Journal of Business Cycle Research (2023) 19:241–273
https://doi.org/10.1007/s41549-023-00088-y

RESEARCH PAPER

Imputing Monthly Values for Quarterly Time Series: An Application Performed with Swiss Business Cycle Data

Klaus Abberger[1,3] · **Michael Graff**[1] · **Oliver Müller**[1] · **Boriss Siliverstovs**[1,2]

Received: 22 December 2022 / Accepted: 31 August 2023 / Published online: 23 September 2023

Abstract

This paper compares algorithms to deal with the problem of missing values in higher frequency data. We refer to Swiss business tendency survey data at monthly and quarterly frequency. There is a wide range of imputation algorithms. To evaluate the different approaches, we apply them to series that are de facto monthly, from which we create quarterly data by deleting two out of three data points from each quarter. At the same time, the monthly series are ideal to deliver higher frequency information for multivariate imputation algorithms. With this set of indicators, we conduct imputations of monthly values, resorting to two univariate and four multivariate algorithms. We then run tests of forecasting accuracy by comparing the imputed monthly data with the actual values. Finally, we take a look at the congruence of an imputed monthly series from the quarterly survey question on firms' capacity utilisation with other monthly data reflecting the Swiss business cycle. The results show that an algorithm based on the Chow and Lin approach, amended with a variable pre-selection procedure, delivers the most precise imputations, closely followed by the standard Chow-Lin algorithm and then multiple regression. The cubic spline and the EM algorithm do not prove useful.

Keywords Temporal disaggregation · Imputation · Business tendency surveys · Out-of-sample validation · Mixed-frequency data

JEL Classification C19 · C22 · C53

✉ Michael Graff
graff@kof.ethz.ch

1 ETH Zürich, KOF Swiss Economic Institute, Leonhardstrasse 21, CH-8092 Zurich, Switzerland

2 Latvijas Banka, Riga, Latvia

3 CESifo, Munich, Germany

1 Introduction

Access to timely and reliable information is instrumental for continuous monitoring and for forming sound judgements on the short-term outlook of economic conditions. This is especially true when economic conditions change rapidly, with elevated levels of uncertainty, as the experience gained during the Great Financial Crisis as well as the more recent COVID-19 pandemic has illustrated.

Some economic key indicators, however, are released at quarterly or annual frequencies only, and in some cases with substantial publication delays, which often amount to several months. In such situations, to alleviate the informational constraints imposed by low-frequencies and publication delays, temporal disaggregation may be indicated, for the most current values as well as those referring to the past. To this end, different temporal disaggregation methods, both uni- and multivariate, have been suggested and are applied to convert observed low-frequency time series into higher frequencies. Noteworthy, imputation of lower-frequency series into higher frequencies poses different challenges for missing values between observed data points in the past and the most recent data points (interpolation), which have to be now- or forecasted beyond the last available observation (extrapolation). We refer to the former as "in-sample" and the latter as "out-of-sample".

In this paper we document the performance of a selection of temporal disaggregation techniques, distinguishing in-sample from out-of-sample imputations. Our main data come from the monthly business tendency surveys (BTS) in the Swiss manufacturing industry, conducted by the KOF Swiss Economic Institute. First, we explore which method performs best to impute monthly values within the very same data corpus. To this end we create artificially quarterly series from *de facto* monthly BTS series by removing two out of three quarterly observations, so that we can compare the imputations with the known realisations. Moving from this experimental set-up to a real rather than constructed informational deficit, we turn to the genuinely quarterly series *technical capacity utilisation in per cent*, summarising the responses of firms in Switzerland to the corresponding question in the quarterly KOF BTS in the manufacturing industry. The importance of this indicator lies in the fact that it measures economic slack/pressure at the micro level and—aggregated—helps to assess potential output and hence the output gap at the macro level (Graff & Sturm, 2012).[1]

In particular, we proceed as follows: In the first step, we conduct an experimental simulation exercise based on the eleven monthly BTS question series that can be traced sufficiently long back into the past. Skip-sampling two months of each quarter in these time series provides eleven quarterly time series. To these skip-sampled time series we apply a selection of uni- and multivariate temporal disaggregation algorithms to impute back the removed observations, with the multivariate algorithms referring to the monthly information from the remaining ten time series,

[1] Given the exposure of the small and open Swiss economy to a rapidly changing global economy, this indicator is one of utmost important for the guidance of monetary and fiscal policies. It is in order to remark that whilst our data and their information are relating to Switzerland, the problem is general and applies similarly to the rest of the world.

serving as *auxiliary indicators*. We benchmark the performance of the considered imputation methods against what is in our context the most conservative (naïve) method, namely carrying the last observation forward. The competing temporal disaggregation methods include the univariate cubic spline, and in the multivariate setup the Expectation–Maximisation algorithm, multiple regression, the standard Chow-Lin method, and a modified version of the latter. We refer to this as the *internal* validation step, as we are able to compare the accuracy of imputations across different temporal disaggregation methods because we know the true values of the skip-sampled observations.

In the second step, we apply the best-performing method from the first step to convert a quarterly time series of interest, the KOF survey item *technical capacity utilisation in per cent*, into monthly frequency. Then we take this temporally disaggregated time series and evaluate its congruence with a selection of monthly business cycle indicators other than the KOF BTS that can *a priori* be assumed to be closely related to what is reflected by the technical capacity utilisation BTS question. Since in this exercise the unobserved monthly capacity utilisation values are genuinely unknown, we rely on a forecast encompassing test in order to verify the superiority of the chosen temporal disaggregation method over the benchmark, last observation carried forward. We refer to this as the *external* validation step.

We contribute to the literature in the following ways. First, our experimental simulations with artificially created quarterly time series allows us to compare the performance of different temporal disaggregation techniques along a variety of dimensions going beyond the focus of the related literature. We do not only discriminate (1) *in-sample* interpolations between historically observed data points from *out-of-sample* extrapolations of values, but also (2) the accuracies of imputations of the *first* and the *second* unobserved month within a quarter, (3) the accuracies of *direct* imputations of the missing BTS values of the balance indicators versus *indirect* computation form the imputed positive- and negative shares,[2] and last but not least (3) the accuracy of the directional changes (trends) of the imputed values. In particular, we demonstrate that for our data corpus, temporal disaggregation with an adoption of the procedure suggested by Wang et al. (2007) to the standard Chow-Lin (1971) method, tends to outperform the considered alternatives, although precision is only marginally improved compared to the standard Chow-Lin algorithm. The performance of the Expectation–Maximisation algorithm is revealed as less than satisfying, which is not quite as we expected, given its popularity, whereas multiple regression surprises on the positive side, given that is goes back to the nineteenth century. Last but not least, while the univariate spline may be useful for in-sample-imputations, it clearly fails to outperform the (naïve) carrying forward of the last observation.

The rest of the paper is structured as follows. Section 2 reviews the relevant literature. Section 3 presents the data used in the internal and external validation steps. In Sect. 4 we describe both the internal and external validation procedures in detail,

[2] Our monthly BTS questions are all qualitative (respondents have choose between positive, neutral and negative assessments) and quantified as balance indicators, computed as the differences between the shares of positive and negative answers to each question. For details, see Sect. 3.

and we summarise the results obtained, distinguishing the between pre-COVID-19 period (up to 2019) as well as the COVID-19 pandemic years (2020–2021). Section 5 summarises our findings. The implementations of the imputation algorithms are described in the Appendix, along with reproductions of the KOF monthly and quarterly survey questionnaires.

2 Review of the Literature

In general, temporal disaggregation can be done with either univariate or multivariate approaches. The former type ranges from equal distribution of quarterly outcomes to monthly subdivision or interpolation, assuming a smooth evolution. Non-parametric approaches like spline methods or methods based on the quadratic function minimisation like that of Boot et al. (1967) or Denton (1971) also belong to this category. At the right margin of a series, univariate methods either hold the very last observation fixed (last observation carried forward) or try to exploit the momentum of the series, either by simple extrapolation or by modelling ARIMA processes or univariate state-space models, like in Harvey and Pierce (1984). Yet, even if some of these approaches are technically quite sophisticated, their inherent shortcoming lies in the fact that missing observations are imputed without resorting to any information except for the series itself. While the results may be satisfying when a time series evolves smoothly, these methods by construction cannot capture structural breaks or economic shocks before some time has passed. In other words, the resulting higher frequency breakdowns will be particularly flawed when the economic situation is changing, i.e., when timely information is particularly needed. Feijoó et al. (2003) evaluate the performance of several popular univariate temporal disaggregation procedures. They confirm that univariate higher frequency transformations are informationally inefficient when other data are disregarded that could supply valid information on the true movement of the process through time at points that are unobserved in the target series. This information inefficiency can be successfully addressed by employing one or several auxiliary higher-frequency variables that are related to the low-frequency time series in question.

Let y_t^l be a low-frequency variable that we intend to temporally disaggregate to a higher frequency. The most popular methods for temporal disaggregation assume a linear relationship between an unknown high-frequency representation of the low-frequency variable y_t^h and a related high-frequency variable x_t in the following form:

$$y_t^h = \alpha + \beta x_t + \varepsilon_t \tag{1}$$

where various assumptions are made on the nature of the disturbance term ε_t. For example, Chow and Lin (1971) assume an AR(1) process. Fernández (1981) assumes that ε_t follows an I(1) process and in Litterman (1983) an ARIMA(1,1,0) process is suggested. Proietti (2006) proposes a general state-space model that encompasses these three temporal disaggregation approaches. Another approach to multivariate temporal disaggregation, based on structural time series models, is suggested by Moauro and Savio (2005). In a Monte Carlo simulation, their approach compares favourably with several traditional methods. They conclude (p. 230) that "what this limited experiment seems to indicate is that the choice of the method for time aggregation can be even more relevant than the use of a good reference series, even if the use of this series can substantially add in terms of accuracy when an appropriate framework for time aggregation is chosen."

The multivariate methods for temporal disaggregation mentioned above employ one, or at most, a handful of high-frequency indicators. When more high-frequency indicators are available, possibly exceeding the number of observations of a low-frequency variable, the quality of temporal disaggregation can be negatively influenced by data over-fitting. Hence, a variable selection procedure is required that retains only the most informative high-frequency variables regarding the dynamics of the low-frequency variable of interest. Several variable selection procedures, like forward, backward or stepwise, are common. Alternatively, penalised regressions can be used for variable selection, like the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996).

Alternatively, Stock and Watson (2002b) propose an approach to impute values in large data panels involving variables sampled at heterogeneous frequencies. This is based on the Expectation-Maximization (EM) algorithm, which typically consists of recursive iterations. In the first step, the missing values are substituted with the best guess for some given initial parameter values defining the common factors, which are extracted from the data set by means of principal components analysis. In the second step, the imputed values are updated, conditional on the specified parameter values. Then the missing values are imputed again, conditional on the new parameter values. The iterating procedure is continued until some convergence criteria are met. This approach is used, for example, in Schumacher and Breitung (2008) for imputation of missing values in monthly and quarterly German economic data. Since September 2015, the EM procedure is applied for temporal disaggregation of the quarterly components of the KOF Economic Barometer, a composite leading economic indicator for the Swiss economy (Abberger et al., 2018) and since January 2020 for the KOF/FGV Global Barometers (Abberger et al., 2022).

In this paper we show that it is straightforward to incorporate the procedure developed in Wang et al. (2007) for estimating the LASSO regression of Tibshirani (1996) in the presence of autocorrelated disturbances into the temporal disaggregation procedure suggestion in Chow and Lin (1971). The advantage of such a combination is that the LASSO regression performs selection of the most informative auxiliary indicators that will be used for temporal disaggregation of the target time series. In doing so, the penalised regression also solves the curse of dimensionality problem.

In a parallel contribution to ours, Mosley, Eckley, and Gibberd (2022) suggest a *sparse temporal aggregation* approach that combines the standard Chow-Lin (1971) method with the LASSO penalised regression of Tibshirani (1996) as well as the adaptive LASSO algorithm of Zou (2006). Yet, our approach is different from Mosley et al. (2022) in several important aspects. First and foremost, we apply the temporal disaggregation procedures to a completely different data set, involving business tendency surveys collected in Switzerland. Second, Mosley et al. (2022) focus exclusively on interpolation, leaving the assessment of extrapolation for future

Table 1 Questions from the monthly KOF manufacturing BTS

Questionnaire item	Topic; answer options: up/too high, constant/about right, down/too low
1a	Incoming orders compared to previous month
1b	Incoming orders compared to 12 months ago
2a	Order backlog compared to previous month
2b	Assessment of order backlog
3a	Production compared to previous month
3b	Production compared to 12 months ago
5a	Inventories of final goods compared to previous month
5b	Assessment of final goods' inventories
8a	Expected incoming orders in the following three months
8c	Expected production in the following three months
8d	Expected purchase of intermediary goods in the following three months

research, while we assess the accuracy of both interpolation and extrapolation, thus addressing the problem of "ragged edge" at the right margin. Thirdly, we compare the quality of the sparse temporal disaggregation technique with that based on the Expectation–Maximisation algorithm of Stock and Watson (2002a) developed for the factor-type models. This exercise, comparing algorithms designed for the sparse versus dense data, was also left for the future research in Mosley et al. (2022). Our findings thus complement and reinforce those reached in Mosley et al. (2022) on the usefulness of temporal disaggregation techniques in data-rich environments.

3 Data

Our data are primarily from the monthly KOF Swiss Economic Institute's BTS in the Swiss manufacturing sector. Presently, more than 1,000 firms are surveyed. The latest monthly questionnaire comprises 21 items. The response rate is about 70 per cent. Due to changes to the questionnaire in the past, we have access to eleven monthly series going back more than 50 years.[3] They are summarized in Table 1.

The underlying questions are all qualitative, with three options to answer: down/too low (−), no change/about right (=), up/too high (+). For quantification, the standard is to compute the balance indicator (percentage share + minus percentage share −). The balance tends to move up and down over the business cycle, and while the series' averages may not be exactly constant in the long run, the quantification has an upper bound of 100% and a lower bound of − 100%. These variables can thus increase or decrease only within narrow limits, so that we treat them as

[3] The 10 questionnaire items with shorter time series, hence omitted from the analyses, are 2c (assessment of export order backlog), 4a (change of intermediate products inventory), 4b (assessment of intermediate products inventory), 6 (assessment of employment level), 7a to 7d (assessment of current business situation, assessment of future business situation, difficulty to predict future development of own business, uncertainty of future development of own business), 8b (expected export orders) and 8e to 8g (expectations regarding number of employees, selling prices and purchasing prices).

mean-stationary. The internal validation is conducted exclusively with eleven series from the monthly BTS. These variables also constitute the higher frequency information set for the external validation procedure.

The target series for the external validation procedure with *genuine* imputations of *truly* unobserved monthly values, *technical capacity utilisation in per cent*, comes from the quarterly KOF manufacturing BTS, which is conducted during the first month of each quarter. It is quantitative, computed as the weighted average of the responses of the surveyed firms on their technical capacity utilisation in per cent.[4] During the observation period, it fluctuates around 82% with a maximum of 88.2% and a minimum of 75.9%. Analogous to the monthly BTS series, we treat it as mean-stationary. The remaining variables for the external validation are monthly data reflecting the economic situation, taken from other sources than the KOF BTS, namely the KOF Economic Barometer, the KOF/FGV Leading and Coincident Global Barometers, the Swiss National Bank's Business Cycle Index, the Swiss unemployment rate and the Swiss inflation rate. We use seasonally adjusted data.

The latest observations referred to in this paper relate to December 2021 or the fourth quarter of 2021, respectively. Considering the COVID-19 pandemic and its drastic impact on the economies around the world, we first restrict the analyses to data points up to 2019m12. Then, we verify the robustness of the results referring to the years 2020–2021, when the pandemic affected the economy most.

4 Methods and Results

In this section, we present the temporal disaggregation methods evaluated in this paper, their implementation, and the resulting accuracy statistics. The first subsection deals with the internal validation procedure, the second looks at the external validation exercise. Before proceeding, a few preliminary considerations are in order.

First, imputing unobserved higher-frequency values only makes sense when the process that is observed at the lower frequency is in fact occurring at the higher (target) frequency. This is evident for continuous processes like the production of goods (output) and services or their use (consumption, investment). For genuinely discrete processes, however, the highest frequency to be estimated in sensible way is that of their occurrence. Retail trade sales preceding Christmas, for example, occur once per year. It would not make sense to impute quarterly or monthly values, although this may be possible in technical terms. Hence, before a series is submitted to a procedure to generate a corresponding higher frequency series, a reality check must be performed to confirm that the process indeed occurs at the higher frequency and the fact that there are unobserved data points is solely due to the lower measurement frequency. In our case, the process is the economic situation, which is clearly

[4] For the latest version of the questionnaires, including the quarterly BTS, refereed to later in this paper, see https://ethz.ch/content/dam/ethz/special-interest/dual/kof-dam/documents/FragebogenArchive/imt/inu_en_q.pdf.

evolving in a continuous fashion, but its measurement is performed at discrete intervals, so that temporal disaggregation is justified.

Second, temporal disaggregation has to distinguish between stock and flow variables. Temporal disaggregation of flow variables should ensure that the higher frequency data points add up to the lower frequency values covering the corresponding period. This requires revisions of imputations beyond the last know observation (extrapolations) as soon as the corresponding lower frequency observation becomes available, and they technically change from out-of-sample to in-sample. Stock variables are different, as it not obvious whether or in which way lower frequency data impose restrictions on higher frequency imputations. For some applications, it may make sense that the averages of a disaggregated series should be equal to the value of the corresponding lower frequency observations, but his depends on the nature of the data generating process, so that a general answer cannot be given. For our purposes however, as we are dealing with BTS data reflecting the assessment of a continuously changing environment, coming as stock variables at discrete intervals, there is no need to restrict the joint values of the monthly breakdowns to a corresponding observed quarterly value.

Consequently, for our purposes there is no requirement to distinguish between imputation methods for flow and stock variables, and we do not impose any restrictions on the imputed higher frequency values to be derived from the corresponding lower frequency values when or as soon as they are available.

Last but not least, as we distinguish between ex-post and simulated real-time, we are faced with two options, namely rolling or expanding windows, where the simulated real-time imputation is the last observation at the right margin. For our purposes, we chose to refer to 20-years rolling windows, which keeps the degrees of freedom constant as the window is subsequently moved forward in one-month steps. The length of the period is a compromise between "sufficiently long" to cover more than one business cycle and "sufficiently short" to avoid losing too many simulated real-time imputations.

4.1 Internal Validation

The setup of our simulation exercise is shown in Fig. 1, summarizing the temporal disaggregation methods and terminology of our paper as well as the time pattern of the observed and imputed values.

To illustrate the time pattern, the figure reflects the information set available to an observer in the end of the last month of a year, looking back four quarters/twelve months into the past. The monthly series (KOF BTS series, $k=11$) relate to the months in which they are released. Accordingly, towards the end of December (m12) the monthly BTS data releases for the year that is coming to its end are complete. The quarterly BTS series consist of data points, which are released in and relate to the first month of each quarter. This corresponds to the actual release

												Observer
quarter	q1			q2			q3			q4		
month	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12
monthly series 1	x	x	x	x	x	x	x	x	x	x	x	x
monthly series 2	x	x	x	x	x	x	x	x	x	x	x	x
...	x	x	x	x	x	x	x	x	x	x	x	x
monthly series *k*	x	x	x	x	x	x	x	x	x	x	x	x
quarterly series	x			x			x			x		
	temporal disaggregation											
*locf**	actual	as m1	as m1	actual	as m4	as m4	actual	as m7	as m7	actual	as m10	as m10
cubic spline	actual	symmetric	symmetric	actual	symmetric	symmetric	actual	symmetric	symmetric	actual	asymmetric	asymmetric
multivariate	actual	symmetric	symmetric	actual	symmetric	symmetric	actual	symmetric	symmetric	actual	asymmetric	asymmetric
	simulated real-time											
temporally disaggregated quarterly series	ex post										ex ante	
	actual	imputation	imputation	actual	imputation	imputation	actual	imputation	imputation	actual	postcast	nowcast
		in-sample			in-sample			in-sample			out-of-sample	

Fig. 1 Timeline and terminology

calendar of the KOF BTS.[5] Correspondingly, whilst the four "actual" values (m1, m4, m7 and m10) of the quarterly series are known, the disaggregation challenge is to impute the eight unobservable values for the second and third month of each quarter (m2, m3, m5, m6, m8, m9, m11 and m12). We distinguish between "symmetric" and "asymmetric" imputation environments. Imputations are symmetric when or as soon as they fall in between two observed values (m2, m3, m5, m6, m9 and m9), and we refer to them as "ex-post" (or "in-sample"), whereas imputations are asymmetric when or as long as all actual values lie in the past, and we refer to them as "ex-ante" or "out-of-sample". In Fig. 1, those two are labelled "postcast" (m11) and "nowcast" (m12), because in technical terms, they are forecasts (extensions of time series beyond the last known value), but as they relate to the present (m12) or the recent past (m11), the terminology denotes the chronology. All ex-ante imputations are conducted in simulated real-time. In particular, these imputations are calculated referring to repeated 20-years rolling windows ranging from m*t*–240 to m*t*, where the results for m*t* are the nowcasts and those for m*t*–1 correspond to the postcasts. Moreover, notice that the actual are values preserved in the temporally disaggregated series, even when the algorithms would compute imputed values to replace the known ones.

We consider two univariate and four multivariate imputation algorithms. The computationally simplest algorithm is to carry the last observation forward (*locf*), or what is known as the "hot-deck" estimate. As *locf* is an implementation of a random walk without drift and hence the most "conservative" estimate, it will also serve as our benchmark, which the competing methods have to outperform in order to be considered as serious alternatives.[6] The competing temporal disaggregation methods include a univariate implementation of the cubic spline (spline), the Expectation-Maximisation (EM) algorithm of Stock and Watson (2002a), generalised multiple regression (GLS), the traditional Chow-Lin (1971) algorithm, and our modification of the latter, which is based on the following considerations: Allowing multivariate procedures to incorporate information from *all* available auxiliary indicators may lead to excessive numbers of indicators. For the factor type EM algorithm, this is less likely to pose a serious problem, but GLS and Chow-Lin may suffer from the curse-of-dimensionality phenomenon. In order to mitigate this, we combine the standard Chow-Lin (1971) procedure with the least absolute shrinkage and selection operator (LASSO) regression of Tibshirani (1996) based on approaches suggested in Wang et al. (2007). Because of the selection properties of the LASSO method, only a subset of the available auxiliary indicators is used in temporal disaggregation.[7]

[5] Typically, in the Europe, monthly BTS questionnaires are distributed and collected during the month they relate to and published before the end of this month. During the first month of a quarter, the quarterly BTS are conducted in parallel with the monthly BTS.

[6] As a default option to impute missing values, many statistical packages compute the variable average. While this may be an appropriate default for cross-sectional analyses, it would be *too naïve* for our time series, as it ignores the predictive quality of the last known observation, the "hottest" punched card on top of the file.

[7] Detailed descriptions of the imputation algorithms as implemented in our paper are given in the Appendix.

Table 2 Observation periods

Period	Imputable data points	Evaluation focus
1967m2–2021m12	4,840	Entire data base
1991m3–2019m12	2,541	Ex-ante, direct & indirect imputations
2001m1–2019m12	1,760	Ex-ante & ex-post, direct & indirect imputations
2020m1–2021m12	176	Sensitivity check for the COVID-19 pandemic

A further distinction within our simulation exercise concerns the quantification of the qualitative BTS survey items. As the monthly BTS questions are qualitative and the common quantification is the balance, we can either *directly* impute the missing observations of the balance indicators, or we can impute the positive and negative shares and *indirectly* derive the imputed target values from the difference of the two imputed series. A priori, it is not evident which approach should be superior. The indirect imputation increases the number of observations to be imputed by 100%, but it is informationally more efficient, as the same value for the balance statistics can result from widely different positive and negative shares. Accordingly, we let the data speak and proceed along both paths.

The observations of the eleven monthly balance BTS time series cover the period 1967m2–2021m12, amounting to 659 observations per time series. Correspondingly, the artificially created quarterly time series by the skip-sampling procedure retain 219 observations each, implying that 440 data points can be imputed, which would amount to 4840 data points in total. However, the negative and positive shares have been stored in the KOF data base from 1971m4 only, and unfortunately, there is no way to recover them. In order to be able to compare the accuracy of the *direct* and *indirect* temporal disaggregation, in this paper we refer to the shorter sample period. Accordingly, the first 20-year rolling covers the period 1971m4–1991m3, and the resulting simulated real-time data start in 1991m3. Carrying this forward to 2019m12 yields 231 simulated real-time (ex-ante) imputations per survey item and a total of 2541 pre-COVID-19 imputations for the eleven series. A further consideration concerns how we define ex-post. In this paper, under ex-post, we understand the imputations from one and the same 20-years rolling window (vintage), namely the one that ends before 2020, when the COVID-19 hit. The ex-post vintage thus covers the period of 2000m1–2019m12, comprising 160 imputations per survey item and a total of 1,760 imputations for our eleven series. Finally, to verify the robustness of our results with respect to the inclusion of the COVID-19 years, we separately conduct the analysis over the period 2020m1–2021m12. Here, the number of imputations years is 16 per series, amounting to 176 for the eleven indicators in total. The different observation periods and the corresponding types of analyses are shown in Table 2.

The accuracy of the imputations is measured with the root mean squared error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\mathrm{Predicted}_i - \mathrm{Actual}_i)^2}{N}} \tag{2}$$

The RMSE is computed for different sets of imputed data points, distinguished by the following dimensions: Firstly, for all imputations taken together (jointly for the second and third month of a quarter) versus separately for the second and the third month of a quarter. Secondly, we distinguish between ex-ante and ex-post imputations. Thirdly, direct imputations of the balance are distinguished from indirect ones via positive and negative shares.

To evaluate the precision of the different measures against our benchmark (last observation carried forward), we compute the *relative* RMSE (RRMSE), defined as the ratio of the RMSE imputation method under consideration against the RMSE of the benchmark. Ratios below one indicate higher precision of the imputation method under consideration, ratios above one indicate that the benchmark method delivers superior results.

In addition to this, taking into account the interest of economic observers and policy makers in upward or downward movements and trends in real-time, corresponding to the right margins of economic time series, we also analyse the directional accuracy of the imputations, both ex-post and ex-ante. With the actual values being at hand for comparison, it is straightforward to compute the percentage of corresponding up and down direction changes recorded for the second and third months in each quarter as well as for both taken together. Here, values above 50% indicate that the imputations are reflecting direction changes better than simple guessing by flipping a coin.

4.2 Internal Validation: Results

4.2.1 Imputation Accuracy in Terms of (R)RMSE

The accuracy of the imputations is summarised in Table 3. The panel on the left reports the RMSE statistics computed for interpolated values in-sample (ex-post analysis), the panel on the right shows the extrapolated values out-of-sample (ex-ante analysis), respectively. To facilitate the comparison of the imputations in terms of average accuracy, the lowest RMSE of the respective cells in **bold** numbers are highlighted (best method for imputations of second and third month of a quarter taken together and separate for both imputed months).

Obviously, in terms of RMSE, Chow-Lin and the Chow-Lin/LASSO imputations outperform all alternative approaches, but several additional conclusions can be drawn from Table 2.

For the benchmark, last observation carried forward, three regularities stand out: First, the RMSE are always larger for m3 compared to m2 (with the RMSE for both months together lying in between by construction). This is in line with expectations, as the lag to the last recorded observation increases, and therefore, if a series is not quickly reverting to come steady state, the error is prone to increase. Second, the

Table 3 Comparison of imputation accuracy, RMSE

Method	Period	Ex-post (in-sample)						Ex-ante (out-of-sample)					
		1991–2019		2000–2019		2020–2021		1991–2019		2000–2019		2020–2021	
	Imputation/Months	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect
	2nd and 3rd	–	–	7.26	7.05	10.25	9.94	6.98	6.78	7.26	7.05	10.25	9.94
locf	2nd	–	–	7.11	6.91	9.03	8.76	6.93	6.72	7.11	6.91	9.03	8.76
	3rd	–	–	7.40	7.19	11.34	11.00	7.02	6.85	7.40	7.19	11.34	11.00
cubic spline	2nd and 3rd	–		6.87	6.66	8.98	8.75	8.91	8.61	9.41	9.11	19.07	18.74
	2nd	–		6.73	6.54	9.11	8.92	7.98	7.71	8.30	8.04	13.16	12.92
	3rd	–	–	7.00	6.78	8.85	8.57	9.75	9.42	10.40	10.07	23.54	23.14
	2nd and 3rd	–	–	8.35	8.23	11.59	11.14	8.50	8.57	8.76	8.70	11.78	11.40
EM	2nd	–	–	8.04	7.93	11.83	11.61	8.35	8.42	8.60	8.57	11.94	11.77
	3rd	–	–	8.64	8.52	11.34	10.65	8.65	8.72	8.91	8.83	11.62	11.02
	2nd and 3rd	–	–	4.62	4.72	5.84	5.76	4.79	4.75	4.94	4.92	6.02	6.01
GLS	2nd	–	–	4.65	4.71	6.06	6.41	4.88	4.77	5.01	4.91	6.10	6.35
	3rd	–	–	4.59	4.73	5.62	5.02	4.71	4.73	4.87	4.92	5.95	5.65
	2nd and 3rd	–	–	3.79	3.74	5.31	**4.68**	**3.93**	3.80	3.99	3.91	**5.64**	**5.19**
Chow-Lin	2nd	–	–	3.74	3.64	5.49	5.16	**4.00**	3.84	3.95	3.83	**5.49**	**5.14**
	3rd	–	–	**3.84**	3.84	5.13	**4.15**	**3.85**	3.76	**4.02**	3.98	**5.79**	**5.24**
Chow-Lin/ LASSO	2nd and 3rd	–	–	**3.76**	**3.68**	**5.14**	**4.68**	3.96	**3.75**	**3.98**	**3.85**	5.73	5.38
	2nd	–	–	**3.66**	**3.59**	**5.27**	**4.96**	**4.00**	**3.79**	**3.92**	**3.78**	5.64	5.28
	3rd	–	–	3.85	**3.78**	**5.00**	4.39	3.91	**3.70**	4.04	**3.92**	5.82	5.48

indirect imputations via the difference of imputed positive and negative shares turn out to be slightly but consistently more precise than the direct imputations of the balance. Third, compared to the periods up to 2019 the accuracy of the imputation is markedly worse in the years 2020–2021, when the COVID-19 pandemic had severe and rapidly changing impacts on the economy. Notice that the RMSE are equal ex-ante and ex-post, as *locf* does not revise imputations when new information becomes available.

For the only other univariate method considered here, the cubic spline, the RMSE across all specifications range from 6.54 to 23.54 and the median equals 8.92, which is high compared to the benchmark (*locf*: RMSE 6.54–23.54, median 7.62). Ex-post, the accuracy is somewhat better than ex-ante for 2000–2019, and considerably better for 2020–2021. The cubic spline is particularly inferior the benchmark ex-ante, and this holds especially for the COVID-19 years, when the economic situation changed quickly and more than once. In our data corpus, the spline's extrapolation of the most recent trend is hence more misleading than the assumption of a random walk. Ex-post, the interpolations tend to outperform the benchmark slightly, but the dismal results at the right margin rule out the spline as an alternative to *locf* when real-time accuracy matters most. Furthermore, similar to what we find for the benchmark, the RMSE tend to be larger for m3 compared to m2, and they are consistently larger for the direct compared to the indirect imputations.

For the EM algorithm, the RMSE range from 7.93 to 11.94 with a median of 8.74, which is again higher than for the benchmark. The ex-post accuracy is only slightly better than ex-ante, and the tendency for m2 to be imputed more precisely than m3 as well as the superiority of the indirect imputation of the balances is less visible than with *locf* or the cubic spline. In our data corpus, the EM algorithm imputation hence does not hold the promise to outperform the naïve *locf*. This applies to all periods considered as well as to ex-ante and ex-post.

For the GLS regressions, the RMSE range from 4.59 to 6.41 (median 4.92). In our data corpus, the regression imputations consistently outperform the benchmark. Moreover, GLS is also consistently superior to EM. This applies to all periods considered, ex-ante and ex-post. Interestingly, the third month of a quarter is not clearly imputed less precisely than the second. As the GLS imputations are cross-sectional and therefore disregarding the time series-properties of the data, they should exploit the information from the ten auxiliary variables for all imputed months equally well. In other words, while the relatively stable accuracy across the various imputation environments may appear as an advantage of GLS, the drawback is that without neglecting the time series properties, the imputations could be even better, in particular in-sample and, although to lesser degrees, for the second and third months of the final quarter (ex-ante postcasts and nowcasts). Chow-Lin and our modification will address this.

For the Chow-Lin algorithm, both with and without LASSO, the RMSE range from 3.59 to 5.82 with a median of 3.98. Not only do they consistently outperform the benchmark in terms of accuracy; they are also consistently superior to all other methods under consideration. The extension of the standard Chow-Lin method by LASSO proves its viability, as for both long periods (1991–2019 and

2000–2019) the combination Chow-Lin/LASSO tends to deliver somewhat higher imputation accuracy compared to the standard implementation of the Chow-Lin method. On a side note, in our simulation exercise, the number of auxiliary indicators was not excessively large, such that the standard Chow-Lin method was also feasible. In other settings, however, when the number of auxiliary monthly indicators may well exceed the number of low frequency observations, the modified Chow-Lin procedure would be the only option. In addition, during the COVID-19 years, the standard Chow-Lin method delivered a slightly better ex-ante forecast accuracy but given the comparatively small number of observations available for the evaluation, this finding is to be taken rather precautionary.

Taken together, we observe a clear ranking of the different approaches: Chow-Lin without and with the LASSO modification produce the most accurate imputations. GLS delivers the third best results, whereas at least in our data corpus, the EM and the cubic spline exhibit strongly inferior performance.

Across methods the indirect imputations via imputed positive and negative shares tend to be more precise than the direct imputations of the balance. Ad ad-hoc explanations of this surprising finding may point to the fact that indirect imputations are informationally more efficient, as the same value for a balance can result from combinations of different positive and negative shares with different neutral shares. Another reason may lie in the fact that differences of seasonally adjusted series are usually not identical to the seasonally adjusted difference of two series. If the seasonality of the underlying series is more likely to be captured by a filter than seasonal pattern from the combined series, it may be easier to impute the underlying series. Further research on this is called for, as this finding may be especially useful in practical applications.

Last but not least, the RMSE are larger for m3 compared to m2 for the univariate methods and the EM algorithm, but not for the other multivariate methods, which apparently successfully exploit the information from the auxiliary variables likewise for all imputed months.

To illustrate the relative performance of the imputation algorithms with respect to the benchmark, Table 3 replicates the structure of Table 2, but it reports the RRMSE (*relative* RMSE, computed as the ratio of the RMSE in each cell of Table 2 to the corresponding value of *locf*). It thus quantifies the gains (RRMSE $<$ 1) or losses (RRMSE $>$ 1) in imputation precision in comparison to the benchmark.

As can be inferred from Table 4, both implementations of the Chow-Lin method achieve about 45% reduction in the RMSE relative to the benchmark. For the GLS regressions, the improvement is less pronounced, but it is still amounting to about 30% reduction in the corresponding metric. EM method is nearly always less precise than the benchmark, with most RRMSE clearly exceeding one. The cubic spline method is only slightly better than the benchmark for the interpolating of data points ex-post, but its performance completely deteriorates out-of-sample with all reported RRMSE values markedly exceeding unity.

Comparing the *improvements* of accuracy of imputations relative to the benchmark computed in-sample (ex-post) versus out-of-sample (ex-ante), the former are generally more pronounced than the latter. The reason for this is that the RMSE are

Table 4 Comparison of imputation accuracy, relative RMSE (RRMSE)

Method	Period	Ex–post (in–sample)						Ex–ante (out–of–sample)					
		1991–2019		2000–2019		2020–2021		1991–2019		2000–2019		2020–2021	
	Imputation/Months	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect
locf	2nd and 3rd	–	–	–	–	–	–	–	–	–	–	–	–
	2nd	–	–	–	–	–	–	–	–	–	–	–	–
	3rd	–	–	–	–	–	–	–	–	–	–	–	–
Cubic spline	2nd and 3rd	–	–	0.95	0.94	0.88	0.88	1.28	1.27	1.30	1.29	1.86	1.88
	2nd	–	–	0.95	0.95	1.01	1.02	1.15	1.15	1.17	1.16	1.46	1.48
	3rd	–	–	0.95	0.94	0.78	0.78	1.39	1.37	1.40	1.40	2.08	2.10
EM	2nd and 3rd	–	–	1.15	1.17	1.13	1.12	1.22	1.26	1.21	1.23	1.15	1.15
	2nd	–	–	1.13	1.15	1.31	1.33	1.20	1.25	1.21	1.24	1.32	1.34
	3rd	–	–	1.17	1.18	1.00	0.97	1.23	1.27	1.20	1.23	1.02	1.00
GLS	2nd and 3rd	–	–	0.64	0.67	0.57	0.58	0.69	0.70	0.68	0.70	0.59	0.60
	2nd	–	–	0.65	0.68	0.67	0.73	0.70	0.71	0.70	0.71	0.68	0.73
	3rd	–	–	0.62	0.66	0.50	0.46	0.67	0.69	0.66	0.68	0.52	0.51
Chow–Lin	2nd and 3rd	–	–	0.52	0.53	0.52	0.47	0.56	0.56	0.55	0.55	0.55	0.52
	2nd	–	–	0.53	0.53	0.61	0.59	0.58	0.57	0.56	0.55	0.61	0.59
	3rd	–	–	0.52	0.53	0.45	0.38	0.55	0.55	0.54	0.55	0.51	0.48
Chow–Lin/ LASSO	2nd and 3rd	–	–	0.52	0.52	0.50	0.47		0.57	0.55	0.55	0.56	0.54
	2nd	–	–	0.52	0.52	0.58	0.57	0.58	0.56	0.55	0.55	0.62	0.60
	3rd	–	–	0.52	0.52	0.44	0.40	0.56	0.54	0.55	0.54	0.51	0.50

generally higher for ex-ante imputations than ex-post for all imputation algorithms except *locf*, where they are equal ex-post and ex-ante (see Table 2).

Moreover, comparing precision of imputing values for the second versus the third month in each quarter. Lastly, the imputation accuracy somewhat deteriorated during the COVID-19 period, reflecting the elevated level of uncertainty and turbulence brought about by the pandemic.

4.2.2 Directional Accuracy

Our second quantification of the accuracy of the imputations relates to turning points. Quarterly series as well as imputed monthly *locf* series are silent on changes of directions for the unobserved/imputed months. Yet, economic observers may be particularly interested in upward or downward movements at the right margins of economic time series. The directional accuracy of the imputations resulting from the different disaggregation methods is summarised in Table 5. It shows the percentage of correctly predicted up- or downward movements from one month to the other. Values above 50% indicate that the imputations are reflecting direction changes better than random guessing. Turning points are not computed for the RMSE benchmark, as imputations by carrying the last observed value forward do not have direction changes by construction. Accordingly, here the benchmark for the usefulness of an algorithm is the 50%-hurdle.

The ranking is very similar to that for the point forecast accuracy (Tables 2 and 3). Chow-Lin/LASSO takes the lead, closely followed by Chow-Lin, and then by GLS and EM. The cubic spline shows the strongest contrast when comparing the directional forecast accuracy for ex-post versus ex-ante imputations. For the imputations computed in-sample, the hit ratio is about 60%, whereas for imputations out-of-sample it drops to 50% and below, which is comparable to what one would expect from random coin flipping, or worse For the remaining procedures the hit ratios are comparable across ex-post/ex-ante computations.

Based on our data corpus, the implied recommendations regarding the preferred algorithms for imputing monthly values for quarterly economic time series remain the same when the focus is on the correct direction from one month to the other rather than on the precision of the level: Chow-Lin and in particular Chow-Lin/LASSO stand out, and GLS is not much inferior. The cubic spline algorithm, on the other hand, exhibits clearly unsatisfactory performance ex-ante and is at best moderately useful ex-post. The only qualitative difference to the findings based on the point forecast accuracy measured in terms of RMSE is that the EM algorithm proves to be relatively accurate in terms of directional accuracy. Yet, the percentage of hits is markedly less than for Chow-Lin/LASSO, Chow-Lin or GLS, so again, EM leaves much room for improvement.

Table 5 Directional accuracy, share of correctly predicted monthly up-/downward movements

	Ex-post (in-sample)						Ex-ante (out-of-sample)					
Period	1991–2019		2000–2019		2020–2021		1991–2019		2000–2019		2020–2021	
Imputation	direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect
Cubic spline	–	–	60.9%	60.6%	61.9%	60.2%	50.5%	51.6%	49.4%	51.5%	47.7%	46.0%
EM	–	–	62.8%	63.7%	66.5%	66.5%	64.3%	65.6%	63.5%	65.9%	64.2%	65.9%
GLS	–	–	74.4%	73.9%	75.0%	72.7%	74.8%	73.7%	75.3%	74.2%	72.7%	71.6%
Chow-Lin	–	–	**77.7%**	76.1%	75.6%	72.7%	**75.9%**	75.0%	**77.1%**	76.0%	70.5%	72.2%
Chow-Lin/LASSO	–	–	77.5%	**76.8%**	**77.8%**	**74.4%**	75.2%	**75.6%**	76.2%	**77.4%**	**73.3%**	**72.7%**

4.3 External Validation

In this section, we apply the best-performing method from the preceding internal validation, Chow-Lin/LASSO to convert our quarterly time series of interest, *technical capacity utilisation in per cent* from the quarterly KOF BTS ($CapU^q$), into monthly frequency ($CapU^m$). Then, we verify whether the resulting temporally disaggregated time series is superior to the benchmark, capacity utilisation temporally aggregated with the last value observation carried forward algorithm ($CapU^{locf}$) in explaining the within-quarter variance of a selection of monthly business cycle indicators reflecting economic conditions that can *a priori* be assumed to be closely related to what is measured by *CapU*. Since in this exercise the monthly *CapU* values are genuinely unknown, a forecast encompassing test is applied.

The external validation consists of the following steps. First, we select six monthly economic indicators as reference time series. They cannot be taken from the monthly KOF BTS, as this is where the within quarter variance of the disaggregated series comes from, but they should still have a close relation to the Swiss business cycle. We identified six suitable and publicly available time series:

1.1. The KOF Economic Barometer, one of the most prominent monthly composite leading indicators for the Swiss economy.[8]
1.2. The Leading Global Barometer, a monthly composite leading indicator for the world economy, developed and published jointly be the KOF Swiss Economic Institute and the Brazilian Fundação Getúlio Vargas (FGV).[9]
1.3. The Coincident Global Barometer, which corresponds to the Leading Indicator with the exception that it does not target a lead to the world economy but instead a stronger congruence with it.
1.4. Swiss National Bank's Business Cycle Index (SNB BCI), a monthly composite indicator designed to reflect the Swiss GDP growth cycle.[10]
1.5. The official Unemployment Rate, a monthly series based on the number of registered unemployed persons in Switzerland.[11]
1.6. Inflation, measured as the year-on-year growth rate of the Swiss Consumer Price Index.[12]

Where seasonally and calendar day adjusted data are available, we take these, otherwise we use the series unadjusted as published. We refer to these six series as "external" ($E^m i$). To verify that the external time series and *CapU* are indeed reflecting the same – or at least a similar – data generating process and to identify the lead/lag structure, we compute pairwise cross-correlations of $CapU^{locf}$ and the $E^m i$,

8 For details, see https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-economic-barometer.html.

9 For details regarding this and the next indicator on the list, see https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalbaro.html.

10 For details, see https://data.snb.ch/en/topics/snb/chart/snbbcich.

11 For details, see https://www.bfs.admin.ch/bfs/en/home/statistics/work-income/unemployment-underemployment.html.

12 For details, see https://www.bfs.admin.ch/bfs/en/home/statistics/prices/surveys/lik.html.

Table 6 Cross correlations, capacity utilisation and external series

External series	*CapU*	Max abs. cross correlation	Lead/lag of external series
KOF Economic Barometer			
Level	y–o-y growth rate	0.86	7 months lead
KOF/FGV Leading Global Barometer			
Level	y–o-y growth rate	0.85	6 months lead
KOF/FGV Coincident Global Barometer			
Level	y–o-y growth rate	0.84	4 months lead
SNB Business Cycle Indicator (BCI)			
y–o-y difference	y–o-y growth rate	0.80	9 months lead
Unemployment Rate			
y–o-y difference	y–o-y growth rate	–0.65	3 months lag
Consumer Price Index (CPI) Inflation			
Level	Level	0.50	2 months lead

allowing up to 15 months lag/lead. We refer to $CapU^{locf}$ alternatively in levels and in year-on-year growth rates (y–o-y). The same applies for the six $E^m i$, but here we compute annual differences rather than growth rates, when the scale of a series mandates this. The cross correlations are computed with $CapU^{locf}$ covering the period 2000m1–2019m12, but we refer to all available data points of the six external series to identify the leads or lags.[13] For each combination of $CapU^{locf}$ and $E^m i$, we determine the highest absolute cross correlation, along with the corresponding variable specifications (levels or growth rates/differences) and leads/lags. Table 6 summarises the findings, ordered according to the maximum absolute cross correlations.

The magnitude of the highest (absolute) cross-correlations ranges from 0.86 to 0.50. Of the six external series, five are leading with respect to *CapU*. The highest cross-correlations (0.80 and above) are recorded for the domestic and international versions of the composite business cycle indicators. The cross-correlation with the unemployment rate is negative (–0.65), which makes sense, as unemployment is the inverse of labour utilisation and it is known to react to changing conditions with a certain inertia, so that it is a lagging indicator of the business cycle. The lowest absolute cross-correlation (0.50) is recorded with the CPI inflation. As pressure on technical capacity utilisation can be expected to go along with rising output prices, a positive association is to be expected, but changes to the price level are also caused by many other factors, so that a more than moderate correlation would have come as a surprise.

What we can take from this is that *CapU* is indeed a key indicator for the Swiss business cycle, and accordingly, we find it to correlate highly with the six external series that are also directly or indirectly reflecting the state of the Swiss economy.

What remains to be seen is whether within-quarter variance created by the imputation of monthly *CapU* values strengthens the associations. Our criterion for this

[13] The external series all go back to the 1990s and more than 15 months beyond 2019m12.

Table 7 Encompassing tests (2000–2019), out-of-sample imputations

Series/Statistics	Chow-Lin/LASSO *g*	t-stat	p-value	*locf* *h*	t-stat	p-value	R^2
KOF Economic Barometer	2.06	4.36	**<1%**	0.57	1.26	0.21	0.74
Leading Global Barometer	1.89	4.46	**<1%**	0.38	0.94	0.35	0.72
Coincident Global Barometer	1.84	4.50	**<1%**	0.16	0.40	0.69	0.68
SNB BCI	0.19	2.81	**<1%**	0.10	1.57	0.12	0.64
Unemployment Rate	-0.10	-2.90	**<1%**	-0.002	-0.06	0.95	0.44
CPI Inflation	0.20	7.50	**<1%**	-0.01	-0.84	0.41	0.25

is that the preferred Chow/Lin-LASSO algorithm must be “superior” to the naïve *locf* approach, but since we do not know what would have been the true observed monthly values, had a monthly survey been conducted, we define superiority as beating *locf* in predicting (forecasting) the external reference series. In particular, we let the monthly series $CapU^m$ resulting from temporal aggregation with the Chow-Lin/LASSO algorithm compete with the benchmark $CapU^{locf}$ in predicting the six reference series. This amounts to comparisons of non-nested models, for which the J-test is adequate (Davidson & MacKinnon, 1981; Mizon & Richard, 1986).

For the imputation of the missing monthly values of the $CapU^q$ time series with the Chow-Lin/LASSO algorithm, the auxiliary data are the eleven monthly BTS series (Table 1). The imputations cover the 20 years before the pandemic, 2000m1–2019m12, comprising 240 months, for which 160 unobserved values are imputed. Notice that $CapU^q$ is not a balance indicator but a quantitative survey item, so we do not conduct indirect imputations (as there are no positive- and negative shares).

The J-test is conducted as follows: Let H_1 and H_2 denote two rival models $Y = X_1\ g$ and $Y = X_2\ h$. Then the J-test will evaluate whether the predicted values of an alternative model ($X_2\ \hat{h}$ or $X_1\ \hat{g}$) significantly improve the fit of the rival model in the two following regressions:

$$Y = X_1 g + \varphi\left(X_2 \hat{h}\right) \tag{3}$$

$$Y = X_2 h + \tau\left(X_1 \hat{g}\right) \tag{4}$$

The test statistics are the t-values for φ and τ coefficients. Significance of φ along with insignificance of τ implies rejection of H_1 by H_2. Significance of τ only means that H_2 is rejected by H_1. When neither φ nor τ are significant, the test does not result in any particular model selection. When both φ and τ are significantly different from zero, both models must be considered as partly useful, but ultimately deficient, given the available information. Since our rival models are the different estimates of the known data points from the original monthly reference series R_t, so that X_1 and X_1 are not bundles of time series (vectors) but two single time series,

Table 8 Encompassing tests, COVID-19 years (2020–2021)

	Chow – Lin/LASSO			*locf*			R^2
Series/statistics	*g*	t-stat	p-value	*h*	t-stat	p-value	
KOF Economic Barometer	– 2.58	– 0.86	0.40	5.13	1.61	0.12	0.38
Leading Global Barometer	– 1.42	– 0.70	0.49	4.34	2.01	0.06	0.64
Coincident Global Barometer	– 1.78	– 0.92	0.37	4.65	2.27	**< 5%**	0.66
SNB BCI	0.24	0.62	0.54	– 0.004	– 0.009	0.99	0.27
Unemployment Rate	– 1.24	– 2.21	**< 5%**	– 0.02	– 0.40	0.40	0.87
CPI Inflation	0.26	4.75	**< 1%**	– 0.01	– 0.35	0.73	0.84

$CapU^m{}_{t-L}$ and $CapU^{locf}{}_{t-L}$, the J-test is here identical with the simpler *encompassing test* (E-test), which consists of running the regression

$$R_t = g\,\mathrm{Cap}\,U^m_{t-L} + h\,\mathrm{Cap}\,U^{locf}_{t-L} + \mu_t \tag{5}$$

and addressing the significance of the estimated coefficients *g* and *h* by the usual t-tests, where *L* is the optimal lead/lag parameter shown in Table 6, and μ_t is the error term. The decision rule equals that of the J-test. We refer to the *out-of-sample* imputations of *CapU*, since this is what matters most when economic observers try to understand what is happening in real-time as long as their preferred quarterly data are not updated. The results are shown in Table 7.

The E-tests are unambiguously in favour of the Chow-Lin/LASSO imputations for all six reference series. The regression coefficients *g* all have the expected signs, and the t-statistics are high in absolute term, so that the associated p-values indicate significance at the 1%-level (highlighted **bold**). For the competing *locf* imputation, the *h*-coefficients are all considerably smaller and the t-statistics do not even come close to indicating statistical significance. In other words, the *external* validation confirms that Chow-Lin/LASSO imputations of *CapU* are *significantly* superior to the *locf* series in predicting the six external series, so that we can conclude that the temporally disaggregated time series of technical capacity utilisation created by the Chow-Lin/LASSO imputation is a valid reflection of the monthly dynamics of the Swiss economy.

4.3.1 Robustness of the Results During the COVID-19 Period

The last part of our analyses is to run the same six E-tests for the COVID-19 years 2020–2021. While the number of observations (24, of which 16 are imputed for *CapU*) is small for statistical inference, we can at least verify whether our conclusions based on the pre-COVID-19 results are not reverted. The results are shown in Table 8.

Looking at the COVID-19 years, the E-tests for two of the six external variables (unemployment and inflation) are unambiguously in favour of the Chow-Lin/

LASSO imputations at the 5%-significance level (highlighted in **bold**). For the KOF Economic Barometer, the Leading Global Barometer and the Swiss National Bank's BCI, the E-tests indicate that neither of the two competing imputations can be said to explain the variation in the external series at conventional significance levels. Lastly, for the Coincident Global Barometer, the E-test is favouring *locf*.

Accordingly, the results from the pre-COVID-19 period are confirmed for the time of the pandemic regarding the "hard" data only (official unemployment and inflation statistics), which is still remarkable, as with n = 16 statistical significance requires very clear results. On the other hand, the more sophisticated imputation fails to outperform the benchmark to achieve congruence with the "soft data" (composite sentiment indicators). The uncertainty in 2021 and 2022 apparently to some degree invalidated the composite sentiment indicators, while the auxiliary variables for the imputation of the missing *CapU* observations maintained much of their usefulness.

All in all, our results suggest that the monthly variation in the temporally disaggregated time series surveying technical capacity utilisation by means of the Chow-Lin/LASSO procedure is generally rather closely related to the monthly dynamics of the Swiss economy. The overwhelming evidence in favour of this conclusion that is reported in Table 6 for the longer period is, however, less apparent for the COVID-19 period, in which also many other economic indicators failed to perform as previously.

5 Summary and Conclusions

In this paper, we compare algorithms to deal with the problem of missing values in higher frequency data sets. To this end, we refer to the Swiss KOF business tendency surveys amongst manufacturing firms. They are conducted at both monthly and quarterly frequency, where an information sub-set is collected at quarterly frequency only.

In order to determine the best temporal disaggregation method, we design two verification procedures. The first procedure, referred to as *internal* validation, compares the performance of several univariate as well as multivariate temporal disaggregation techniques using eleven monthly time series from the KOF Swiss manufacturing survey. For each monthly time series in turn, we construct *artificially* quarterly series by skip-sampling two out of three monthly observations in each quarter. Then we use a selection of temporal disaggregation methods to recover the deleted data points. More specifically, we use the following methods: last observation carried forward as the benchmark, a cubic spline, linear regression, the Expectation–Maximisation algorithm, the standard Chow-Lin procedure and the Chow-Lin procedure augmented with the least absolute shrinkage and selection operator (LASSO) regression of Tibshirani (1996) based on approaches suggested by Wang et al. (2007). Since we know both the original and imputed values, we can run standard tests of forecasting accuracy to rank these methods by their imputation precision. We compare the accuracy of imputations both in terms of the root mean squared error as well as of directional accuracy.

As the result of the internal validation procedure, the temporal aggregation algorithm that combines the standard Chow-Lin method with the LASSO approach of Tibshirani (1996) is found to the best performing algorithm. It has a slight edge over the standard Chow-Lin method and much superior performance when compared to other competing methods. It is important to note that the combined Chow-Lin/LASSO method would also work in data-rich environments when the number of auxiliary monthly indicators may well exceed the number of data points in the low-frequency variable, i.e., in situations where the standard Chow-Lin method would be infeasible to implement.

The *internal* validation is amended by an *external* validation, where we evaluate the congruence of genuinely imputed monthly values from the quarterly survey question about firms' technical capacity utilisation in per cent with existing monthly time series that can be expected to be related to technical capacity utilisation: the KOF Economic Barometer, the KOF/FGV Leading and Coincident Global Barometers, the Swiss National Bank's Business Cycle Index, the Swiss unemployment rate and inflation. To this end, we run the forecast encompassing to determine whether the temporally disaggregated technical capacity utilisation time series by our preferred combined Chow-Lin/LASSO method can capture the monthly variation in the six external time series better than the capacity utilisation series temporally disaggregated with the benchmark algorithm last observation carry forward.

For the pre-COVID-19 decades going back to the 1990s, we find the forecast encompassing test statistics unambiguously in favour of our amended Chow-Lin/LASSO imputations for all six reference series. This is strong evidence that the imputed monthly variance of the technical capacity utilisation series is meaningfully related to the true monthly dynamics of the Swiss economy.

For the COVID-19 years 2020–2021, the results are not as unambiguous as those obtained for the longer pre-COVID-19 sample, which we attribute to the rather small sample size available for testing (in total 24 monthly observations, of which 16 are imputed) as well as the elevated degree of general uncertainty and turbulence in the underlying monthly indicators.

The rule that findings from simulated forecasting exercises are particular to the data used applies here, too. Having said this, we trust that the ranking of the algorithms will be similar for comparable applications, i.e. imputations of monthly values for quarterly time series relating to the business cycle. Also, the superiority of indirect imputations of balance indicators via the underlying shares compared to the direct approach may apply for qualitative business cycle survey data in general, but more research should be conducted before drawing too firm conclusions.

Appendix

Cubic Spline

The specification of the cubic spline procedure used in this paper imputes the missing monthly values with the "natural" cubic spline interpolation. A cubic spline is natural if the bending moment (the second derivative of the spline function) at the end points equals zero and the slope (the first derivative of the spline function) is constant. Consequently, the natural cubic spline ends in a straight line, as missing monthly values at the end of the series are extrapolated by continuing a straight line.

EM Algorithm

We adopt an approximate static factor model like the one presented by Stock and Watson (2002a) that allows modelling the co-movement of numerous variables in terms of a few latent factors. The approximate static factor model given a $T \times N$ matrix X of N time series assumes the following factor model representation:

$$X = FL + \epsilon \tag{A1}$$

where L is a $K \times N$ matrix of the factor loading coefficients, and F is a $T \times K$ matrix of K common factors. The idiosyncratic error term ϵ is variable-specific and has the corresponding dimension $T \times N$. The idiosyncratic disturbances can be serially and cross-sectionally correlated. The approximate static factor model relaxes restrictive assumptions of the classic factor analysis that requires cross-sectional and temporal independence of the idiosyncratic disturbances. Stock and Watson (2002a) showed that under fairly general conditions on the error terms the latent factors can be consistently estimated using the principal components (PC) analysis. Observe that in order to rule out scale effects, we perform the principal components extraction referring to the correlation matrix rather than to the covariance matrix of the selected indicator variables. This is mandatory, as the variances of our transformed indicators series differ greatly for purely technical reasons that should not affect the weight given to a particular variable.

For any arbitrary number of common factors K ($K < \min(N, T)$) estimates of L and F are obtained as a solution to the following nonlinear least squares minimisation problem:

$$\hat{L}, \hat{F} = \underset{F,L}{\operatorname{argmin}} \sum_{t=1}^{T} (X_t - F_t L)\prime(X_t - F_t L), \text{subject to } LL' = I_k \tag{A2}$$

The optimisation problem is solved by setting L equal to the eigenvectors corresponding to the K largest eigenvalues of the sample correlation matrix X. The estimator of the common factors is given by $\hat{F} = X\hat{L}'$.

Alternatively, the first principal component can be defined as the linear combination of variables with maximal variance. The subsequent principal components are similarly defined with an additional restriction that their loadings must be orthogonal to all previously calculated principal components. Formally,

$$\hat{L}_k = \underset{L_k}{\operatorname{argmax}}\ var\left(F_k\right), \text{subject to } L_k L_k' = 1 \text{ and } L_k L_j' = 0 \text{ for all } j < k. \tag{A3}$$

Factors are estimated as before by $\hat{F}_k = X\hat{L}_k'$, where F_k is the k-th column of F and L_k the k-th row of L. Hence principal components analysis has the following interpretation. The first principal component explains as much variation in the data as possible. The second explains as much of the remaining variation in the data PC as possible after extraction of the first, and so on. In this way principal component analyses reduces the dimensionality of a large set of interrelated variables, while retaining as far as possible the information (variation) present in the data set.

Our panel always contains one quarterly series whose non-quarter months are missing and $N-1$ monthly indicators. We employ the Expectation–Maximisation (EM) algorithm following Stock and Watson (2002b) to estimate the common factors and missing observations simultaneously.

The steps of this algorithm are:

- Fill the missing values in X with their initial estimates (mean imputation is commonly used, i.e. the missing values in variable X_i are filled with the mean of X_i).
- Repeat the following steps until convergence is reached:

1. Compute the factors $\hat{F}$ and factor loadings $\hat{L}$ (M-step).
2. Reconstruct X with $\hat{X} = \hat{F}\hat{L}$ (E-step).
3. If the absolute differences between the missing values from the first imputations in $\hat{X}$ and the corresponding values in X are below a certain threshold, stop.
4. Update the initial iterations in X with the new estimates in $\hat{X}$ and go to step 1.

In the above procedure, we limit ourselves to 2 principal components, because the quality of the imputation in terms of *RMSE* did not improve with a higher number of components.

Multiple Regression

The missing monthly values of a quarterly series are imputed with the help of the remaining ten available monthly indicators. For this purpose, the following multiple regression is estimated:

$$y^l = CX\beta + Cu \tag{A4}$$

y^l is the $T \times 1$ quarterly target series, X the $3T \times N$ matrix of monthly indicators, C a $N \times 3T$ conversion matrix that converts the monthly matrix X to a $T \times N$ quarterly matrix by extracting the quarter months, and $u \sim N(0, \Sigma)$ is a $3T \times 1$ vector of errors with $\Sigma = \sigma^2 I$. We drop the requirement of i.i.d. errors u_t and assume they follow an AR-process. The ordinary least squares solution is given by

$$\hat{\beta} = (X'C'CX)^{-1}X'C'y^l. \tag{A5}$$

If $N > T$, the matrix $X'C'CX$ is singular and no unique solution for $\hat{\beta}$ exists. When $N \leq T$ but N is large relative to T, ordinary least squares is prone to overfitting, such that the model fits the in-sample data and its noise well but fails to give good out-of-sample predictions. Since $N = 10$ and our in-sample window is $T = 80$ quarters long, overfitting might prove to be an issue.

We then estimate the $T \times 1$ monthly series y^h with $\hat{y}^h = X\hat{\beta}$ and replace the missing monthly values in y^h with $\hat{y}^h$ to arrive at the imputed monthly series.

Chow and Lin (1971) Procedure

The Chow and Lin (1971) methodology is a least-squares optimal solution for temporal disaggregation on the basis of a linear regression model. In that respect it formalises and generalises the ad-hoc ordinary least squares solution. Notwithstanding that is has been suggested close to half a century ago, it is arguably still the most popular imputation method when high frequency indicators are at hand.[14]

The Chow/Lin procedure seeks to exploit a statistical relationship between low frequency data and higher frequency indicator variables through a high frequency regression equation.

$$y^h = X\beta + u \tag{A6}$$

Where X are high frequency indicators and u is an error term with variance covariance matrix V. With C the conversion matrix which includes the distribution or interpolation restrictions the low frequency regression equation becomes.

$$y^l = Cy^h = CX\beta + Cu \tag{A7}$$

The regression coefficients are computed using the Generalised Least Squares (GLS) estimator.

$$\hat{\beta} = \left[X'C'(CVC')^{-1}CX\right]^{-1}X'C'(CVC')^{-1}y^l \tag{A8}$$

The high frequency values can then be calculated by.

$$\hat{y}^h = X\hat{\beta} + D(y^l - CX\hat{\beta}) \tag{A9}$$

Where D is the distribution matrix (distributing the low frequency residuals to the high frequency values).

$$D = VC'(CVC')^{-1} \tag{A10}$$

The part CVC' is the low frequency variance covariance matrix.

[14] See, amongst many others, Bagzibagli (2014), Čižmešija et al. (2018) and Stuart (2018),

In the context of this article the low frequency is quarterly, and the high frequency is monthly. The problem is an interpolation problem where the value of the quarter is identical to the value of the first month in the respective quarter. So, the conversion matric C is.

$$C = \begin{bmatrix} 1 & 0 & 0 & & & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & & 0 \\ & & & & \vdots & & & \\ 0 & \cdots & & & & 1 & 0 & 0 \end{bmatrix} \tag{A11}$$

This case was also discussed in the original paper of Chow and Lin (1971). The simplest case is to assume that the monthly regression residuals are serially uncorrelated. In that case $D = C'$ so that the distribution matrix assigns the quarterly residual fully to the first month of the quarter.

Chow and Lin assume for the high frequency residuals an AR(1) process, that is.

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{A12}$$

Where ϵ_t is white noise with variance σ_ϵ^2. In this case V is of the form

$$V = \frac{\sigma_\epsilon^2}{1-\rho} \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ & \vdots & \vdots & \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix} \tag{A13}$$

So, for the procedure one needs an estimate of ρ. For this, the first order autocorrelation of the quarterly (the low frequency) residuals is estimated. With the AR(1) assumption for the monthly residuals the first order autocorrelation of the quarterly residuals is ρ^3. The estimates are then plug into the distribution matrix D to estimate the high frequency values.

Combined Chow-Lin and LASSO Approach

We combine Chow and Lin (1971) with the penalised regression or least absolute shrinkage and selection operator (LASSO) approach suggested in Tibshirani (1996). It turns out that is it a straightforward way to do as all the necessary steps were already described in contribution of Wang et al. (2007) that shows how the penalised LASSO regression can be modified in the presence of the autocorrelated error term. The main motivation for extending the standard Chow-Lin procedure is that it becomes operational in data-rich environments with a very large number of high-frequency indicators available as potential candidates to be used in the temporal aggregation process. As noted in Tibshirani (1996), the LASSO combines both shrinkage and model selection aspects.

The LASSO method estimates the linear regression coefficients by minimizing the sum of least squares subject to an l_1 penalty function:

$$\min_{\beta} ||\left(y_t^l - X_t\beta\right)||_2^2 + \lambda||\beta||_2 \tag{A14}$$

The use of the l_1 penalty in conjunction with the squared objective function leads to many corner solutions for which the parameter estimates are zero. λ is a tuning parameter that captures the relative weight on the penalty function.

Wang et al. (2007) extend the LASSO for regressions with autoregressive errors by reformulating the LASSO criterion as follows.

$$\min_{\beta} ||\left(y_t^l - X_t\beta\right) - \sum_{j=1}^{q} \varphi_j\left(y_{t-j}^l - X_{t-j}\beta\right)||_2^2 + \lambda||\beta||_2 + \gamma||\varphi||_2 \tag{A15}$$

Here (λ, γ) are the shrinkage parameters. The authors allow for two different shrinkage parameters, one for the regression coefficients and one for the autoregressive coefficients. In our implementation, the problem is simplified. Instead of a general AR(q) process for the error term, a parsimonious first-order autoregressive dynamics is assumed, AR(1), for which no shrinkage is required. Penalty terms are computed only for the regression slope parameters β.

To solve the optimization problem Wang et al. (2007) propose an iterative algorithm. This algorithm simplifies in the present context to:

In step i of the iteration:

2. Use LASSO with fixed $\widehat{\phi}(i)$:

$$\hat{\beta}(i+1) = \min_{\beta} ||\left(y_t^l - X_t\beta\right) - \hat{\phi}(i)\left(y_{t-1}^l - X_{t-1}\beta\right)||_2^2 + \lambda||\beta||_2 \tag{A16}$$

3. Use $\widehat{\beta}(i+1)$ to estimate $\widehat{\phi}(i+1)$ with ordinary least squares.

Steps 1 and 2 are repeated until convergence. The initial values for iterations $\widehat{\beta}(0)$ are estimated with a ridge regression $\widehat{\phi}(0)$ with ordinary least squares. The shrinkage parameter λ is chosen according to BIC, as proposed by Wang et al. (2007).

After convergence, the estimates $\widehat{\beta}$ and $\widehat{\rho} = \widehat{\phi}^3$ are again plugged into the Chow and Lin estimation equation for $\widehat{y^h} = X\widehat{\beta} + D(y^l - CX\widehat{\beta})$ to obtain high frequency estimates.

Latest Version of KOF the Manufacturing Business Tendency Surveys

37280

KOF **Business tendency survey Industry**

KOF Swiss Economic Institute
ETH Zürich, LEE F 101, 8092 Zürich
http://www.kof.ethz.ch

Tel: 044 632 43 26
ind@kof.ethz.ch

Sector name:

clasification:

Survey INU

Company-ID

Contact-ID

Sector-ID

Please note

- Your responses should refer only to the branch named above
- The questions refer to the activities of domestic branches
- Do not use a red pencil
- Tick the appropriate box
- The notes are on the back of the sheet
- Your responses are treated strictly confidential.

Review and Assessment of the Current Situation

1. Incoming orders

a) In the past month compared to the previous* they have
○ increased ○ remained the same ○ declined

b) Compared to the same past month one year ago they were
○ higher ○ the same ○ lower

2. Order backlog ☐ no order backlog

a) In the past month compared to the previous* month orders have
○ increased ○ remained the same ○ declined

b) How would you assess the present order backlog* overall? As
○ large ○ normal ○ too low

c) How would you assess the present order backlog* for exports? As ☐ no exports
○ large ○ normal ○ too low

3. Production

a) In the past month compared to the previous* it has
○ increased ○ not changed ○ decreased

b) Compared to the same past month one year ago it was
○ higher ○ the same ○ lower

4. Intermediate products inventory ☐ no inventory

a) In the past month compared to the previous* it has been
○ higher ○ the same ○ lower

b) How would you assess the intermediate product inventory* ? As
○ too high ○ normal ○ too low

5. Finished products inventory ☐ no inventory

a) In the past month compared to the previous* it has
○ increased ○ the same ○ dropped

b) How would you assess the finished product inventory* ? As
○ too high ○ normal ○ too low

6. Employment levels

We would assess the current number of employees* as
○ too large ○ normal ○ too small

7. Business situation

a) How would you assess your current overall business situation*?
○ good ○ satisfactory ○ poor

b) In **the next 6 months*** our business situation will
○ improve ○ remain the same ○ get worse

c) To predict the **future development** of our business situation is currently
○ easy ○ rather easy ○ rather difficult ○ difficult

d) The **uncertainty** about the **future development** of our business situation is currently
○ higher than usual ○ normal/as usual ○ lower than usual

Expectations

8. It is likely that in the next 3 months

a) incoming orders will*
○ increase ○ remain the same ○ decrease

b) export orders will* ☐ no export
○ increase ○ remain the same ○ decrease

c) production will*
○ increase ○ remain the same ○ decrease

d) the purchase of intermediate products* will
○ increase ○ remain the same ○ decrease

e) the number of employees (FTEs) will*
○ increase ○ remain the same ○ decrease

f) our selling prices will*
○ increase ○ remain the same ○ decrease

g) our purchase prices will*
○ increase ○ remain the same ○ decrease

* Excluding seasonal fluctuations **Continue on the back page**

Comments

37280

Additional quarterly questions

9. Technical capacity

a) In the past 3 months* it

○ expanded ○ remained the same ○ reduced

b) We assess our current technical capacity* as

○ too large ○ adequate ○ too small

c) the average utilisation of capacity was in the past 3 months (in %)

<=50	55	60	65	70	75	80	85	90	95	100	105	>=110
○	○	○	○	○	○	○	○	○	○	○	○	○

10. Incoming orders

In the last 3 months* they

○ increased ○ remained the same ○ decreased

11. Production

In the last 3 months* it has

○ increased ○ remained the same ○ decreased

12. Finished products inventory

In the past 3 months* it

○ increased ○ remained the same ○ decreased

13. Sale prices

expressed in Swiss francs, in the past 3 months* they

○ increased ○ remained the same ○ decreased

14. Profitability

in the last 3 months* it

○ improved ○ remained the same ○ deteriorated

15. Level of backorders

We have currently sufficient production backorders for: ☐☐ months

16. Competitive position

a) In the past 3 months our domestic competitve* position has

○ improved ○ not changed ○ deteriorated

b) In the past 3 months our compeive position* in the EU has ☐ no EU exports

○ improved ○ not changed ○ deteriorated

c) In the past 3 months our competitve position* outside the EU has ☐ no exports outside EU

○ improved ○ not changed ○ deteriorated

17. Production obstacles

The main factors currently limiting our business are (multiple answers possible)

- no obstacles ☐
- insufficient demand ☐
- shortage of labor force ☐
- shortage of material/intermediate products ☐
- insufficient technical capacity ☐
- financial restrictions ☐
- other factors ☐

18. Wages and inflation

a) How much do you expect the **average gross wage** of employees in your company will change between now and in one year's time? Please enter your estimate as a **percentage** (with a negative sign if it is a decrease). ☐☐☐.☐

b) What do you expect the **inflation rate** (for the consumer price index) will approximately be in Switzerland in the **next twelve months**? Please enter your estimate (with a negative sign if the inflation rate is below zero). ☐☐☐.☐

c) Approximately how high do you think the annual **inflation rate** (for the consumer price index) will be in Switzerland in **five years**? Please enter your estimate (with a negative sign if the inflation rate is below zero). ☐☐☐.☐

19. Weighting information

Number of employees in **full-time equivalent** positions incl. apprentices in **Switzerland** (in the company or the company division entered in the questionnaire)
Example: 2 full-time positions and 1 part-time position at 40% correspon to a total of 2.4 employees

Number of employees: ☐☐☐☐☐.☐☐

* Excluding seasonal fluctuations

Many thanks for your participation

Explanations on the survey and the questionnaire can be found on the website:
https://u.ethz.ch/fi9dh

29578

Regarding the questions

1. Incoming orders
are considered to be orders from customers; internal orders should not be taken into account. Basically, the quantities ordered (primarily for standaridised products) should be used for reporting purposes. Whre this is not possible, the value of the orders can be used as the reporting basis (pure, price driven changes should be excluded).

2. Order backlog
This includes the quantity or the (price-adjusted) value of the customer orders still waiting to be worked on. The order backlog is to low, when the normal utilisation of capacity is not possible or endangered in the future. It is considered as too large, when the backlogged orders cannot be fulfilled within the desired (normal) period. When you regularly deliver abroad, please also answer question 2c. For this question, take into account orders, which are not directly exported but rather are channeled through external export firms.

3. Production
This is understood to include the quantity or the(price-adjusted) value of the intermediate and final products produced and possibly the sum of the work and machine hours used.

4. Intermediate products inventory
Here, only the stocks of commodities and goods in process obtained through third parties are to be considered. We are solely interested in quantitative changes. The stocks are considered to be too high or too low if ther usual - perhaps seasonal diverse - proportion to the planned production is significantly distorted in the respective direction.

5. Finished products inventory
What is meant is only inventory, which is not applied to an order or contract. Customer inventory or final products, which are kept in storage at your facillity for scheduling or technical reasons, are not counted as part of the finished product inventory. Inventory is considered too high when the current inventory levels are an indication of market stagnation and too low, when orders cannot be fulfilled from the inventory within the desired period.

6. Employment levels
This has to do with the average number of workers(converted to full-time equivalents or FTEs) in the corresponding product groups and possibly the number of work hours expended. For compamnies with only one survey, this corresponds to the overall employment trend. The assessment should be made with regard to the order backlog or the finished product inventory and expected incoming orders.

7. Business situation
This quation is intentionally vague. The overall economic condition of the company should be reported as the buisiness outlook. The respondent may decide, this assessment using revenues, income and number of empolyees or a combination of all of these factors.

Funding Open access funding provided by Swiss Federal Institute of Technology Zurich.

References

Abberger, K., Graff, M., Siliverstovs, B., & Sturm, J.-E. (2018). Using rule-based updating procedures to improve the performance of composite indicators. *Economic Modelling, 68*, 127–144.

Abberger, K., Graff, M., Müller, O., & Sturm, J.-E. (2022). Composite global indicators from survey data. The Global Economic Barometers. *Review of World Economics, 158*, 917–945.

Bagzibagli, K. (2014). Monetary transmission mechanism and time variation in the Euro area. *Empirical Economics, 47*, 781–823.

Boot, J. C. G., Feibes, W., & Lisman, J. H. C. (1967). Further methods of derivation of quarterly Figures from annual data. *Applied Statistics, 16*(1), 65–75.

Chow, G. C., & Lin, A.-L. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics, 53*(4), 372–375.

Čižmešija, M., Sorić, P., & Lolić, I. (2018). Consumer surveys and the EU statistics on income and living conditions: Friends or foes? *International Journal of Sustainable Economy, 10*(1), 78–98.

Davidson, R., & MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica, 49*(3), 781–893.

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association, 66*, 99–102.

Eurostat,. (2018). *ESS guidelines on temporal disaggregation, benchmarking and reconciliation*. Publications Office of the European Union.

Feijoó, S. R., Rodríguez Caro, A., & Quintana, D. D. (2003). Methods for quarterly disaggregation without indicators; A comparative study using simulation. *Computational Statistics & Data Analysis, 43*, 63–78.

Fernández, R. B. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics, 63*(3), 471–476.

Graff, M., & Sturm, J.-E. (2012). The information content of capacity utilisation rates for output gap estimates. *Cesifo Economic Studies, 58*(1), 220–251.

Harvey, A. C., & Pierce, R. G. (1984). Estimating missing observations in economic time series. *Journal of the American Statistical Association, 79*, 125–131.

Litterman, R. B. (1983). A random walk, Markov model for the distribution of time series. *Journal of Business & Economic Statistics, 1*(2), 169–173.

Mizon, G., & Richard, J.-F. (1986). The encompassing principle and its application to testing non-nested hypotheses. *Econometrica, 54*(3), 657–678.

Moauro, F., & Savio, G. (2005). Temporal disaggregation using multivariate structural time series models. *Econometrics Journal, Royal Economic Society, 8*(2), 214–234.

Mosley, L., Eckley, I., & Gibbert, A. (2022). Sparse temporal disaggregation. *Journal of the Royal Statistical Society: Series A (statistics in Society), 185*, 2203–2233.

Proietti, T. (2006). Temporal disaggregation by state space methods: Dynamic regression methods revisited. *Econometrics Journal, Royal Economic Society, 9*(3), 357–372.

Schumacher, C., & Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting, 24*(3), 386–398.

Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association, 97*(460), 1167–1179.

Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics, 20*(2), 147–162.

Stuart, R. (2018). A quarterly Phillips curve for Switzerland using interpolated data, 1963–2016. *Economic Modelling, 70*, 78–86.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B, 58*(1), 267–288.

Wang, H., Li, G., & Tsai, C.-L. (2007). Regression coefficients and autoregressive order shrinkage and selection via the lasso. *Journal of Royal Statistical Society, Series B, 69*, 63–78.