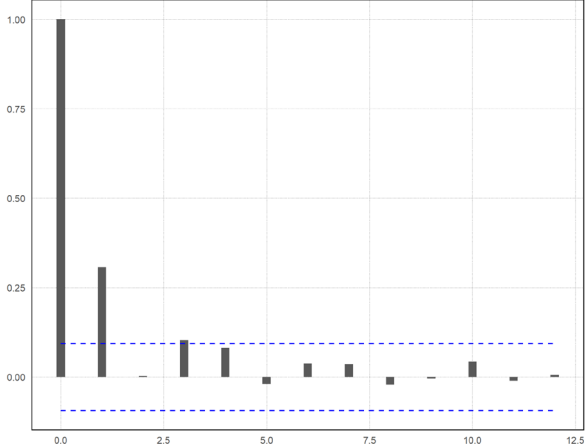
German consumer confidence
Histogram of linear retro. error and normal distribution
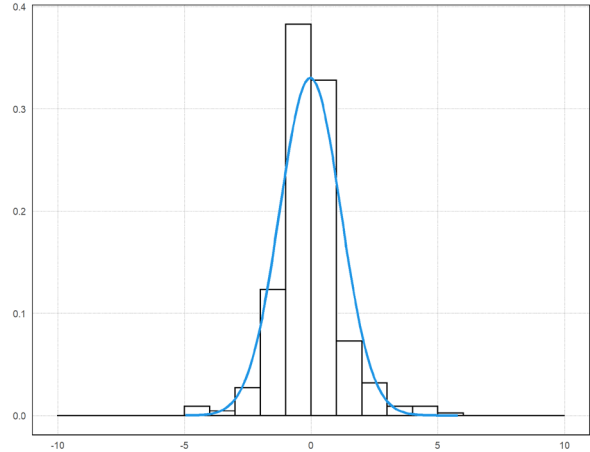
## Table A.2: Comparison with Curtin (2002)

| Questions | Bias 1991-2002 | Bias 2000-2020 | MAE 1991-2002 | MAE 2000-2020 | Correlation 1991-2002 | Correlation 2000-2020 |
|---|---|---|---|---|---|---|
| Consumer sentiment | 0.00 | 0.40 | 1.00 | 1.20 | 0.99 | 0.99 |
| Personal financial situation | 0.60 | 0.80 | 2.50 | 2.40 | 0.97 | 0.99 |

Notes: Statistics based on the revision computed as the difference of the final and preliminary release. Bias: average revision; MAE: average absoulte revision; Correlation between preliminary and final value.

*Figure A.4: Cross-correlation with selected indicators 1985-2020*
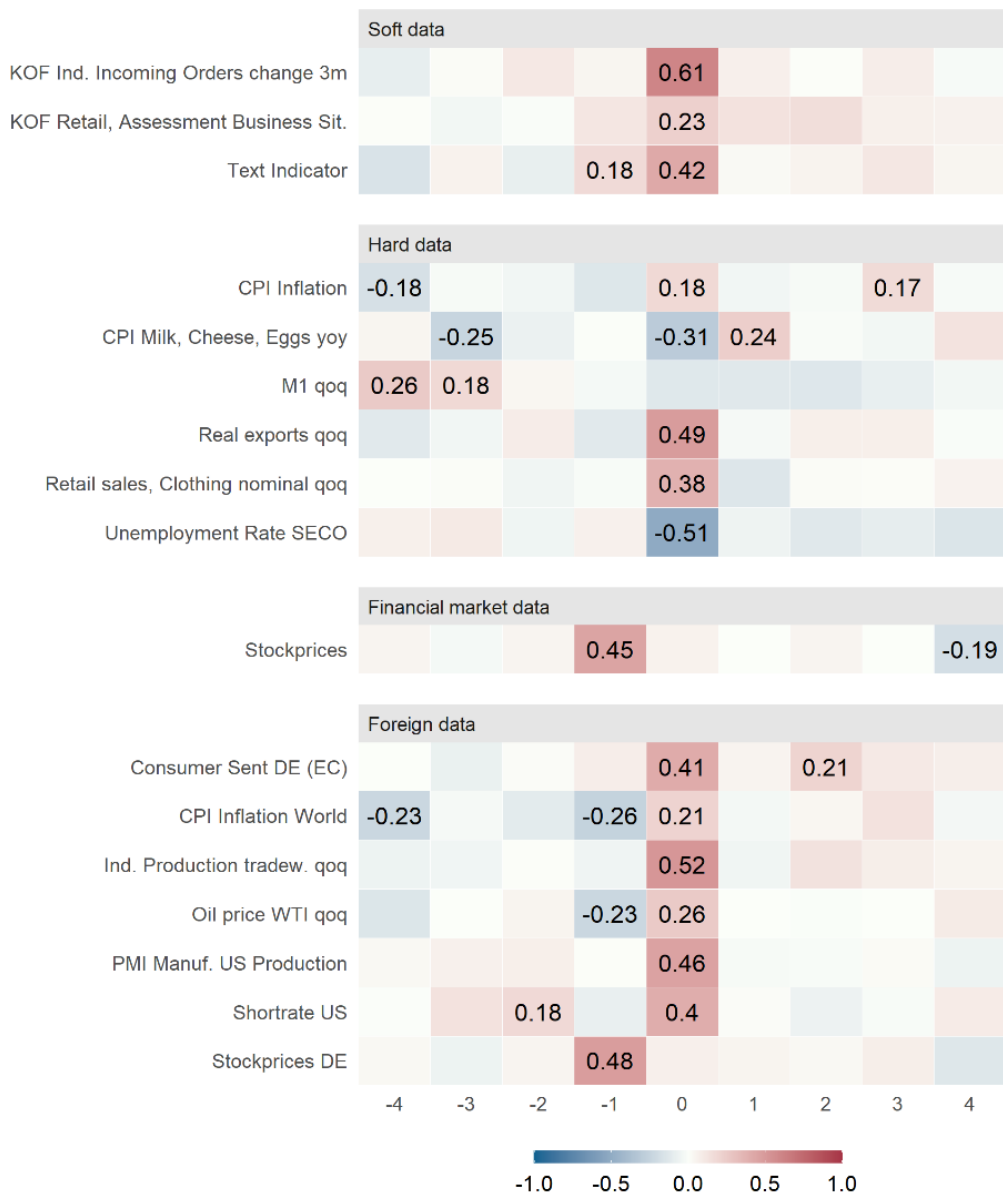


Notes: Cross-correlation between the Swiss consumer sentiment index and a selection of relevant monthly indicators. If the correlation-coefficient is labelled, it's statistically significant different from zero at the 5% level. Quarterly leads/lags are given on the x-axis. Before computing the cross-correlation, the series have been aggregated to quarterly frequency and pre-whitened with an AR(p) model (see Ch. 12.1, Neusser, 2016). The lag order has been determined using the Bayesian information criterion.

*Figure A.5: Cross-correlation with selected indicators 1995-2020*
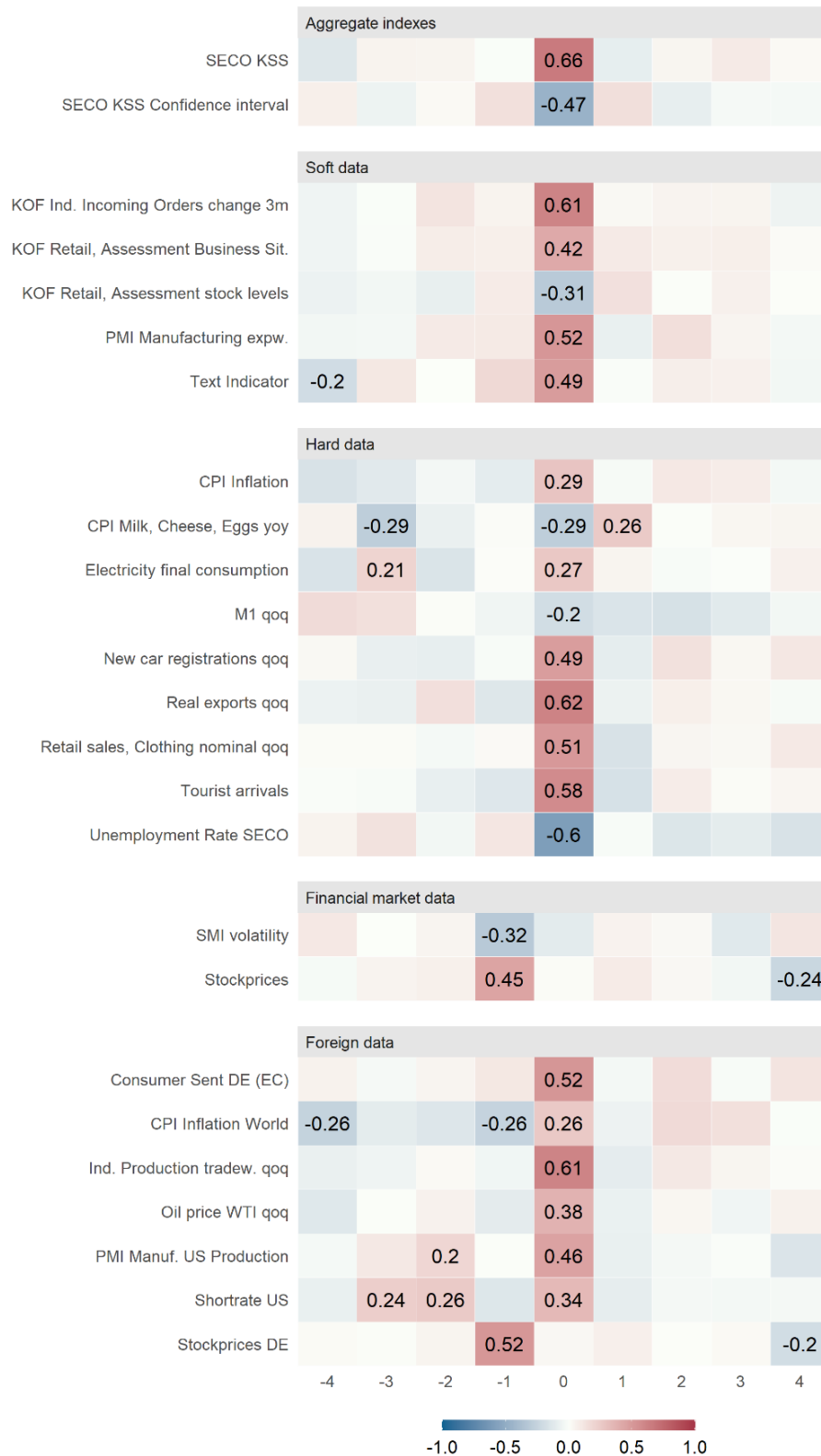
| | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|
| **Aggregate indexes** | | | | | | | | | |
| SECO KSS | | | | | 0.66 | | | | |
| SECO KSS Confidence interval | | | | | -0.47 | | | | |
| **Soft data** | | | | | | | | | |
| KOF Ind. Incoming Orders change 3m | | | | | 0.61 | | | | |
| KOF Retail, Assessment Business Sit. | | | | | 0.42 | | | | |
| KOF Retail, Assessment stock levels | | | | | -0.31 | | | | |
| PMI Manufacturing expw. | | | | | 0.52 | | | | |
| Text Indicator | -0.2 | | | | 0.49 | | | | |
| **Hard data** | | | | | | | | | |
| CPI Inflation | | | | | 0.29 | | | | |
| CPI Milk, Cheese, Eggs yoy | | -0.29 | | | -0.29 | 0.26 | | | |
| Electricity final consumption | | 0.21 | | | 0.27 | | | | |
| M1 qoq | | | | | -0.2 | | | | |
| New car registrations qoq | | | | | 0.49 | | | | |
| Real exports qoq | | | | | 0.62 | | | | |
| Retail sales, Clothing nominal qoq | | | | | 0.51 | | | | |
| Tourist arrivals | | | | | 0.58 | | | | |
| Unemployment Rate SECO | | | | | -0.6 | | | | |
| **Financial market data** | | | | | | | | | |
| SMI volatility | | | | -0.32 | | | | | |
| Stockprices | | | | 0.45 | | | | | -0.24 |
| **Foreign data** | | | | | | | | | |
| Consumer Sent DE (EC) | | | | | 0.52 | | | | |
| CPI Inflation World | -0.26 | | | -0.26 | 0.26 | | | | |
| Ind. Production tradew. qoq | | | | | 0.61 | | | | |
| Oil price WTI qoq | | | | | 0.38 | | | | |
| PMI Manuf. US Production | | | 0.2 | | 0.46 | | | | |
| Shortrate US | | 0.24 | 0.26 | | 0.34 | | | | |
| Stockprices DE | | | | 0.52 | | | | | -0.2 |

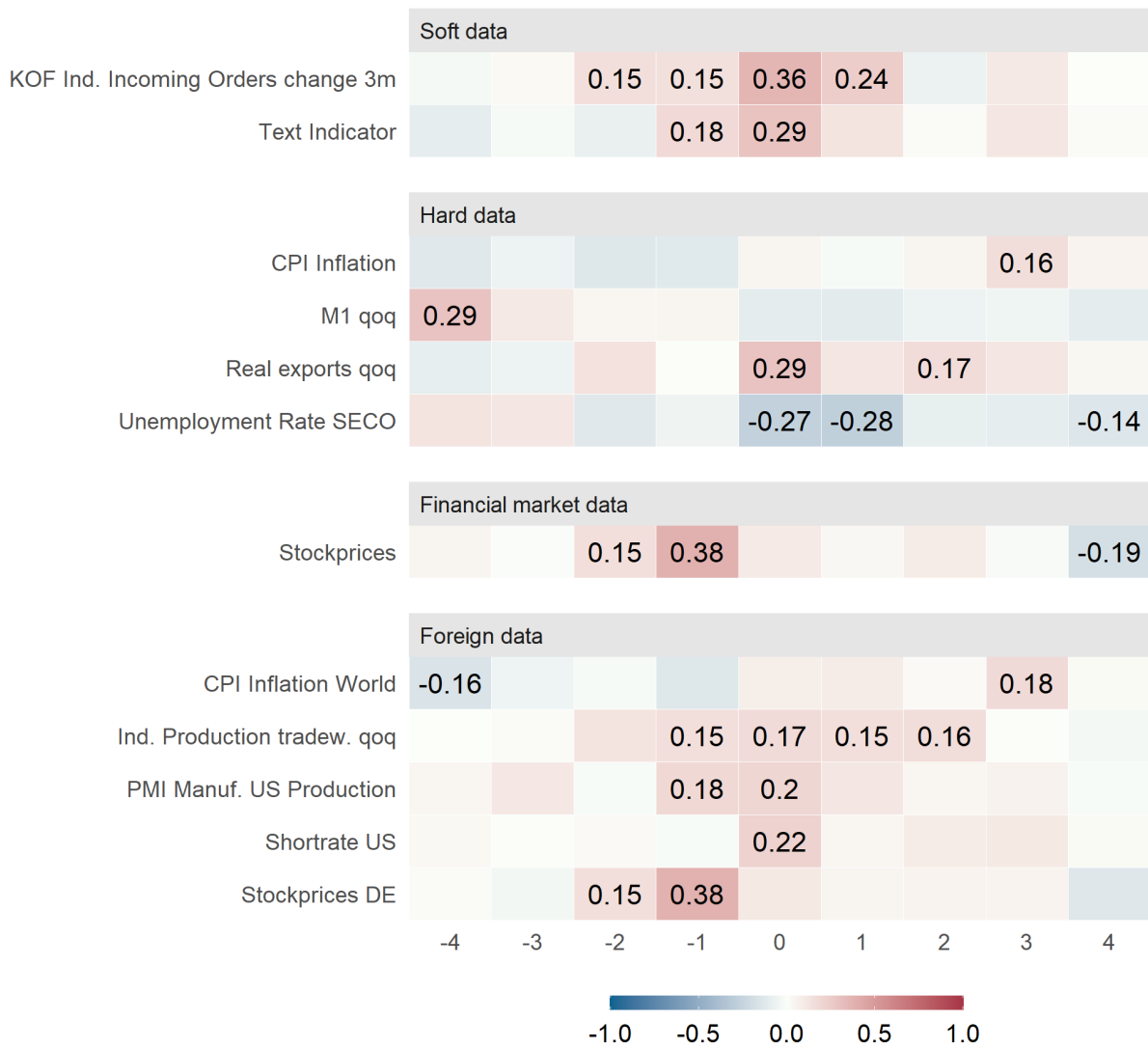Notes: Cross-correlation between the Swiss consumer sentiment index and a selection of relevant monthly indicators. If the correlation-coefficient is labelled, it's statistically significant different from zero at the 5% level. Quarterly leads/lags are given on the x-axis. Before computing the cross-correlation, the series have been aggregated to quarterly frequency and pre-whitened with an AR(p) model (see Ch. 12.1, Neusser, 2016). The lag order has been determined using the Bayesian information criterion.

## Figure A.6: Cross-correlation with selected indicators 1972 - 2019

| | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|
| **Soft data** | | | | | | | | | |
| KOF Ind. Incoming Orders change 3m | | | 0.15 | 0.15 | 0.36 | 0.24 | | | |
| Text Indicator | | | | 0.18 | 0.29 | | | | |
| **Hard data** | | | | | | | | | |
| CPI Inflation | | | | | | | | 0.16 | |
| M1 qoq | 0.29 | | | | | | | | |
| Real exports qoq | | | | | 0.29 | | 0.17 | | |
| Unemployment Rate SECO | | | | | -0.27 | -0.28 | | | -0.14 |
| **Financial market data** | | | | | | | | | |
| Stockprices | | | 0.15 | 0.38 | | | | | -0.19 |
| **Foreign data** | | | | | | | | | |
| CPI Inflation World | -0.16 | | | | | | | 0.18 | |
| Ind. Production tradew. qoq | | | | 0.15 | 0.17 | 0.15 | 0.16 | | |
| PMI Manuf. US Production | | | | 0.18 | 0.2 | | | | |
| Shortrate US | | | | | 0.22 | | | | |
| Stockprices DE | | | 0.15 | 0.38 | | | | | |

Color scale: -1.0  -0.5  0.0  0.5  1.0

Notes: Cross-correlation between the Swiss consumer sentiment index and a selection of relevant monthly indicators. If the correlation-coefficient is labelled, it's statistically significant different from zero at the 5% level. Quarterly leads/lags are given on the x-axis. Before computing the cross-correlation, the series have been aggregated to quarterly frequency and pre-whitened with an AR(p) model (see Ch. 12.1, Neusser, 2016). The lag order has been determined using the Bayesian information criterion.

## Table A.3: Selected series in cross-correlation analysis

| ID | Name | Source | Datatype | Description |
|---|---|---|---|---|
| KSSdev | SECO KSS Confidence interval | SECO | Aggregate indexes | Size of confidence interval SECO-KSS |
| KSSmean | SECO KSS | SECO | Aggregate indexes | SECO-KSS |
| ind_111 | KOF Ind. Incoming Orders change 3m | KOF | BTS Industry | Total, Incoming Orders, Expected Change Next 3 Months, Balance |
| retail_36 | KOF Retail, Assessment Business Sit. | KOF | BTS Retail | Total, Business Situation, Assessment, Balance |
| retail_45 | KOF Retail, Assessment stock levels | KOF | BTS Retail | Total, Stock Levels, Assessment, Balance |
| elstat_enverbr_sa | Electricity final consumption | BFE | Electricity | Final consumption |
| SMI_VOL | SMI volatility | SIX | Financial market | SMI Volatility |
| STOCKP_CH | Stockprices | St. Louis Fred | Financial market | Total Share Prices for All Shares for Switzerland |
| csent1_int_6 | Consumer Sent DE (EC) | European Comission | Foreign Consumer Sentiment | Germany |
| SHORT_US | Shortrate US | St. Louis Fred | Foreign financial market | 3 month interst rate US |
| STOCKP_DE | Stockprices DE | St. Louis Fred | Foreign financial market | Total Share Prices for All Shares for the Germany |
| pmi_int_6 | PMI Manuf. US Production | ISM | Foreign PMI | United States, Business Surveys, Report on Business, Manufacturing, PMI, Production, SA, Index |
| pmi_man_expw | PMI Manufacturing expw. | SECO | Foreign PMI | Export weighted PMI |
| text | Text Indicator | UNINE,Tamedia, e-newspaperarchives.ch | Text Indicator | Newspaper sentiment calculated using KWIC method |
| touris_1_sa | Tourist arrivals | BFS | Tourism | Arrivals in tourism sector, Total |
| unemp_144_sa | Unemployment Rate SECO | SECO | Unemployment | Unemployment, Rate, SA |
| cpi_35_sa | CPI Milk, Cheese, Eggs yoy | BFS | Prices | Inflation for Milk, Cheese, Eggs |
| cpi_1_sa | CPI inflation | BFS | Prices | |
| M1 | M1 | SNB | Monetary Aggregates | |
| cars_new_pass_sa | New car registrations | ACEA | Car Registrations | Number of new Vehicle Registrations |
| exp_real_sa | Real exports | SFCA | Export/Import | |
| dhu_kleid_n_sa | Retail sales clothing | BFS | Retail sales | |
| CPI_W | CPI inflation world | UNINE | Prices | |
| INPRTW | Industrial production tradeweighted | St. Louis Fred | Global economic activity | Trade weighted EU and US |
| OIL_WTI | Oil price WTI | St. Louis Fred | Foreign financial market | |

Notes: Metadata to all tested series will be provided in a separate excel file.

## 4.2 The project team

**Prof. Dr. Georg Lutz** ist seit 2016 Direktor von FORS. Er ist zudem Professor an der Universität Lausanne und zwischen 2008 und 2016 war er Leiter der Schweizer Wahlstudie Selects. In seiner Forschung befasst er sich mit dem politischen Verhalten, insbesondere dem Wahl- und Abstimmungsverhalten, der Wahlbeteiligung, den politischen Institutionen, dem Wahlsystem, der Schweizer Politik sowie mit Methoden der empirischen Sozialforschung. Georg Lutz unterrichtet zudem politisches Verhalten und Umfrageforschung an der Universität Lausanne.

**Prof. Dr. Boris Wernli** ist seit 2008 Leiter der Abteilung *Befragungen* und Mitglied der Direktion von FORS. Zudem unterrichtet er seit 2009 Umfrageforschung an der Universität Lausanne. Boris Wernli verfügt über 25 Jahre Erfahrung in der Umfrageforschung in verschiedenen Feldern der Sozialwissenschaften mit einem Fokus auf die längsschnittliche Datenanalyse. Er ist zudem verantwortlich für Umfrageprojekte in Zusammenarbeit mit dem Institut für Sozialwissenschaften der Universität Lausanne und arbeitet eng mit dem Nationalen Forschungsschwerpunkt LIVES zusammen.

**Prof. Dr. Daniel Kaufmann** ist Assistenzprofessor in angewandter Makroökonomie an der Universität Neuenburg und Research Fellow an der KOF Konjunkturforschungsstelle der ETH Zürich. In der Vergangenheit untersuchte er, ob Firmenumfragen nützlich sind, um makroökonomische Grössen vorherzusagen und entwickelte einen täglichen Konjunkturindikator für die Schweizer Wirtschaft während der Covid-19 Krise. In seiner aktuellen Forschung untersucht er die Auswirkungen von nominellen Rigiditäten auf die reale Wirtschaft, die Zinsentwicklung in der Schweiz seit dem 19. Jahrhundert, sowie den Informationsgehalt von Texten für die Konjunkturanalyse.

**PD Dr. Oliver Lipps** arbeitete für das Deutsche Mobilitätspanel (MoP), den Survey on Health, Ageing and Retirement in Europe (SHARE) und für das Schweizer Haushaltspanel, bevor er Leiter der methodischen Forschungseinheit FORS in Lausanne wurde. Als Privatdozent am Institut für Soziologie der Universität Bern unterrichtet er Methoden der empirischen Sozialforschung und an der FORS Sommerschule in Lugano Methoden der Paneldatenanalysen. Oliver Lipps hat vor allem Artikel zu Fragen der Umfragedatenqualität veröffentlicht, mit einem Schwerpunkt auf Nicht-Teilnahme und Attrition in Panelbefragungen.

**Dr. Erika Antal** ist verantwortlich für das Gewichtungssystem der Schweizer Haushaltspanel-Erhebung. Die Forschungsinteressen von Erika Antal liegen im Bereich der Stichprobenziehung für Befragungen. Als Verantwortliche für das Gewichtungssystem des SHP baut sie die verschiedenen Gewichte auf. Sie trägt auch zur Entwicklung von Stichprobenstrategien für andere verwandte Erhebungen bei (Lives-Cohort und die Zusatzstichprobe des Kantons Waadt).

**MA Marc Burri** ist Doktorand in angewandter Makroökonomie an der Universität Neuenburg. In seiner Dissertation befasst er sich damit, wie Textdaten zur Untersuchung der Konjunktur und für Kausalitätsanalysen verwendet werden können. Er hat einen täglichen Konjunkturindikator für die Schweizer Wirtschaft während der Covid-19 Krise entwickelt.

**MA Victor Legler** ist seit 2016 wissenschaftlicher Mitarbeiter in der Gruppe Data Collection and Analysis bei FORS und verfügt über einen Master-Abschluss in Public Management and Policy (PMP) vom Hochschulinstitut Für Öffentliche Verwaltung (IDHEAP). Bei FORS arbeitet er hauptsächlich an der Programmierung und Organisation des Datenerhebungsprozesses (Web/Papier) für die MOSAiCH-Studie sowie der Datenaufbereitung.

## 4.3   Literature

Andersson, C., Andersson, K., & Lundquist, P. (2011). Estimation of change in a rotation panel design. Proceeding of the 58th session of International Statistical Institute, Dublin.

Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013): "Now-Casting and the Real-Time Data Flow," in Handbook of Economic Forecasting, ed. by G. Elliot, and A. Timmermann, vol. 2A, chap. 4, pp. 195–237. Elsevier.

Biemer, P. P., & Lyberg, L. E. (2003). Introduction to survey quality (Vol. 335). John Wiley & Sons.

Bierbaumer-Polly, J, A Czaloun, C Glocker, S. Kaniovski (2019), "An evaluation of the European Commission's concept of the consumer confidence index for Switzerland," study commissioned by the Swiss State Secretariat for Economic Affairs (SECO), Austrian Institute of Economic Research (WIFO), https://www.seco.admin.ch/dam/seco/en/dokumente/Wirtschaft/Wirtschaftslage/Konsumentenstimmung/Evaluation.pdf.download.pdf/Evaluation.pdf

Bürkl, Rolf (undated), "GfK Konsumklima - Zur Methode," Presentation, GfK Marketing & Consumer Intelligence

Burri, M., Kaufmann, D. (2020) A daily fever curve for the Swiss economy. Swiss J Economics Statistics 156, 6 . Available at https://doi.org/10.1186/s41937-020-00051-z

Bundesamt für Statistik (2021). Statistique de la population active occupée (SPAO): bases méthodologiques. Neuenburg.

Curtin, R.T. (undated). "Surveys of Consumers." University of Michigan, Survey Research Center. Available at https://www.oecd.org/sdd/leading-indicators/33650864.pdf

Curtin, Richard (2002), "Surveys of consumers: The accuracy of the preliminary estimates", https://data.sca.isr.umich.edu/fetchdoc.php?docid=25918

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a), 427–431.

EC (2018), "A revised consumer confidence indicator," European Commission, https://ec.europa.eu/info/sites/default/files/new_cci.pdf

EC (2021), "The joint harmonised EU programme of business and consumer surveys: User guide," European Commission, https://ec.europa.eu/info/files/user-guide-joint-harmonised-eu-programme-business-and-consumer-surveys_en

EC (2021a), "Flash consumer confidence indicator for EU and Euro Area," Press release 22.09.2021, https://ec.europa.eu/info/sites/default/files/flash_consumer_2021_09_en.pdf

The European Commission 2014 .The Joint Harmonised EU Programme of Business and Consumer Surveys. List of best practice for the conduct of business and consumer surveys (march 2014). Available at https://ec.europa.eu/info/files/user-guide-joint-harmonised-eu-programme-business-and-consumer-surveys_en

Hauzenberger, N, F Huber, D Kaufmann, R Stuart and C Tille, "Interest rates in Switzerland 1852-2020," Grundlagen für die Wirtschaftspolitik Nr. 24, SECO, https://www.seco.admin.ch/seco/de/home/Publikationen_Dienstleistungen/Publikationen_und_Formulare/Strukturwandel_Wachstum/Wachstum/interest_rates_switzerland_1852-2020.html

Huwiler, M and D Kaufmann, 2013. "Combining disaggregate forecasts for inflation: The SNB's ARIMA model," Economic Studies 2013-07, Swiss National Bank.

Istat (2021), "Consumer and business confidence," Istituto Nazionale di Statistica, https://www.istat.it/it/files//2021/09/Business-and-consumer-confidence_09_2021.pdf

Johnston, Richard, and Henry E. Brady. "The rolling cross-section design." Electoral Studies 21.2 (2002): 283-295.

Kaniovski and Glocker, 2019 . AN EVALUATION OF THE EUROPEAN COMMISSION'S CONCEPT OF THE CONSUMER CONFIDENCE INDEX FOR SWITZERLAND. Available at https://www.seco.ad-min.ch/dam/seco/it/dokumente/Wirtschaft/Wirtschaftslage/Konsumentenstimmung/Evalua-tion.pdf.download.pdf/Evaluation.pdf

Kaufmann, D and R Scheufele (2017), "Business tendency surveys and macroeconomic fluctuations," International Journal of Forecasting, 33(4), 878-893.

Kemeny, F., & Bachmann, A. (2019). Konsumentenumfrage liefert frühe Hinweise auf BIP-Entwicklung. Die Volkswirtschaft 11 / 2019.

Lipps, O., & Pekari, N. (2021). Sequentially mixing modes in an election survey. Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=15281 DOI:10.13094/SMIF-2021-00003

Lutz, G., De Rocchi, T., & Pekari, N. (2013). The Swiss 2011 Rolling Cross-Section study: Design, field work, and data quality. FORS Working Papers, (4/2013).

Mariano, R. S. and Y. Murasawa, (2003). A new coincident index of business cycles based on monthly and quarterly series. Journal of applied Econometrics, 18(4):427–443, 2003. URL https://doi.org/10.1002/jae.695.

Marcellino, Massimiliano, (2006), Leading Indicators, ch. 16, p. 879-960 in Handbook of Economic Forecasting, Elliott, G., Granger, C. and Timmermann, A. eds., Elsevier.

Merkle, D., Langer G. and Sussman D. (2004). Consumer Confidence: Measurement and Meaning. . Available at https://www.langerresearch.com/wp-content/uploads/Consumer-Confidence-Measurement-and-Meaning.pdf

Michigan Survey (2021a), "Recent reports," Website, acessed on 21.10.2021, https://data.sca.isr.umich.edu/reports.php

Michigan Survey (2021b), "Time Series Data", Website, accessed on 21.10.2021, https://data.sca.isr.umich.edu/data-archive/mine.php

Michigan Survey (1978), "Surveys of consumer attitudes - Monitoring economic change program", 03.07.1978, https://data.sca.isr.umich.edu/fetchdoc.php?docid=4199

Michigan Survey (undated), "Index calculations," https://data.sca.isr.umich.edu/fetchdoc.php?do-cid=24770

Neusser, K. (2016). *Time Series Econometrics*, (pp. 207–214). Cham: Springer. https://doi.org/10.1007/978-3-319-32862-1_11.

Phillips, G. W., & Jiang, T. (2016). Measurement error and equating error in power analysis. Practical Assessment, Research, and Evaluation, 21(1), 9.

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. Journal of management, 36(1), 94-120.

Rast, P., & Hofer, S. M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: simulation results based on actual longitudinal studies. Psychological Methods, 19(1), 133.

Sax, C., & Eddelbuettel, D. (2018). Seasonal Adjustment by X-13ARIMA-SEATS in R. Journal of Statistical Software, 87(11), 1–17. https://doi.org/10.18637/jss.v087.i11

Schorfheide, F and D Song, (2015), "Real-Time Forecasting With a Mixed-Frequency VAR," Journal of Business & Economic Statistics, 33(3), 366–380

Scott R. Baker, Nicholas Bloom, Steven J. Davis, Measuring Economic Policy Uncertainty, *The Quarterly Journal of Economics*, Volume 131, Issue 4, November 2016, Pages 1593–1636, https://doi.org/10.1093/qje/qjw024

SECO (2020), "Consumer sentiment at record low," Press release 15.04.2020, https://www.seco.ad-min.ch/seco/en/home/seco/nsb-news.msg-id-78784.html

Stock, J H, and M W Watson (2002), "Macroeconomic Forecasting Using Diffusion Indexes," Journal of Business & Economic Statistics, 20(2), 147-162, DOI: 10.1198/073500102317351921

Tomczak, M., Tomczak, E., Kleka, P., & Lew, R. (2014). Using power analysis to estimate appropriate sample size. Trends in Sport Sciences, 21(4), 195-206.

The European Commission 2018. A REVISED CONSUMER CONFIDENCE INDICATOR (21 DECEMBER 2018). Available at https://ec.europa.eu/info/sites/default/files/new_cci.pdf

The European Commission 2021 .The Joint Harmonised EU Programme of Business and Consumer. Surveys User Guide (updated July 2021). Available at https://ec.europa.eu/info/files/user-guide-joint-harmonised-eu-programme-business-and-consumer-surveys_en

Voorpostel, M., & Lipps, O. (2011). Attrition in the Swiss Household Panel: Is change associated with later drop-out? Journal of Official Statistics, 27(2), 301-318.