CESifo WORKING PAPERS

10191
2022

December 2022

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

Klaus Abberger, Michael Graff, Oliver Müller, Boriss Siliverstovs

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10191*

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

Abstract

This paper documents a comparative application of algorithms to deal with the problem of missing values in higher frequency data sets. We refer to Swiss business tendency survey (BTS) data which are conducted in both monthly and quarterly frequency, where an information sub-set is collected at quarterly frequency only. This occurs in many countries, for example, the harmonised survey programme of the European Union also has this frequency pattern. There is a wide range of ways to address this problem, comprising univariate and multivariate approaches. To evaluate the suitability of the different approaches, we apply them to series that are artificially quarterly, i.e., de facto monthly, from which we create quarterly data by deleting two out of three data points from each quarter. The target series for imputation of missing (deleted) observations comprise the set of time series from the monthly KOF manufacturing BTS survey. At the same time, theses series are ideal to deliver higher frequency information for multivariate imputation algorithms, as they share a common theme, the Swiss business cycle. With this set of indicators, we conduct the different imputations. On this basis, we then run standard tests of forecasting accuracy by comparing the imputed monthly series to the original monthly series. Finally, we take a look at the congruence of the imputed monthly series from the quarterly survey question on firms' technical capacities with existing monthly data on the Swiss economy. The results show that for our data corpus, algorithms based on the approach suggested by Chow and Lin deliver the most precise imputations, followed by multiple OLS regressions.

JEL-Codes: C190, C220, C530.

Keywords: temporal disaggregation, business tendency surveys, out-of-sample validation, mixed-frequency data.

Klaus Abberger
ETH Zürich, KOF Swiss Economic Institute
Zurich / Switzerland
abberger@kof.ethz.ch

*Michael Graff**
KOF Swiss Economic Institute
Zurich / Switzerland
graff@kof.ethz.ch

Oliver Müller
ETH Zürich, KOF Swiss Economic Institute
Zurich / Switzerland
omueller@kof.ethz.ch

Boriss Siliverstovs
Latvijas Banka
Riga / Latvija
Boriss.Siliverstovs@bank.lv

*corresponding author

1. Introduction

Analyses of the present economic situation are all too often plagued by missing data for the current or even recent periods. Moreover, many economic indicators are available as time series with less than desirable frequencies. As economic activity is a continuous process, there is in principle no limit to the desirable frequency of the data to draw on. In practice, financial data are frequently of daily or even higher frequency, but amongst macroeconomic data such frequencies are mostly unavailable. For macroeconomic assessments, typical series are annual and quarterly (relating to national accounting), or monthly (indices of production, inflation, unemployment statistics and business tendency data, amongst others), and publication lags are considerable and vary greatly. Given this, observers of the economic situation are faced with two related problems: mixed frequency and ragged edge.

Temporal disaggregation denotes the process of imputing high frequency from low frequency data, either from one and the same series or including information from higher frequency series that are informationally related to the target series. Temporal disaggregation can in principle address both mixed frequency and ragged edge, where the former problem calls for imputations between known data points in the past, while the latter amounts to forecasting higher frequency data points at the right margin.

This paper documents an applied investigation into strategies and algorithms to deal with this in the case of Swiss business tendency survey (BTS) data. BTS data are generally some of the earliest available macroeconomic indicators; and they are commonly collected in both monthly and quarterly frequency, where the lower frequency aims at limiting the burden for the respondents. As a result, for Switzerland and elsewhere, some of the information is available at quarterly frequency only.

Our data are mainly taken from the monthly KOF Swiss Economic Institute's BTS in the Swiss manufacturing sector.[1] Presently more than 1'000 firms are surveyed. The response rate is about 70 per cent. The latest questionnaire comprises 21 items. Due to changes to the questionnaire in the past, we have access to 11 complete series going back more than 50 years.[2]

All underlying questions are qualitative, with three options to answer: down/too low (−), no change/about right (=), up/too high (+). For quantification, we resort to the traditional and simple but robust net balance indicator (percentage share + minus percentage share −). The balance should go up and down over the business cycle, and while the mean may not be exactly constant in the long run, the quantification has an upper bound of 100 per cent and a lower bound of −100, so these series cannot increase or decrease forever. We thus refrain from testing for stationarity, as test results pointing to nonstationary processes would be misleading due to specific short-term outcomes of processes that in the end must be mean stationary.

The last observations at the time of finalising this paper relate to December 2021 or the fourth quarter of 2021, respectively. Obviously, the Covid-19 pandemic and its drastic impact on economies around the world pose a challenge to economic time series analyses unprecedented in the absence of wars, revolutions, or similar shocking disruptions. We therefore restrict the in-sample analysis to the period

[1] For the latest version of the questionnaire, 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.

[2] The 10 questionnaire items with shorter time series and 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) and 8b (expected export orders) and 8e to 8g (expectations regarding number of employees, selling prices and purchasing prices).

1967m2 to 2019m12. This notwithstanding, at the end of the paper, we will take a look at how well our imputations would have fared in real-time during the pandemic in 2020 and 2021.

Table 1.1 lists the 11 KOF monthly manufacturing BTS questions that form the basis of our analyses.

Table 1.1: Questions from the monthly KOF manufacturing BTS going back to 1967m2

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

Apart from this, the KOF quarterly manufacturing BTS, which is conducted during the first month of a quarter, address additional issues like bottlenecks, and potentially most interesting, as a micro foundation of the macroeconomic output gap, numerical estimates of firms' technical capacity utilisation.

Facing situations like this, applied business cycle researchers have explored ways to transform quarterly into monthly time series. There is a wide range of ways to address the frequency imputation task.

A first condition to be met is that the process that is only observed (measured) at the lower frequency is in fact occurring at the higher (target) frequency. This is evident for continuous processes like the production of goods and services (GDP) 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 X-mas, for example, occur exactly once per year, and it would not make sense to construct quarterly or monthly Christmas sales series, although this may be possible in technical terms. Hence, before series are 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 missing data points is solely due to the lower measurement frequency. In our case, the processes (relating to the respondents' firms' situation) are clearly continuous, whilst measurement comes at discrete intervals, so that temporal disaggregation is justifiable.

Another time series distinction to be aware of is between stock and flow variables. Temporal disaggregation of flows is usually performed in ways ensuring that the higher frequency flows add up to the lower frequency values. For stock variables, the consideration is not so obvious. Should the higher frequency values' average correspond to the lower frequency one, or should the match rather correspond to the first or last the of the higher frequency values? This will depend on the time series process, so that a general answer cannot be given. As we are dealing with BTS items reflecting the assessment of continuous phenomena, given at a particular point in time, we do not find it mandatory to restrict the joint values of the monthly breakdowns to a corresponding value from the quarterly series.

With these considerations taken care of, a first procedural distinction is between univariate and multivariate approaches. The former ranges from equal distribution of quarterly outcomes to a monthly subdivision to interpolations that implicitly assume a smooth evolution over the months of a quarter. Nonparametric approaches which ground on such smooth evolution assumptions are for example spline methods or the so-called Denton method, which is a classical approach from temporal disaggregation. At the right margin, univariate methods either hold the last observation fixed (last observation carried forward) or try to exploit the momentum of the series either by simple extrapolation or by ARIMA or univariate state-space models.

Yet, even if some of these approaches are technically quite sophisticated, the inherent shortcoming lies in the fact that missing observations are generated 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 valid data are particularly needed.

In addition to this, univariate imputations are informationally inefficient when other data that could supply valid information on the true movement of the partially unobserved process are available but disregarded. Specifically, if there are variables that are directly or partially correlated with the lower frequency series of interest, this information can be exploited. To this end, the (sets of) time series have to be identified and thence pre-selections to be carried out regarding *meaningful* correlations with the series of interest, where *meaningful* refers to the expected signs and leads/lags of the associations to eliminate spurious correlations.

We thus resort to a range of univariate and multivariate imputation algorithms. Obviously, the results from the monthly imputations will differ to some degree. Also, it is an established fact that no single imputation or forecasting algorithm beats all others under all circumstances. And with the "true" values of the imputed ones unknown, how shall we compare and evaluate the results?[3]

A first step is to compare the statistical properties of the monthly series (for example the moments of the distribution) and perform visual checks of time series plots to assess the plausibility of the outcomes.

Yet, with no monthly reference series at hand, the definitive yardstick for the identification of the most appropriate transformation is missing. We therefore resort to the following strategy to compare the suitability of the different approaches for our survey data: Instead of attempting to evaluate the monthly imputations from genuinely quarterly data, we make sure that we *do possess* adequate reference series for the model selection stage. To this end, we create artificially quarterly data by deleting the two out of three data points from each quarter of our monthly BTS data. This essentially replicates the situation that one would face if the survey was quarterly rather than monthly in the first place.

Related to this is the problem that for multivariate approaches there is no general procedure to determine which indicators to consider. For our purposes, however, the answer is straight-forward: The candidate series are the ones listed in Table 1.1. We thus resort to a set 11 of indicators that share a common theme, which is the reflection of the Swiss business cycle. They result from the same data generating process. Finally, they have the same publication dates, so the ragged edge problem does not occur.

In particular, we first construct the 11 *artificially* quarterly series. We then generate imputed monthly values for the (seemingly) missing ones, where for each series the other 10 *monthly* series form the pool of potential high frequency indicator variables. We then conduct the different imputations. On this basis,

[3] For a detailed discussion, see Eurostat (2018).

we can then run standard tests of forecasting accuracy by comparing the imputed monthly values to the original ones. We refer to this as *internal* validation. It will deliver the first pieced of information, the algorithms recommended to apply re-establish erased monthly values within out monthly BTS data set.

We will extend this 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. There are not many such series for the Swiss economy, but we can resort to six important ones: the KOF Economic Barometer, the KOF/FGV Leading and Coincident Global Barometers, the Swiss National Bank's Business Cycle Index, the Unemployment Rate and Inflation.

The remainder of our paper is structured as follows. The next Section will look in more detail at some related work on the topic of temporal disaggregation. Section 3 will present the imputation methods to be evaluated, their implementations and the outcomes. Section 4 will look at the different outcomes in terms of descriptive statistics as well as statistically significantly differences. Section 5 summarises, concludes, makes some careful on recommendations for temporal disaggregation of series like ours and identifies promising lines for further investigations into this topic.

2. Review of the literature

A common difficulty faced by practitioners is missing high-frequency data, or transformation of low-frequency time series into a high-frequency time series, respectively.[4]

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 the 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, univariate methods either hold the 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).

Practitioners can refer to various statistical packages to perform some of the most common univariate und multivariate implementations of temporal disaggregation.[5]

Feijoó et al. (2003) evaluate the performance of several most 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 high-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:

[4] A recent example relating to Swiss data is the annual to quarterly imputation for GDP data from the 1960s and 1970s performed by Stuart (2018).
[5] See for example Sax and Steiner (2013).

$$y_t^h = \alpha + \beta x_t + \varepsilon_t ,$$

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 the three aforementioned temporal disaggregation approaches. The main advantage of formulating the temporal disaggregation procedures in a general state-space modelling framework is that several previously unattended inferential issues can be addressed in a straightforward manner. 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 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 model over-fitting becomes likely. Hence, some kind of variable selection procedure is required that retains only the most informative high-frequency variables regarding the dynamics of the low-frequency variable of interest. In this respect, imputation procedures based on regression are commonly applied. In these procedures, the missing values of the low-frequency variables are substituted with the predicted values from regressions of the original low-frequency variables on a selection of high-frequency variables. Several variable selection procedures, like forward, backward or stepwise, are common. Alternatively, penalised regressions can be used for variable selection. An example of the latter is Tibshirani's (1996) Lasso (Least Absolute Shrinkage and Selection Operator).

An approach to impute values in larger data panels involving variables sampled at heterogeneous frequencies is proposed by Stock and Watson (2002b). This is based on the Expectation-Maximization (EM) algorithm, which typically consist of stepwise 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 composite leading economic indicator (KOF Economic Barometer) for the Swiss economy (see Abberger et al., 2018) and since January 2020 for the Global Barometers (see Abberger et al., 2022).

Labonne and Weale (2020) also conduct temporal disaggregation, but for a different data environment. They derive monthly estimates of business sector output in the UK from rolling quarterly value-added tax (VAT) based turnover data. The VAT data exhibit substantial noise. When observed data are noisy, the authors suggest that the interpolants should not be restricted to the data themselves, but to the underlying "clean" signal. Therefore, the authors choose an unobserved components approach, where the VAT observations are modelled as the sum of latent components, one of which is an observation error. Zult et al. (2020) also analyze VAT data, but for their purposes, from the Netherlands. They also face the problem of "noisy" VAT data, as firms declare VAT at different frequencies. The majority declares VAT at quarterly rhythm (a minority monthly or annually), and the aim is to dispose of monthly figures. The authors use classical temporal disaggregation methods (Chow/Lin, Denton) within gaps of available quarterly data. They augment these with nowcasting methods (ARIMA, bridge models, structural time series models) for the most recent months, where quarterly data are not yet available. Quenneville et al. (2013) deal with the problem of calendarization. Calendarization is the process of transforming the values of a flow time series that is observed over varying time intervals into values that cover calendar intervals such as day, week, month, quarter and year. This problem involves temporal distribution of the reported values into, say, daily values and aggregation of the resulting daily interpolations into the desired frequency (monthly, quarterly or annual). An application at Statistics Canada deals with their monthly business surveys. They have collection agreements with some respondents that allow them to report periods other than the standard calendar month. The authors propose an innovative spline approach to address the calendarization problem. Although these papers fall broadly within the same field as ours, the objectives and the data situation are very different, so that we do not derive any practical conclusions from them at this stage.

Mosley et al. (2021), similarly to what we undertake in this paper, adapt the Chow/Lin procedure to a data-rich environment. After demonstrating that the traditional Chow/Lin procedure has shortcomings when used with a large number of high frequency indicators, they introduce an approach, which they call "generalized regularized M-estimation framework for temporal disaggregation". In particular, they add a penalty term ("regularizer") to the Chow/Lin cost function. The approach used in our paper can be seen as a special case of this general approach. Also, our data situation is special since the quarterly data are observed in the first month of each quarter.

3. Methods and results

In this section, we present the methods evaluated in this paper, their implementation, and some descriptive statistics. We first evaluate two univariate imputations of monthly data points, the last observation carried forward procedure and the cubic spline. As the former 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.

We then evaluate the following multivariate imputations: multiple regression, the EM algorithm, the Chow/Lin estimator and, as our theoretically preferred algorithm, a combined implementation of Chow/Lin with Lasso and first-order autoregressive term (AR1). The traditional Chow/Lin method does not directly address the problem of variable selection. However, in a more and more data-rich environment, data selection is becoming increasingly important. We address this with a new approach combining Lasso regression with the explicit error structure of the Chow/Lin approach. This combination allows the Chow/Lin approach to be applied to a large number of high frequency indicators.

The data corpus and the steps of computations are the same for all methods. We refer to seasonally adjusted data only, as otherwise we would have to impute seasonality, which unnecessarily complicates the task. First, we compute artificially quarterly series Q from the 11 monthly KOF manufacturing BTS questions M presented above, where we delete two out of three data points in a row. Numbering the months of a year from 1 to 12 we arrive at a theoretical maximum of three quarterly series from each monthly series with the following monthly origins:

1-4-7-10, 2-5-8-11 and 3-6-9-12.

However, for our imputation problem, due to the chronology of the data generating process in our surveys, we can disregard the second and third options, as for our monthly and quarterly surveys, the release and reference period coincide. Monthly results for M^i are released before end of month i and quarterly results are released for Q^j before end of the first month of quarter j; accordingly, the first month of each quarter delivers two observations (*M* and *Q*). Imputations of missing monthly values are thus needed for the second and third months of a quarter. The series of deleted data points to be imputed hence has the following monthly sequence:

2-3-5-6-8-9-11-12.

As this sequence reflects the information that our imputation methods have to recover, we will refer to the 11 series of this type as “out of sample”. They are the same for all imputation approaches to be evaluated. The monthly BTS series from 1967m2 to 2021m12 that we will look at in this paper cover 659 monthly observations, the quarterly series *Q* derived from them cover 219 data points, 440 data points can hence be imputed per series, amounting to a total of 4,840 data pints for the 11 series.

The following Figure illustrates the data generating processes of the KOF BTS and the terminology that we apply in this paper, assuming the perspective of the third month of the third quarter (m9). “Symmetric” relates to imputations for months lying between observed values in the past, whereas “asymmetric” imputations are “nowcasts” for months two and three of the current quarter, i.e. technically forecasts.

Figure 1: Timeline and terminology

									Observer
	Ex post						real time		
	q1			q2			q3		
monthly series	m1	m2	m3	m4	m5	m6	m7	m8	m9
quarterly series	new			new			new		
last obs. c. forw.	new	as m1	as m1	new	as m4	as m4	new	as m7	as m7
interpolation	new	symmetric	symmetric	new	symmetric	symmetric	new		
spline	new	symmetric	symmetric	new	symmetric	symmetric	new	asymmetric	asymmetric
multivariate	new	symmetric	symmetric	new	symmetric	symmetric	new	forecast	forecast

simulated real time								
	out of sample			out of sample			out of sample	
reference			reference			reference		
in sample								

The targets for our simulations are the seasonally adjusted balance indicators. For this, we can either directly impute the missing observation of this series. Alternatively, we can impute the seasonally adjusted positive and negative shares and compute the imputed target values indirectly from the difference of the two. *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 can result from greatly different positive and negative shares. Accordingly, we let the data speak and proceed along both paths.[6]

We compute simulated pseudo real-time (simulated real-time) values based on a rolling window of 20 years up to 2019. Stored values of monthly balance indicator series in the KOF data base start at 1967m2. However, the data base records negative and positive shares from 1971m4 only, so that including these, which is required for the indirect approach, the first rolling 20 years window ends in 1991m3. Accordingly, out-of-sample imputations can be calculated starting 1991m3, yielding 231 imputations per survey item and series and a total of 2,541 imputations for our 11 series.

[6] Notice that the target series, against which the imputations for the two approaches have to be evaluated, are not exactly the same. The direct target series is the seasonally adjusted balance from the differences of the unadjusted positive and negative shares, whereas the indirect target series is the difference of the seasonally adjusted positive and negative shares.

An additional data restriction applies for comparisons of real-time and ex-post imputations, given that our imputations are performed in 20 years rolling windows. This implies that consistent ex-post series cover exactly 20 years. The vintage that ends before the Covid-19 pandemic hit the Swiss economy thus ranges from 2000m1 to 2019m12, comprising 160 imputations per survey item and series and a total of 1,760 imputations for our 11 series.

Finally, to shed light on the performance of our imputations in 2020 and 2021, when the economy operated in crises modus and BTS data may not reflect the same as under normal conditions, our Covid-19 period ranges from 2020m1 to 2021q12, the ex-post imputations are taken from the 2002–2021 vintage. The number of imputations for these two years is 8 per series, amounting to 88 for the 11 indicators.

The observation periods and their characteristics are summarised in Table 3.1.

Table 3.1: Observation periods

Range	Imputations	Evaluations
1967m2–2021m12	4,840	Entire data base
1991m3–2019m12	2,541	Real-time, direct & indirect imputations
2001m1–2019m12	1,760	Real- time & ex-post, direct & indirect imputations
2020m1–2021m12	88	Sensitivity check for Covid-19 pandemic

For descriptive statistics and comparisons based thereon, we first look at the root mean squared error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

.

We can discriminate between *RMSE* statistics referring to asymmetric estimations (real-time out-of-sample total, real-time out-of-sample second months and real-time out-of-sample third months) as well as to the corresponding *RMSE* referring to symmetric ex-post imputations.

The *RMSE* measures the overall precision of the imputations, penalising larger deviations. To evaluate the precision of the different measures against our benchmark (last observation carried forward), we compute the relative *RMSE* (*RRMSE*), defined the ratio of the *RMSE* imputation method under consideration against the *RMSE* of the benchmark. Ratios above one indicate superiority of the method under consideration, ratios below one indicate that the benchmark is superior.

Our second descriptive statistics addresses turning points. In particular, quarterly series as well as imputed monthly series with the last observation carried forward 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. Our monthly imputations provide this information, and with the actual values at hand for comparison, it is straightforward to compute the percentage of corresponding up and down direction changes of the first and second imputed values within each quarter or for both taken together. For these statistics, values above 50% indicate that the imputations are reflecting direction changes better than simple guessing.

Notice that with chronological distance of the imputed value from the last observation increasing, asymmetric imputations for the third month of a quarter can be expected to be inferior to those of the second.

3.1 Last observation carried forward

Linear interpolation and carrying the last observed value forward until new information becomes available are the simplest imputation methods. The former, however, is impossible when real-time values have to be imputed at the end of a series, in our case for the second and third months of the current quarter. The last observation carried forward procedure (henceforth: *locf*) copies the latest available observation of the same series, in our case the first month of the quarter (see Figure 1). Notice that with chronological distance of the imputed value from the last observation increasing, *locf* imputations for the third month of a quarter should be expected to be particularly inferior to those of the second month, as no new information is taken into account.

The *RMSE* for the 11 real-time out-of-sample imputations, broken down by total, first month out-of-sample (m2) and second month out-of-sample (m3), are shown in Table 3.1.

Table 3.1: Descriptive statistics for the last value carried forward (*locf*) imputations

Period	1991–2019		2000–2019		2020–2021	
Imputation	direct	indirect	direct	indirect	direct	indirect
RMSE all	6.98	6.78	7.26	7.05	10.25	9.94
RMSE m2	6.93	6.72	7.11	6.91	9.03	8.76
RMSE m3	7.02	6.85	7.40	7.19	11.34	11.00

Table 3.1 shows that the *RMSE* for the *locf* imputations in the pre-Covid-19 periods range from 6.7 to 7.4. For the longer period from 1991 as well as for the shorter period from 2000, two regularities stand out: (1) the *RMSE* are larger for m3 compared to m2, and (2) larger for the direct comparted to the indirect imputations. This holds without exception, and also for the two Covid-19 years. For the latter, however, the errors are remarkably larger (9.0–11.3) than for the pre-Covid-19 years. Turning points are not computed, as *locf* imputations do not have direction changes by construction.

What we can take from this table alone is that the imputation errors are higher on average for the third months of a quarter compared to the seconds months, which is in line with expectations, as the lag to the last recorded observation increases. The finding that the indirect imputations via imputed positive and negative shares turn out to be slightly but consistently more precise than the direct imputations of the balance can only be attributed to the seasonal adjustment procedure.

3.2 Cubic spline

With our specification of the cubic spline procedure, missing monthly values of a quarterly series are imputed with natural cubic spline interpolation. A cubic spline is natural if the bending moment (the second derivative of the spline function) at the end points is equal zero and the slope (the first derivative of the spline function) is constant. Consequently, natural cubic spline ends in a straight line at the end points. Missing monthly values at the end of the series are extrapolated by continuing the straight line at the end point. The descriptive statistics are shown in Table 3.2.

Table 3.2 shows that the real time *RMSE* for the cubic spline imputations in the pre-Covid-19 periods range from 7.7 to 10.4, which is considerably higher than for *locf* (6.7–7.4). The *RMSE* for the Covid-

19 years 2020–2021 (19.7–23.5) are again much higher than for the preceding periods, about twice as high as for *locf* (9.0–11.3). Regarding the turning points, the percentage of hits in real time is close to 50% (and even below this for the Covid-19 period), so that random guessing would have about the same expected outcome.

Table 3.2: Descriptive statistics for the cubic spline imputation

Period	1991–2019		2000–2019		2020–2021	
Imputation	direct	indirect	direct	indirect	direct	indirect
real time						
RMSE all	8.91	8.61	9.41	9.11	19.07	18.74
RMSE m2	7.98	7.71	8.30	8.04	13.16	12.92
RMSE m3	9.75	9.42	10.40	10.07	23.54	23.14
Direction change	50.5%	51.6%	49.4%	51.5%	47.7%	46.0%
ex post						
RMSE all	n/a	n/a	6.87	6.66	8.98	8.75
RMSE m2	n/a	n/a	6.73	6.54	9.11	8.92
RMSE m3	n/a	n/a	7.00	6.78	8.85	8.57
Direction change	n/a	n/a	60.9%	60.6%	61.9%	60.2%

The ex-post performance is somewhat better than real-time in terms of *RMSE* for 2000–2019, and considerably better for 2020–2021. Also, the ex-post direction changes record about 60% hits.

The two regularities that were observed for *locf* also hold for the cubic spline imputations, and this both in real time and ex post: the *RMSE* are larger for m3 compared to m2, and larger for the direct comparted to the indirect imputations.

What we can take from this is that our implementation of the cubic spline imputation is clearly inferior to *locf* in real time, the variance created by the imputations is misleading, and this hold especially for the dramatic Covid-19 years. Ex post, the *RMSE* still do not recommend the cubic spline as an alternative to *locf*, but the direction changes are captured a bit better.

3.3 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:

$$X = FL + \epsilon$$

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:

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

The optimisation problem is solved by setting L equal to the eigenvectors corresponding to the K largest eigenvalues of the sample correlation matrix of 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,

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

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.

The *RMSE* for the real-time out-of-sample imputations, broken down by total, first month out-of-sample and second month out-of-sample, are shown in Table 3.3.

Table 3.3: Descriptive statistics for the EM algorithm

Period	1991–2019		2000–2019		2020–2021	
Imputation	direct	indirect	direct	indirect	direct	indirect
real time						
RMSE all	8.50	8.57	8.76	8.70	11.78	11.40
RMSE m2	8.35	8.42	8.60	8.57	11.94	11.77
RMSE m3	8.65	8.72	8.91	8.83	11.62	11.02
Direction change	64.3%	65.6%	63.5%	65.9%	64.2%	65.9%
ex post						
RMSE all	n/a	n/a	8.35	8.23	11.59	11.14
RMSE m2	n/a	n/a	8.04	7.93	11.83	11.61
RMSE m3	n/a	n/a	8.64	8.52	11.34	10.65
Direction change	n/a	n/a	62.8%	63.7%	66.5%	66.5%

Table 3.3 shows that the real time *RMSE* for the EM algorithm imputations in the pre-Covid-19 periods range from 8.4 to 8.9, which is considerably higher than for *locf* (6.7–7.4). The *RMSE* for the Covid-19 years 2020–2021 (11.0–11.9) are again much higher than for the preceding periods and somewhat higher than for *locf* (9.0–11.3). For direction changes, the percentage of hits in real time is about 65%, irrespectively of the period, so that in this respect, the EM algorithm performs better than random guessing.

The ex-post performance is only slightly better than real-time in terms of *RMSE* for 2000–2019, and the same holds for 2020–2021. Ex-post imputed direction changes are not better than in real time, with the exception of the Covid-19 period.

Of the two regularities that were observed for *locf*, only one also holds for the EM algorithm imputations and this only for the pre-Covid-19 periods (*RMSE* larger for m3 compared to m2). The second regularity (*RMSE* larger for direct than indirect imputations) is reverted for the 1991–2019 real-time imputations.

What we can take from this is that the EM algorithm imputation does not hold the promise to outperform the naïve *locf*, and this applies to all periods considered as well as to real time and ex post. The only improvement is for directions changes, where this approach comes close to two thirds hit ratio.

3.4 OLS regression

The missing monthly values of a quarterly series are imputed with the help of the remaining 10 available monthly indicators. For this purpose, the following multiple regression is estimated:

$$y^l = CX\beta + Cu$$

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 $\Sigma = \sigma^2 I$. In section 3.5 we drop the requirement of i.i.d. errors u_t and assume they follow an AR-process. The ordinary least squares (OLS) solution is given by

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

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, OLS 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.

The *RMSE* for the real-time out-of-sample imputations, broken down by total, first month out-of-sample and second month out-of-sample, are shown in Table 3.4.

Table 3.4: Descriptive statistics for the OLS regression imputation

Period	1991–2019		2000–2019		2020–2021	
Imputation	direct	indirect	direct	indirect	direct	indirect
real time						
RMSE all	4.79	4.75	4.94	4.92	6.02	6.01
RMSE m2	4.88	4.77	5.01	4.91	6.10	6.35
RMSE m3	4.71	4.73	4.87	4.92	5.95	5.65
Direction change	74.8%	73.7%	75.3%	74.2%	72.7%	71.6%
ex post						
RMSE all	n/a	n/a	4.62	4.72	5.84	5.76
RMSE m2	n/a	n/a	4.65	4.71	6.06	6.41
RMSE m3	n/a	n/a	4.59	4.73	5.62	5.02
Direction change	n/a	n/a	74.4%	73.9%	75.0%	72.7%

Table 3.4 shows that the real time *RMSE* for the OLS regression in the pre-Covid-19 periods range from 4.7 to 5.9, which is clearly lower than for *locf* (6.7–7.4). The *RMSE* for the Covid-19 years 2020–2021 (5.7–6.4) are somewhat higher than for the preceding periods, but again clearly lower than for *locf* (9.0–11.3). For direction changes, the percentage of hits in real time comes up to 75%, so that in this respect, OLS performs much better than random guessing.

The ex-post performance is only slightly better than real-time in terms of *RMSE* for 2000–2019, and the same holds for 2020–2021. Ex-post imputed direction changes are not better imputed than in real time for 2020–2021, but they are for Covid-19 years.

The two regularities that were observed for *locf*, largely hold for OLS, too, with a few exceptions: *RMSE* are *not* larger for m3 compared to m2 for the direct real time imputations as well as direct ex-post for 2000–2019. Also *RMSE* are *not* larger for any of the pre-Covid-19 direct real-time imputations compared to indirect, and the same holds for ex-post 2000–2019.

What we can take from this is that for our data the OLS regression imputations consistently outperform the naïve *locf*. This applies to all periods considered as well as to real time and ex post.

3.5 Chow and Lin

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 above-described ad-hoc OLS 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.[7]

The main idea of the Chow/Lin approach is that indicator and target variable satisfy a regression model that is valid both for high and low frequency, with the exception of the error structure. From the available low frequency, the procedure derives the estimates of the parameters of the regression model. These parameters are then applied to the high frequency model to derive the high frequency figures, including the extrapolation for the periods after the last low frequency value.

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

$$y^h = X\beta + u$$

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 is similar to the OLS expression

$$y^l = Cy^h = CX\beta + Cu$$

In contrast to the OLS solution the regression coefficients are calculated using the Generalised Least Squares (GLS) estimator

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

The high frequency values can then be calculated by

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

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

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

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

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

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

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

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

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

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

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

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 auto-correlation of the quarterly residuals is ρ^3. The estimates are then plug into the distribution matrix D to estimate the high frequency values.

This approach is used to address the problem presented in this article. Quarterly data where we observe the quarterly value in the first month of the quarter are interpolated with other monthly indicators.

The *RMSE* for the real-time out-of-sample imputations, broken down by total, first month out-of-sample and second month out-of-sample, are shown in Table 3.5

Table 3.5 shows that the real-time *RMSE* for the Chow/Lin algorithm in the pre-Covid-19 periods range from 3.8 to 4.0, which is remarkably lower than for *locf* (6.7–7.4). The *RMSE* for the Covid-19 years 2020–2021 (5.1–5.8) are somewhat higher than for the preceding periods, but again remarkably lower than for *locf* (9.0–11.3). For direction changes, the percentage of hits in real time comes up to 77% before Covid-19 and up to 72% for 2020–2021 so that in this respect, Chow/Lin also performs much better than random guessing.

The ex-post performance is consistently better than in real time, and this holds for *RMSE* as well as direction changes.

What we take from this is that for our data – comparable to OLS – the Chow/Lin imputations consistently outperform the naïve *locf*. This applies to all periods considered as well as to real time and ex post.

Table 3.5: Descriptive statistics for the Chow/Lin algorithm

Period	1991–2019		2000–2019		2020–2021	
Imputation	direct	indirect	direct	indirect	direct	indirect
real time						
RMSE all	3.93	3.80	3.99	3.91	5.64	5.19
RMSE m2	4.00	3.84	3.95	3.83	5.49	5.14
RMSE m3	3.85	3.76	4.02	3.98	5.79	5.24
Direction change	75.9%	75.0%	77.1%	76.0%	70.5%	72.2%
ex post						
RMSE all	n/a	n/a	3.79	3.74	5.31	4.68
RMSE m2	n/a	n/a	3.74	3.64	5.49	5.16
RMSE m3	n/a	n/a	3.84	3.84	5.13	4.15
Direction change	n/a	n/a	77.7%	76.1%	75.6%	72.7%

3.6 Chow and Lin with Lasso

Our last approach is a combination of Chow/Lin and Lasso, which we refer to based on the assumption that it has a potential to improve the traditional Chow/Lin approach by targeting the reference series more directly. The Lasso (least absolute shrinkage and selection operator) was introduced by Tibshirani (1996). Although, it can be used for estimation as shrinkage estimator it also can be used as model selection technique. The Lasso method estimates the linear regression coefficients by minimizing the sum of least squares subject to a l_1 penalty function:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

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. This behaviour leads to the model selection behaviour of the Lasso.

λ is a tuning parameter that captures the relative weight on the penalty function. Selecting a value for λ is crucial for Lasso. James et al. (2013) suggest cross-validation as a simple way to tackle this problem. We follow their approach and also use cross-validation for the selection of lambda in our calculations.

In this paper, the Lasso approach is used as selection method. Having selected a model by Lasso, a post-Lasso step is added which means that after model selection the parameters of the model are estimated by ordinary least squares without shrinkage.

The Chow and Lin methodology relies on a regression model to make use of high frequency indicators for temporal disaggregation. As always in regression models the problem of variable selection arises.

Regressing on all available high frequency series is likely to lead to overfitting. To avoid this, repeated combinations of a limited number of regressors can by chosen with the results subsequently submitted to model averaging (potentially along with other imputations as in Stuart, 2018). For our purposes, however, comparing specific imputation results is more informative (which of course does not rule out that in a practical implementation, model averaging may be appropriate). Accordingly, we would rather select variables by ordinary low frequency regression, where the usual variable selection methods can be applied. The elected variables can then be submitted to the Chow/Lin disaggregation algorithm.

In this two-step procedure, however, the two problems can also be linked more closely.[8] Based on the work of Wang et al. (2007) we combine the Chow and Lin procedure with Lasso shrinkage. Wang, Li and Tsay introduce Lasso for regressions with autoregressive errors. They obtain estimators by minimizing the Lasso criterion (for the present context the equation adapted to use the low frequency values)

$$\sum_{t=q+1}^{n_0}\left[(y_t^l - x'^{l}_{t}{}'\beta) - \sum_{j=1}^{q}\rho_j(\,y_{t-j}^l - x'^{l}_{t-j}\beta)\right]^2 + n\sum_{j=1}^{p}\lambda|\beta_j| + n\sum_{j=1}^{q}\gamma|\rho_j|$$

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. An AR(1) term is introduced for which no shrinkage is required. Penalty terms are computed only for the other regression parameters.

To solve the optimization problem Wang, Li and Tsay propose an iterative algorithm. This algorithm simplifies in the present context to:

In step i of the iteration:

1. Use Lasso with fixed $\hat{\phi}(i)$:

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

2. Use $\hat{\beta}(i+1)$ to estimate $\hat{\phi}(i+1)$ with OLS

Steps 1 and 2 are repeated until convergence. $\hat{\beta}(0)$ is estimated with standard Lasso, $\hat{\phi}(0)$ with OLS. The shrinkage parameter λ is chosen according to BIC, as proposed by Wang, Li, Tsay.

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

The *RMSE* for the real-time out-of-sample imputations, broken down by total, first month out-of-sample and second month out-of-sample, are shown in Table 3.6.

Table 3.6 shows that the real time *RMSE* for the Chow/Lin-Lasso algorithm in the pre-Covid-19 periods range from 3.7 to 4.0, which is again remarkably lower than for *locf* (6.7–7.4). The *RMSE* for the Covid-19 years 2020–2021 (5.3–5.8) are somewhat higher than for the preceding periods, but again remarkably lower than for *locf* (9.0–11.3). For direction changes, the percentage of hits in real time comes up to

[8] For a comprehensive discussion of the underlying principles, see Di Fonzo (2003).

77% before Covid-19 and up to 73% for 2020–2021 so that in this respect, Chow/Lin-Lasso also performs much better than random guessing.

Table 3.6: Descriptive statistics for the Chow/Lin-Lasso selection algorithm

Period	1991–2019		2000–2019		2020–2021	
Imputation	direct	indirect	direct	indirect	direct	indirect
real time						
RMSE all	3.96	3.75	3.98	3.85	5.73	5.38
RMSE m2	4.00	3.79	3.92	3.78	5.64	5.28
RMSE m3	3.91	3.70	4.04	3.92	5.82	5.48
Direction change	75.2%	75.6%	76.2%	77.4%	73.3%	72.7%
ex post						
RMSE all	n/a	n/a	3.76	3.68	5.14	4.68
RMSE m2	n/a	n/a	3.66	3.59	5.27	4.96
RMSE m3	n/a	n/a	3.85	3.78	5.00	4.39
Direction change	n/a	n/a	77.5%	76.8%	77.8%	74.4%

The ex-post performance is consistently better than in real time, and this again holds for *RMSE* as well as direction changes.

The two regularities that were observed for *locf*, largely hold for Chow/Lin-Lasso too, with a few exceptions: *RMSE* are *not* larger for m3 compared to m2 for the direct real time imputations covering the 1991–2019 period as well as the ex-post imputations for 2021–2022. *Also, RMSE* are larger for direct compared to compared to indirect imputations without exception.

What we take from this is that for our data – comparable to OLS and Chow/Lin without Lasso – the Chow/Lin-Lasso imputations consistently outperform the naïve *locf*. This applies to all periods considered as well as to real time and ex post.

In the next section, we will take a closer look at the relative performance of the different imputations.

4.1 Which approach is statistically significantly superior?

4.1 Comparative internal validation

In the previous section, we reported the results of standard forecasting accuracy evaluation, based on comparisons of the imputed monthly values with the known (true) ones. We refer to this as *internal* validation, as it is based solely on our 11 monthly BTS data series that underly the imputations. To facilitate the evaluation the relative performance of the different imputation algorithms, we report *relative* root mean square errors (*RRMSE*), defined as

$$RRMSE = RMSE^{k} / RMSE^{locf} ,$$

where *k* is the evaluated method and *locf* (last value carried forward) is the "naïve" benchmark model. A *RRMSE* above 1 indicates that method *k* fares worse than the benchmark across the 11 imputed survey items. Values below 1 signal the opposite, and the lower the value, the more promising the method.

As before, we also look at descriptive statistics regarding turning points. However, as the benchmark *locf* by construction does not impute turning points, we refer to the same statistics as above, which we show again in this statistical summary to ease comparison. Recall that values above 50% indicate that the imputations are reflecting direction changes better than simple guessing, and the higher the better.

The *RRMSE* for the real-time (out-of-sample imputations), broken down by total (first and second month), first month and second month, and direct imputation of the seasonally adjusted balance are versus indirect via separate imputations of the seasonally adjusted positive and negative shares, from which the balance is subsequently computed, are shown in Table 4.1.1. The statistics for the best results are highlighted in **bold**, and in *italics* where this applies to more than one imputation method.

Table 4.1.1 shows that the real-time *RRMSE* for the cubic spline imputations without exception clearly exceed one. As, in addition to this the direction change predictions are not superior to random guessing, it is clearly preferable for our data corpus to freeze the last observations (*locf*) than to create out-of-sample variance within quarters with the cubic spline.

The EM algorithm could be expected to outperform the cubic spline, as it refers to the entire data corpus rather than projecting univariate dynamics into the future. For our data corpus, however, this expectation is disappointed when it comes to the *RRMSE*, which do not really look better. Even if they are somewhat less disastrous for the 2020–2021 Covid-19 years in comparison to the cubic spline, where without a clue of the pandemic delivered by real-time data, the forecast errors are going through the roof, they still all exceed one. The percentage of direction change hits, on the other hand, is about 65% and thus better than guessing, so that there is some ambivalence in the outcome of this imputation method.

Turning to OLS regression, our results show that this imputation is preferable to *locf* in all respects considered here. The *RRSME* are all clearly below one, and the percentage of direction change hits reaches up to 75%. Exploiting the information from 10 BTS series via multiple regression to impute monthly values of the eleventh works well in our data corpus and provides improved timely estimates for the second a third months of a quarter at the right margins of our series.

The findings for Chow/Lin and the Chow/Lin-Lasso algorithms are qualitatively the same as for the OLS based imputations, but while the percentage of direction change hits falls in the vicinity of 75% and is on the whole only slightly better than with OLS, the *RRMSE* drop from above 0.65 to below 0.55. The latter result is unambiguous, not even a single *RRMSE* from the OLS imputation is lower than any of the two counterparts from Chow/Lin or Chow/Lin-Lasso. Accordingly, for the real-time imputations within our data corpus, the two Chow/Lin based methods deliver the most accurate results.

The picture that emerges from the descriptive statistics thus shows that Chow/Lin and Chow/Lin-Lasso are unambiguously superior to the other imputation methods considered here. Also, it appears that neither of the two favourites clearly outperforms the other. Yet, two regularities become evident under closer inspection. Firstly, up to 2019, Chow/Lin without Lasso tends to score lower on *RRMSE* for the direct imputations, whereas Chow/Lin-Lasso performs superior in this respect for the indirect approach. Moreover, the percentage of direction change hits up to 2019 reveals the same pattern. Given that the indirect approach delivers more precise results than the direct one, superior performance with the indirect approach can in practice be especially useful.

Table 4.1.1: Comparative statistics with respect to benchmark, real-time imputations

Period	1991–2019		2000–2019		2020–2021	
Imputation	direct	indirect	direct	indirect	direct	indirect
Cubic spline						
RRMSE all	1.28	1.23	1.30	1.26	1.86	1.83
RRMSE m2	1.14	1.10	1.14	1.11	1.28	1.26
RRMSE m3	1.40	1.35	1.43	1.39	2.30	2.26
Direction change	50.5%	51.6%	49.4%	51.5%	47.7%	46.0%
EM algorithm						
RRMSE all	1.22	1.23	1.21	1.20	1.15	1.11
RRMSE m2	1.20	1.21	1.18	1.18	1.17	1.15
RRMSE m3	1.24	1.25	1.23	1.22	1.13	1.08
Direction change	64.3%	65.6%	63.5%	65.9%	64.2%	65.9%
OLS regression						
RRMSE all	0.69	0.68	0.68	0.68	0.59	0.59
RRMSE m2	0.70	0.68	0.69	0.68	0.60	0.62
RRMSE m3	0.67	0.68	0.67	0.68	0.58	0.55
Direction change	74.8%	73.7%	75.3%	74.2%	72.7%	71.6%
Chow/Lin						
RRMSE all	**0.56**	*0.54*	*0.55*	0.54	**0.55**	**0.51**
RRMSE m2	*0.57*	0.55	*0.54*	0.53	**0.54**	**0.50**
RRMSE m3	**0.55**	0.54	**0.55**	0.55	**0.56**	**0.51**
Direction change	**75.9%**	75.0%	**77.1%**	76.0%	70.5%	72.2%
Chow/Lin-Lasso						
RRMSE all	0.57	*0.54*	*0.55*	**0.53**	0.56	0.53
RRMSE m2	*0.57*	**0.54**	*0.54*	**0.52**	0.55	0.52
RRMSE m3	0.56	**0.53**	0.56	**0.54**	0.57	0.54
Direction change	75.2%	**75.6%**	76.2%	**77.4%**	**73.3%**	**72.7%**

For the Covid-19 years, Chow/Lin scores slightly better in terms of *RRMSE*, and Chow/Lin-Lasso outperforms Chow-Lin regarding direction changes. This pattern is different from the one of the pre-Covid-19 years, but one should keep in mind that the Covid-19 results are based on eight observations only.

What we take from this is that our expectation that Chow/Lin-Lasso should in real time prove superior to the other imputation methods is strongly confirmed regarding all alternatives except Chow/Lin, where a difference is detectable, but it appears too small to allow strong conclusions.

The remaining evidence from the internal validation is the comparative performance of the ex-post imputations. The results are shown in Table 4.1.2. As above, the statistics for the best results are highlighted in **bold**, and in *italics* where this applies to more than one imputation method. Also notice that for the pre-Covid-19 years, due to data availability, the table refers to the 2000–2019 period only.

Table 4.1.2 shows that the ex-post *RRMSE* for the cubic spline imputations are without exception clearly below one. In contrast to real time (out of sample), this method apparently works, when it comes to impute missing observation *between* known values. In addition, the direction change predictions are now somewhat superior to random guessing. Accordingly, while it is preferable for our data corpus to freeze the last two observations out of sample, past values (in sample) imputed with the cubic spline are superior to the *locf* benchmark.

Again, the EM algorithm could be expected to outperform the cubic spline, as it not univariate but referring to the entire data corpus. However, this expectation is again disappointed when it comes to the ex-post *RRMSE*, which are all exceeding one. The percentage of direction change hits, on the other hand, is somewhat better than with the cubic spline, but the fact that the errors with *locf* are lower than for EM does not even recommend this method for ex-post imputations of values from the past.

Ex post OLS regression imputations are preferable to *locf* in all respects considered here. All *RRMSE* are clearly below one and they tend to be slightly lower than their real-time counterparts. The percentage of direction change hits reaches up to 75% and is comparable to the real-time equivalents. Multiple regression imputations of monthly values thus work well in our data corpus, providing timely estimates for the second a third months of a quarter at the right margins and also between observed values.

The results for the ex-post Chow/Lin as well as the Chow/Lin-Lasso imputations are remarkably better than for all alternatives considered here, confirming the real-time findings. Moreover, the ex-post *RRMSE* are consistently below their real-time equivalents up to 2019, and better of equal for the two Covid-19 years. In addition, the percentages of direction change hits are consistently higher than the real-time equivalents. Accordingly, the two Chow/Lin based methods deliver the not only the most accurate real-time results, but also ex post, and the ex-post results tend to be more precise than the out-of-sample-forecasts. Also, like in real time, neither of the two favourites consistently outperforms the other, but ex post, the advantages of amending the Chow-Lin algorithm with Lasso and AR1 term appear somewhat more consistent, as they are now also showing up for the two Covid-19 years.

What we take from this is that all imputation methods are delivering more precise results when it comes to impute missing values that lie between known ones in the past, compared to the same task in simulated real time, which technically amounts to forecasting. Moreover, as in simulated real time, Chow/Lin-Lasso proves clearly superior to the other imputation methods except Chow/Lin, where the statistics for the preferred look slightly better, but appear too small to allow strong conclusions. Last but not least, it is not inconceivable that different algorithms could be superior for ex-post and real-rime imputations, which would recommend imputing (nowcast) the last two unobservable months at the right margin of a series with one method and all other missing observations (those that lie between known data points) with another method. For our data corpus and the methods under consideration here, however, this does not apply. The recommended method is Chow/Lin-Lasso for both tasks. On this basis, the next section will look at the performance of the recommended method when we move beyond out 11 BTS series and face truly missing monthly observations.

Table 4.1.2: Comparative statistics with respect to benchmark, ex-post imputations

Period	2000–2019		2020–2021	
Imputation	direct	indirect	direct	indirect
Cubic spline				
RRMSE all	0.95	0.92	0.88	0.85
RRMSE m2	0.93	0.90	0.89	0.87
RRMSE m3	0.96	0.93	0.86	0.84
Direction change	60.9%	60.6%	61.9%	60.2%
EM algorithm				
RRMSE all	1.15	1.13	1.13	1.09
RRMSE m2	1.11	1.09	1.15	1.13
RRMSE m3	1.19	1.17	1.11	1.04
Direction change	62.8%	63.7%	66.5%	66.5%
OLS regression				
RRMSE all	0.64	0.65	0.57	0.56
RRMSE m2	0.64	0.65	0.59	0.63
RRMSE m3	0.63	0.65	0.55	0.49
Direction change	74.4%	73.9%	75.0%	72.7%
Chow/Lin				
RRMSE all	*0.52*	0.52	0.52	*0.46*
RRMSE m2	0.52	0.50	0.54	0.50
RRMSE m3	*0.53*	0.53	0.50	**0.40**
Direction change	**77.7%**	76.1%	75.6%	72.7%
Chow/Lin-Lasso				
RRMSE all	*0.52*	**0.51**	**0.50**	*0.46*
RRMSE m2	**0.50**	**0.49**	**0.51**	**0.48**
RRMSE m3	*0.53*	**0.52**	**0.49**	0.43
Direction change	77.5%	**76.8%**	**77.8%**	**74.4%**

4.2 External validation

The last step of our empirical analyses is an attempt of *external* validation. For this, we evaluate the congruence of a (truly) imputed series from the KOF quarterly BTS question about firms' *technical* *capacity utilisation in per cent* (*CapU*) with the following six external monthly time series that can theoretically be expected to be related to it:

1. The KOF Economic Barometer, perhaps the most prominent monthly composite leading indicator for the Swiss economy.[9]
2. The Leading Global Barometer, a monthly composite leading indicator for the Swiss economy, developed and published jointly be the KOF Swiss Economic Institute and the Brazilian Fundação Getúlio Vargas (FGV).[10]
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.
4. Swiss National Bank's Business Cycle Index (SNB BCI), a monthly composite indicator designed to reflect the Swiss GDP growth cycle.[11]
5. The official Unemployment Rate, a monthly series based on the number of registered unemployed persons in Switzerland.[12]
6. Consumer Price Inflation, measured as the year-on-year growth rate of the Swiss Consumer Price Index.[13]

These are our six monthly *external* reference series.[14] Where the data sources offered the option, we took seasonally and calendar day adjusted data, unadjusted series otherwise.

The (truly) missing values for *CapU* in the second and third months of each quarter are imputed as before, but now all eleven (and not ten out of eleven) monthly BTS series are used for the multivariate approaches. The imputation is conducted in simulated real time with our Chow/Lin-Lasso algorithm, the method that has proven superior in the internal validation process. The window ranges from 2000m1 to 2019m12, which is right up to the pandemic and the same as the main period considered for the internal validations reported above, comprising 240 monthly values, of which 160 are imputed. Notice that *CapU* is one of the few quantitative KOF BTS series, so the imputations are direct only (as there are no positive- and negative shares).

To allow for different types of associations between *CapU* and the six reference series, we refer to *CapU* alternatively in levels and in year-on-year (y-o-y) growth rates. The same applies for the six reference series, but here, we compute annual differences rather than growth rates, when the scale of a series mandates this. Then, we determine the strength and lead/lag relations between the different specifications of *CapU* and the external reference series by means of pairwise cross correlations, for which we allow 15-months windows. The analysed period ranges from 2000m1 to 2019m12, but we refer to all available data points of the six reference series to identify leads or lags.[15] For each reference series, we

[9] For details, see https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-economic-barometer.html.
[10] For details regarding this and the next indicator on the list, see https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalbaro.html.
[11] For details, see https://data.snb.ch/en/topics/snb/chart/snbbcich.
[12] For details, see https://www.bfs.admin.ch/bfs/en/home/statistics/work-income/unemployment-underemployment.html.
[13] For details, see https://www.bfs.admin.ch/bfs/en/home/statistics/prices/surveys/lik.html.
[14] The *external* series are not part of the KOF BTS, i.e. they result from a different data generating process. We have resorted to regularly published and freely available data only. To ensure that they are *meaningfully* related to technical capacity utilisation, we carefully inspect whether the signs and leads/lags of the relations correspond to economic expectations and experience.
[15] At the time of writing, all reference series stretched well into 2022 and went back into the 1990s.

identify the highest absolute cross correlation with *CapU* and determine the proper variable specifications (levels or growth rates/differences) and lead/lag structures accordingly. Table 4.2.1 summarises the findings, ordered according to the maximum absolute cross correlations.

Table 4.2.1: Cross correlations between capacity utilisation and reference series

Reference series	*CapU*	Max abs. cross correl.	Lead/lag of ref. 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 growth rate	y-o-y growth rate	-0.65	3 months lag
Consumer Price Index (CPI) Inflation			
y-o-y growth rate	Level	0.50	2 months lead

The table shows that the highest absolute cross correlation (0.86) is between the y-o-y growth rate of *CapU* and the KOF Economic Barometer, where the latter has a lead of 7 months. This is according to expectations, as the KOF Economic Barometer is designed as a leading indicator for the Swiss economy.

The second highest cross correlation (0.85) is between the y-o-y growth rate of *CapU* and the KOF/FGV Leading Global Barometer, where the latter has a lead of 6 months. This is also plausible, as the KOF/FGV Leading Global Barometer is designed as a leading indicator for the world economy, and the business cycle of Switzerland as a small open economy is largely driven by the world economy.

The third highest cross correlation (0.84) is between the y-o-y growth rate of *CapU* and the KOF/FGV Coincident Global Barometer, where the latter has a lead of 4 months. The reduced lead makes sense, as it corresponds to the lead/lag relationship of the KOF/FGV Global Barometers.

The fourth highest cross correlation (0.80) is between the y-o-y growth rate of *CapU* and the y-o-y difference of the Swiss National Bank's BCI, where the latter has a lead of 10 months. The longer lead compared to the KOF Economic Barometer goes along with a lower correlation, reflecting the trade-off between lead and precision that has to be expected when designing leading indicators.

The fifth highest absolute cross correlation (0.65) is between the y-o-y growth rate of *CapU* and the y-o-y growth rate of the official Swiss Unemployment Rate. The correlation is negative, and *CapU* is leading unemployment by 3 months. This makes perfect 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 sixth and last highest cross correlation (0.50) is between *CapU* and the y-o-y growth rate of the Swiss CPI, i.e. the annual inflation rate, where the latter has a lead of only 3 months. As pressure on technical capacity utilisation can be expected to go along with rising output prices, this also makes sense, even as theory would have capacity lead inflation rather than the other way round, but the empirical lead is so small that it is coming close to reflecting a coincident association.

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 other six series that are also directly or indirectly reflecting the state of the Swiss economy.

What remains to be seen is whether the intra-quarter variance created by the imputation of monthly *CapU* strengthens the associations or not. Our criterion for this 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 Chow/Lin-Lasso imputation of *CapU* compete with the internal benchmark locf in predicting the six reference series, which amounts to comparisons of non-nested models, for which the J-test is adequate.[16]

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:

$Y = X_1\, g + \varphi\, (X_2\, \hat{h})$,

$Y = X_2\, h + \tau\, (X_1\, \hat{g})$.

The test statistics are the t-values for φ and τ. 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, $CapU^{Chow/Lin\text{-}Lasso}{}_{t-L}$ and $CapU^{locf}{}_{t-L}$, the J-test is here identical with the simpler *encompassing test* (E-test), which consists of running the regression

$R_t = g\; CapU^{Chow/Lin\text{-}Lasso}{}_{t-L} + h\; CapU^{locf}{}_{t-L} + \mu_t$,

and submitting g and h to t-tests. The superscripts to *CapU* denote the imputation methods, L is the lead/lag applied to synchronise the series according to Table 4.2.1 and μ is white noise. The decision rule equals that of the J-test. As above, 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 4.2.2.

[16] See Davidson and MacKinnon (1981) and Mizon and Richard (1986).

Table 4.2.2: Encompassing tests (2000–2019)

	Chow/Lin-Lasso			*locf*			
Series/statistics	*ß*	*t-stat.*	*p-value*	*ß*	*t-stat.*	*p-value*	R^2
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	1.67	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

The table shows that the E-test statistics are unambiguously in favour of the Chow/Lin-Lasso imputations for all six reference series. The regression coefficients *ß* all have the expected sign, and the t-statistics are all high in absolute terms so that the associated *p* statistics clearly pass the 1-percent-significance hurdle. For the competing *locf* imputation, the *ß*-coefficients, are all considerably smaller and none of the *t*-statistics comes even close to indicate 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 reference series, so that we can conclude that the variance created by the Chow/Lin-Lasso imputation is meaningfully related to the monthly dynamics of the Swiss economy.

The last step of our analyses is to conduct 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 very small for statistical inference, we can at least see if the pre-Covid-19 results are not reverted, which might cause concerns. The results are shown in Table 4.2.3.

Table 4.2.3: Encompassing tests, Covid-19 years (2020–2021)

	Chow/Lin-Lasso			*locf*			
Series/statistics	*ß*	*t-stat.*	*p-value*	*ß*	*t-stat.*	*p-value*	R^2
Economic Barometer	1.00	1.80	*< 10%*	1.52	2.08	**< 1%**	0.63
Leading Global Barometer	1.29	2.84	**< 1%**	1.01	2.29	*< 5%*	0.67
Coincident Global Barometer	1.23	2.63	**< 1%**	0.78	1.71	*< 10%*	0.68
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	0.41	0.70
CPI Inflation	0.26	4.75	**< 1%**	-0.01	-0.35	0.73	0.84

The table shows that of the six E-tests, three are unambiguously and in favour of the Chow/Lin-Lasso imputations (for predictions of the Leading Global Barometer, the Coincident Global Barometer and CPI inflation). The regression coefficients have the expected sign, the *t*-statistics are all high in absolute terms and the associated *p* statistics clearly pass the 1-percent-significance hurdle. Lowering the significance hurdle to 5 percent gets the Unemployment Rate on the same side. Unless we lower the significance hurdle even more to 10 per cent, when the result for predictions the KOF Economic Barometer become inconclusive, the E-test for this reference series is in favour of *locf*. Also, the KOF barometer regression is the only one where the regression coefficients *ß* is higher for *locf* than for Chow/Lin-Lasso. For the sixths and last reference series, the Swiss National Bank's BCI, neither of the two competing imputation contributes to a significant prediction, where the question must remain open whether this is due distorted BCI signals during the pandemic or to insufficient precision of the Chow/Lin-Lasso procedure for these two years.

In this context, notice the differences between the R^2 for the pre-Covid-10 period and the Covid-19 years (Tables 4.2.2 and 4.2.3). For the three Barometers, they are equal or somewhat smaller for 2020–2021, for unemployment and inflation, they are markedly higher, and it is only for the BCI that we observe a sizable drop (from 0.64 to 0.27). The fact that the BCI proves to be an outlier in this respect point to the interpretation that during the pandemic, the signals from the BCI were more distorted than those of the other reference series for the economic situation in Switzerland.

Taken together, we would argue that the final look at how the variance created by the Chow/Lin-Lasso imputation is by and large meaningfully related to the monthly dynamics of the Swiss economy, although we do not have an explanation for the observation that for 2020–2021, the dynamics of the KOF Economic Barometer in the second and third months of the eight quarters concerned is better reproduced by carrying the first months forward.

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 in both monthly and quarterly frequency, where an information sub-set is collected at quarterly frequency only.

From these data, we construct *artificially* quarterly series. We then impute monthly values for the (seemingly) missing ones. On this basis, we run standard tests of forecasting accuracy by comparing the imputed monthly values to the original ones. We refer to this as *internal* validation. All analyses are performed alternatively in simulated real time (out of sample) and ex post (in sample)

For the simplest imputation method, the last value carried forward procedure, we find that that the imputation errors are higher on average for the third months of a quarter compared to the seconds months, which is in line with expectations, as the distance to the last recorded observation increases. Also, the indirect imputations via imputed seasonally adjusted positive and negative shares turn out to be slightly but consistently more precise than the direct imputations of the seasonally adjusted balance, which shows up repeatedly and also for the more sophisticated imputation algorithms.

The last value carried forward procedure serves as our benchmark against which we evaluate the performance of the more sophisticated imputation algorithms. The other univariate imputation method that we look at is the cubic spline. It is clearly inferior to the benchmark in real time, the variance created by the imputations is misleading, and this holds especially for the dramatic Covid-19 years. Ex post, the precision of the imputations still does not recommend the cubic spline as an alternative to the benchmark, although direction changes are captured a bit better.

Of the multivariate approaches, the EM algorithm imputation does not hold the promise to outperform the benchmark, and this applies to real time and ex post. The only improvement is for directions changes. Multiple OLS regression imputations, on the other hand, consistently outperform the benchmark, both in simulated real time and ex post. The same holds for imputations following Chow and Lin. Moreover, the latter consistently outperform the OLS regression approach.

The last and most sophisticated algorithm is our own specification, a combined implementation of Chow/Lin with Lasso and a first-order autoregressive term. These imputations do not differ greatly from those of the traditional Chow/Lin approach, but taken together, they tend to slightly outperform the latter. Interestingly, traditional Chow/Lin imputations tend to be more precise for the direct imputations of the seasonally adjusted balance indicators, whereas our amended Chow/Lin algorithm performs superior with the indirect approach, which first imputes seasonally adjusted positive and negative shares and computes the imputed target values as the difference between the two. Notice that the indirect imputation is informationally more efficient, as the same value for the balance can result from greatly different positive and negative shares. Given that the indirect approach delivers more precise results than the direct one, superior performance with the indirect approach can in practice be especially useful.

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.

For the pre-Covid-19 years, the statistics are unambiguously in favour of our amended Chow/Lin-Lasso imputations for all six reference series. In particular, the external validation confirms that our preferred imputations of technical capacity utilisation are significantly superior to the benchmark series in predicting the six reference series, so that we can conclude that the variance created by the imputations is meaningfully related to the monthly dynamics of the Swiss economy.

For the Covid-19 years 2020–2021, the number of observations (24, of which for capacity utilisation 16 are imputed) is uncomfortably small for statistical inference, and the results are not unambiguous, but we find that the variance created by the imputations is by and large still meaningfully related to the monthly dynamics of the Swiss economy, although we do not have an explanation for the observation that for 2020–2021, the dynamics of the KOF Economic Barometer in the second and third months of the eight quarters concerned is better reproduced by carrying the first months forward.

It is in order to stress that our findings and generalisations are based on our particular data corpus, which are Swiss BTS and other economic data relating to Switzerland spanning roughly the last three decades. It remains to be seen whether our generalisations can be replicated with data from other countries, but we feel confident enough to argue that cubic spline extrapolations are likely to create misleading signals in real time, that the wide-spread trust in the EM algorithm should be questioned, that traditional OLS regressions may still be very useful for imputations, both in real time and ex post, and finally, the most promising direction for future research as well as practical applications may be to work on refining the Chow/Lin imputation algorithm.

Last but not least, future research should address the unexpected finding that the difference between the widely used balance indicator computed from the seasonally adjusted balance of the unadjusted plus- and minus-shares and the difference of the seasonally adjusted plus- and minus-shares does has systematic effects on the reproducibility of the series, which may have consequences for the nature of the signals that they are giving.

References

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

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

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.

Čižmešija, M., Sorić, P. and 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.

Chow, G. C. and 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.

Davidson, R. and 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.

Di Fonzo, T., 2003. Temporal disaggregation of economic time series: towards a dynamic extension. European Commission Working Papers and Studies, Theme 1, General Statistics.

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

Feijoó, S. R., Rodríguez Caro, A. and 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.

Galli, A. Which Indicators Matter? 2018. Analyzing the Swiss Business Cycle Using a Large-Scale Mixed-Frequency Dynamic Factor Model. Journal of Business Cycle Research, 14: 179–218 (2018).

James, G., Witten D., Hastie T. and Tibshirani, R., 2013. An introduction to statistical learning. Springer, Heidelberg.

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

Labonne; P. and Weale, M., 2020. Temporal disaggregation of overlapping noisy quarterly data: estimation of monthly output from IK value-added tax data. Journal of the Royal Statistical Association, 183(3): 1211–1230.

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. and Richard, J.-F., 1986. The Encompassing Principle and its application to testing nonnested hypotheses, Econometrica, 54(3): 657–678.

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

Mosley, L., Eckley, I. and Gibbert, A., 2021. Sparse Temporal Disaggregation, Mimeo.

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

Quenneville, B., Picard, F. and Fortier, S., 2013. Calendarization with interpolating splines and state space models. Applied Statistics, 62(3): 371–399.

Sax, C. and Steiner, P., 2013. Temporal Disaggregation of Time Series. The R Journal, 5(2): 80–87.

Schumacher, C. and 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. and 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. and 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. and 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.

Zult, D., Krieg, S., Schouten, B., Ouwehand, P. and van den Brakel, J., 2020. From Quarterly to Monthly Turnover Figures Using Nowcasting. CBS Discussion Paper, Statistics Netherlands.

Appendix: 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 compeitve 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. Custemer 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.

<<0>>