A Monte-Carlo Evaluation of Regression-Based Temporal Disaggregation Methods

Christoph Sax*
University of Basel

July 2017

Abstract

Using Monte-Carlo Simulations, this paper evaluates regression-based methods for temporal disaggregation (Chow-Lin, Fernandez, Litterman and dynamic Chow-Lin) as well as procedures for estimating the autoregressive parameter that is used by some of these methods (max-log, min-RSS).

In a first exercise, time series are simulated that are consistent with the simplifying theoretical assumptions of the methods. In a second exercise, real-world SARIMA processes are simulated and used for evaluation.

In line with the theoretical expectations, Chow-Lin performs best for cointegrated series, Litterman or Fernandez for non-cointegrated series. Somewhat surprisingly, the dynamic extension of Chow-Lin under-performs in both cases. If there is no a priori knowledge on the degree of cointegration, the Chow-Lin method with a slightly modified max-log algorithm for finding the autoregressive parameter leads to best results.

*Contact Address: University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland. I would like to thank Peter Steiner, Tommaso Di Fonzo, Bruno Parnisari, Ronald Indergand, Stefan Leist and seminar participants at the CFE Conference 2014 in Pisa, Italy. This reasearch was funded by the State Secretariat for Economic Affairs SECO, Switzerland.

1 Introduction

Temporal disaggregation methods are used to disaggregate low frequency time series to higher frequency series, while either the sum, the average, the first or the last value of the resulting high frequency series is consistent with the low frequency series. In many European countries (e.g. France, Italy, Switzerland), quarterly figures of Gross Domestic Product (GDP) are estimated by temporal disaggregation methods. Together with the selection of the indicator series, the choice of the disaggregation method is essential for the quality of both inter- and extrapolation of quarterly GDP.

Using Monte Carlos simulations, this paper evaluates the standard regression-based methods for temporal disaggregation: **Chow-Lin** (Chow and Lin, 1971), with different algorithms for determining the autoregressive parameter (Bournay and Laroque, 1979; Barbone *et al.*, 1981), **Fernandez** (Fernández, 1981), **Litterman** (Litterman, 1983) and a **dynamic Chow-Lin** formulation by Silva and Cardoso (2001). The inclusion of Kalman-filter-based methods (Proietti, 2006) is beyond the scope of this paper.

Two simulation exercises are performed. In a first exercise, time series are simulated that are consistent with the simplifying theoretical assumptions of the methods. In a second exercise, a range of real-world seasonal autoregressive integrated moving average (SARIMA) processes are simulated. These processes are not consistent with the theoretical assumptions of the methods. The processes are chosen by analyzing actual data, where both the true series and the indicator series is available on a quarterly level. SARIMA models are estimated both for the quarterly error term as well as for the quarterly indicators.

Both exercises show that the degree of cointegration in the series pairs is crucial for the outcome of the forecast competition. Litterman and Fernandez performs best with non-cointegrated, Chow-Lin with cointegrated series. This finding does not depend on whether the residuals follow the simplified processes that are implied by the models or whether real-world SARIMA processes are considered. If there is no a priori knowledge on the degree of cointegration, the Chow-Lin method with a slightly modified max-log algorithm for finding the autoregressive parameter leads to best results.

There are some empirical studies on the performance of temporal disaggregation methods in the literature. Litterman (1983) has compared his method to different Chow-Lin methods using a few real time series. In a Monte Carlo experiment, Miralles *et al.* (2003) have evaluated the interpolation accuracy of Chow-Lin methods in detail. In a similar experiment, a study by Ciammola *et al.* (2005) has evaluated different disaggregation methods. This paper largely confirms and extends the findings of these last two papers.

The paper is organized as follows: The first section describes and discusses the methods evaluated in this paper. Section 3 explains the set-up of the Monte Carlo simulations. Section 4 describes the results. The final section summarizes the results and formulates practical recommendations for the use of disaggregation methods.

Simulations and estimations have been performed in R (R Core Team, 2017). To make the analysis fully reproducible, all data, code and content to build this paper is available in the *tdmc* R package, which can be easily installed from GitHub. See the corresponding package page for details (Sax, 2017).

2 Methods to Evaluate

The aim of temporal disaggregation is to find an unknown high frequency series, y, whose sums, averages, first or last values are consistent with a known low frequency series, y_l (The subscript l denotes low frequency variables). In order to estimate y, one or more high frequency indicator variables can be used. For the ease of exposition, the terms annual and quarterly will be used instead of low frequency and high frequency. For a more extensive discussion using the same notation, see Sax and Steiner (2013).

The basic assumption of Chow-Lin, Fernandez and Litterman is that the relationship between the annual values of the true series and the indicator series also holds on a quarterly level. The methods perform a Generalized Least Squares Regression (GLS) of the annual values on the annualized quarterly indicator series and use the estimated coefficients to calculate at preliminary estimation of the true quarterly series. The GLS estimator is defined as follows:

$$\hat{\beta}(\Sigma) = \left[X'C'(C\Sigma C')^{-1}CX\right]^{-1} X'C'(C\Sigma C')^{-1}y_l\,. \tag{1}$$

where X is a matrix with one or more quarterly indicator variables, including a constant. C is the conversion matrix that aggregates quarterly series to annual series; for flow series, it is defined as the Kronecker product, $I_{n_l} \otimes [1, 1, 1, 1]$; for quarterly index series, the result is multiplied by 0.25. Σ is the quarterly variance-covariance matrix, while $C\Sigma C'$ is its annual equivalent.

With the regression coefficients at hand, a preliminary quarterly series, $\hat{\beta}X$, can be calculated. The final quarterly series is constructed as:

$$\hat{y} = \hat{\beta}X + D\hat{u}_l\,. \tag{2}$$

where $\hat{u}_l$ is a vector containing the remaining differences between the annualized values of $\hat{\beta}X$ and the actual annual values, y_l ($\hat{u}_l \equiv \; y_l - C\hat{\beta}X$). D is the distribution matrix, a function of the same variance-covariance matrix that was used in the GLS estimation, Σ:

$$D = \Sigma\, C'(C\,\Sigma\, C')^{-1}\,. \tag{3}$$

The methods differ in how the variance-covariance matrix, Σ, is constructed. Chow-Lin assumes that the quarterly residuals follow an autoregressive process of order 1 (AR1), i.e., $u_t = \rho u_{t-1} + \epsilon_t$, where ϵ is $\mathrm{WN}(0, \sigma_\epsilon)$ (with WN denoting White Noise) and $|\rho| < 1$. Fernandez and Litterman assume that the residuals follow a non-stationary process, i.e., $u_t = u_{t-1} + v_t$, where v is an AR1 ($v_t = \phi v_{t-1} + \epsilon_t$, where ϵ is $\mathrm{WN}(0, \sigma_\epsilon)$). Fernandez is a special case of Litterman, where $\phi = 0$, and, therefore, u follows a random walk. While the dynamics in these models are modeled through the error term, the dynamic Chow-Lin method by Silva and Cardoso (2001) explicitly models the dynamic structure and estimates the model by a simple transformation of the original Chow-Lin model. An Overview of the methods evaluated in this paper is given in Table 1.

There are several ways to estimate the (quarterly) autoregressive parameter ρ or ϕ in the Chow-Lin and Litterman methods from the annual data. Bournay and Laroque (1979, p. 23) suggest the maximization of the likelihood of the GLS-regression. Another approach is the minimization of the weighted residual sum of squares, as it has been suggested by

Table 1: Temporal Disaggregation Methods to Evaluate

Methods	Parameter	Remarks
Chow-Lin	fixed ρ	
Dynamic	fixed ρ	
Fernandez	$\phi = 0$	
Litterman	fixed ϕ	
Chow-Lin	max-log	
Chow-Lin	min-RSS	
Chow-Lin	max-log	trunc. $\rho > 0.2$
Dynamic	max-log	
Dynamic	min-RSS	
Dynamic	max-log	trunc. $\rho > 0.2$
Litterman	max-log	
Litterman	min-RSS	
Litterman	max-log	trunc. $\phi > 0.2$

Barbone *et al.* (1981).

As will be shown, the max-log algorithm by Bournay and Laroque (1979) generally leads to good estimates of the autoregressive parameter, ρ. Sometimes, however, the algorithm fails dramatically and produces a strong negative estimation, when the true parameter is in fact positive. These negative estimations not only decrease its forecast accuracy, they also increase the in-sample volatility of the series. A reasonable workaround is the truncation of the estimation to 0 or to a low positive value, like 0.2. A positive value, rather than a value of 0, has the advantage that the residuals will be distributed smoothly among the quarters.

3 Simulation Set-Up

For the simulation experiment, a quarterly *indicator series* and a *true series* are required. The true quarterly series is aggregated to an annual series, which is observable to the disaggregation method. The method attempts to estimate the true series using the indicator series.

Consistent with the theoretical structure of the disaggregation methods discussed in section 2, the true relationship between the quarterly indicator series, x, and the true quarterly series of interest, y, is assumed to be the following:

$$y = \beta_0 + \beta_1 x + u \tag{4}$$

where u is a stationary or non-stationary error term. In the simulations, a single indicator series, x, is used, thus the difference in notation to the previous section. The true annual series, y_l, is calculated the following way:

$$y_l = Cy \tag{5}$$

where C is defined as in the previous section.

In order to set up the simulation, a quarterly indicator series, x, and a quarterly error series, u, are simulated. Together, these variables determine the quarterly true series, y. In the following, two main Monte Carlo experiments have been performed. In a first experiment, time series are simulated that are consistent with the simplifying theoretical assumptions of the methods. In a second experiment, several real-world SARIMA processes are simulated and used for evaluation, both with cointegrated and non-cointegrated time series.

Table 2: Overview of the Simulation Set-Up

	Simulation Steps	AR1 simulations	Real-world SARIMA simulations	Outcome
1	Template series	–	Selection of suitable pairs of quarterly template series, Series 1 representing the true series (Table 3, Col. 2), Series 2 the indicator (Table 3, Col. 3).	Template series
2	Inter-series relationship	By assumption: $\beta_0 = 10$, $\beta_1 = 0.5$	OLS Regression of Series 2 on Series 1 (Table 11)	Coefficients β_0, β_1
3	Template residuals	–	Regression residuals from Step 2	Template residual series
4	Model selection	By assumption: Indicator follows a random walk, residual series an AR1 or a (autoregressive) random walk.	Residuals from Step 3 and indicator series from Step 1 serve as a template for the SARIMA model selection. (Table 3, Col. 4, 5)	Time series models describing indicator and residual series
5	Simulation	Simulation of the models from Step 4		1000 artificial quarterly indicator series, x, and residual series, u
6	True series	$y = \beta_0 + \beta_1 x + u$		1000 artifical quarterly true series, y.
7	Aggregation	$y_l = Cy$		Annual series, y_l
8	Disaggregation	Disaggregation methods are applied to the annual series, y_l, and the indicator series, x.		Disaggregated series, $\hat{y}$
9	Evaluation	The disaggregated series, $\hat{y}$, are compared to the quarterly true series, y. (Tables 4 to 10)		Benchmarking statistics

Table 3: Template series for SARIMA modeling. The residual model is based on the residuals from the regressions of the true series on the indicator series, shown in Table 11.

Abbr.	True Series	Indicator Series	Indicator	Residual
Cointegrated Series				
SAL	Sales of Chemicals and Pharmaceuticals	Exports of Chemicals and Pharmaceuticals	(0,1,2)(0,1,1)	(2,0,3)(1,0,1)
IMP	Imports of Chemicals and Pharmaceuticals	Total Imports	(0,1,0)(1,0,0)	(1,0,2)(1,0,1)
ENE	Consumer Prices Energy	Consumer Prices Oil	(1,0,1)(2,0,0)	(2,0,1)(2,0,0)
Non-cointegrated Series				
CON	Construction Index SBV	Construction Employment	(1,1,1)(1,0,0)	(0,0,2)(0,1,0)
SER	Consumer Prices Services	Consumer Prices Total	(0,1,2)(0,0,0)	(2,0,1)(0,1,1)

3.1 Autoregressive Models of Order 1

The methods discussed in section 2 differ in their theoretical assumptions about the behavior of the error term, u. As a starting point, series are simulated that are in line with the theoretical assumptions of the models. The error term, u, is simulated as an AR1 process, using ρ values from -0.5 to 0.85, or as a random walk with an autoregressive innovation term, ϕ, using values from 0 to 0.85. Together, this yields various stationary and non-stationary processes (Table 2, Steps 2 and 4). The simulated series have a length of 92, a typical length of an indicator series in GDP estimation.

A simulated random walk process is used as an indicator series, x. In each run of the experiment, both a new indicator series and a new error term series is drawn (Step 5). As it turns out, the modeling of x is of minor importance. Different specifications are used as a robustness check, without altering the outcome of the performance evaluation.

3.2 Real-world SARIMA Processes

In a second exercise, the simulations are repeated with real-world SARIMA processes, which are not consistent with the theoretical assumptions of the methods. The processes are based on real series, similar to the ones used for quarterly GDP estimation. Because the relationship is modeled on a quarterly base, series have to be found that are available on a quarterly base (Table 2, Step 1). Table 3 shows the five indicator series that are used as a series template in the experiment. They include processes that reflect both economic activity and prices. Also, they include processes with a strong seasonal pattern and processes without a seasonal pattern.

In a next step, the pairs of series are regressed on each other (see Table 11 in the appendix). The regressions are used both to optain estimated coefficients of the series relationship (Table 2, Step 2) and to optain residuals that serve as a template for the SARIMA modeling of the error term (Table 2, Step 3). Of the five error term processes, three SARIMA models are stationary, two are not. Three pairs of series are thus cointegrated, two are not. All error terms also have a seasonal structure.

SARIMA modeling has been done by an automatic procedure implemented in the R-package *forecast* (Hyndman *et al.*, 2014). The degree of non-seasonal integration has been chosen by a KPSS test (Kwiatkowski *et al.*, 1992), the degree of seasonal integration by a OCBS test (Osborn *et al.*, 1988). The number of both seasonal and non-seasonal AR and MA terms has been chosen such that the AICc criterion is minimized. The coefficients are estimated by maximum likelihood.

For simulations, bootstrapped errors of the true series have been used. As a robustness check, calculations were repeated with normally distributed errors, with virtually no effect on the results. Each series has been simulated 1000 times. The length of the simulated time series is equal to the length of the template series, as they represent typical use cases for quarterly GDP estimation (see Table 11 for the number of observations in each pair of series). Temporal disaggregation has been performed by the R-package *tempdisagg* (Sax and Steiner, 2014).

4 Results

4.1 Autoregressive Models of Order 1

A central finding of the experiment is that there is no trade-off between quarterly and annual forecast performance: Whether one looks at the accuracy of the β_1-Coefficient or the annual, the quarterly or the in-sample forecasts, the outcome of the forecast competition is the same. In the following, it is generally referred to the annual forecast accuracy reported in Table 4, but the findings are fully valid for the β_1 accuracy, the quarterly and the in-sample forecast accuracy, which can be found in the appendix. (The single exception concerning the performance of the max-log algorithm is discussed below.) Accuracy is measured by the root mean squared error (RMSE), but all findings in this paper also apply to mean absolute errors (not reported).

In line with expectations, the methods using the true autoregressive parameter (ρ or ϕ) are performing best. This serves as a basic sanity check both for the disaggregation methods and for the simulation design. In practice, however, the true parameter is rarely known, nor is the degree of cointegration.

Of the automatic methods, Chow-Lin with a truncated max-log algorithm performs best for cointegrated series, while Litterman with an unrestricted max-log algorithm performs best for non-cointegrated series. However, unless the first difference autoregressive parameter, ϕ, is not close to 1, Chow-Lin with a truncated max-log algorithm still performs well in the case of non-cointegrated series. In contrast, Litterman with a non-truncated max-log algorithm performs badly for cointegrated series.[1]

Is there a method that is generally preferable, if there is no a priori knowledge on the degree of cointegration? This depends on the frequency of the actual occurrence of the of cointegrated and non-cointegrated series: If there are mostly cointegrated series to analyze, the Chow-Lin method with a truncated max-log algorithm generally out-performs other automated methods. If there are mostly non-cointegrated series to analyze, the Litterman method with a non-truncated max-log algorithm isf best. However, if the frequency of

[1]The non-truncated max-log algorithm is still better than the truncated here, because negative estimation values of ϕ are a crude way to deal with cointegrated series in a non-cointegrated context. With a negative autoregressive parameter, errors are reverted in the following period, therby bringing the level of the series back to the long-term relationship.

Table 4: Annual forecast accuracy (RMSE) for AR1/RW processes

Methods	$\rho : -0.5$	$\rho : 0.0$	$\rho : 0.5$	$\rho : 0.85$	$\phi : 0.0$	$\phi : 0.5$	$\phi : 0.85$
Chow-Lin, $\rho : -0.5$	**1.62**	2.28	3.68	7.86	16.76	32.79	100.00
Chow-Lin, $\rho : 0.0$	1.62	**2.26**	3.57	7.31	15.04	29.38	89.26
Chow-Lin, $\rho : 0.5$	1.68	2.31	**3.49**	6.54	12.56	24.44	73.63
Chow-Lin, $\rho : 0.85$	2.12	2.81	3.95	**5.78**	7.84	14.74	40.74
Dynamic, $\rho : -0.5$	2.62	3.01	4.28	8.18	16.48	32.35	96.24
Dynamic, $\rho : 0.0$	1.62	2.26	3.57	7.31	15.04	29.38	89.26
Dynamic, $\rho : 0.5$	4.62	4.99	5.54	7.84	12.98	24.15	71.13
Dynamic, $\rho : 0.85$	11.78	12.06	12.27	13.01	13.68	18.06	39.23
Fernandez, $\phi : 0.0$	2.55	3.39	4.72	6.25	**6.32**	11.13	23.34
Litterman, $\phi : 0.5$	2.92	3.86	5.26	6.57	6.46	**10.84**	21.12
Litterman, $\phi : 0.85$	4.27	5.49	7.36	8.42	8.02	12.11	**18.26**
Chow-Lin, max. Log.	**1.63**	2.27	3.63	5.95	6.99	12.40	25.00
Chow-Lin, min. RSS.	1.73	2.36	**3.55**	**5.93**	8.39	15.38	36.86
Chow-Lin, max. Log. (≥ 0.2)	1.64	**2.27**	3.58	5.95	6.99	12.40	25.00
Dynamic, max. Log.	1.69	2.37	3.75	7.33	12.97	19.55	30.65
Dynamic, min. RSS.	1.72	2.51	4.23	8.83	12.82	18.70	28.74
Dynamic, max. Log. (≥ 0.2)	2.07	2.66	3.77	7.23	12.79	19.51	30.64
Litterman, max. Log.	2.44	3.25	4.58	6.28	6.48	11.15	18.92
Litterman, min. RSS.	4.53	5.79	7.97	9.39	9.22	14.13	20.59
Litterman, max. Log. (≥ 0.2)	2.63	3.49	4.83	6.34	**6.48**	**11.09**	**18.92**

the actual occurrence is around half-half, the Chow-Lin method with a truncated max-log algorithm performs substantially better than Litterman or Fernandez. This is because Chow-Lin max-log method performs reasonably well with non-cointegrated series, while Litterman or Fernandez perform not so well with cointegrated series.

Interestingly, Chow-Lin with a non-truncated max-log method is out-performed by the min-RSS approach for cointegrated series with a low value of ρ. In line with the findings of Ciammola *et al.* (2005), min-RSS performs substantially better when looking at in-sample accuracy. The relatively bad performance of the unrestricted max-log method for low ρ values is due to the fact that even with a positive ρ, in a few cases, the unrestricted max-log method produces a negative estimation close to -1. These negative estimations not only decrease the forecast and in-sample accuracy, they also lead to a higher volatility of the series.

A simple workaround is to truncate the estimation of the parameter to a low positive value, like 0.2. A positive value, rather than a value of 0, has the advantage that there will be no discrete steps in the final series. With truncation, the max-log method out-performs the min-RSS method for almost all values of ρ. Only for $\rho = 0.85$, the pseudo-estimation of the min-RSS approach is incidentally close to the true value, and min-RSS slightly out-performs the max-log approach.

Figure 1 sheds light into the causes behind the performance differences. It shows the distributions of the ρ estimation for different methods for different processes. Relatively unaffected by the process, the ρ estimates of the the min-RSS method are always around 0.6 to 0.8. This static behavior of the estimator has also in be found by Ciammola *et al.* (2005) and Miralles *et al.* (2003). This is also the reason why for $\rho = 0.85$, the min-RSS has out-performed the max-log approach.

On the other hand, the Chow-Lin max-log method generally provides a consistent estimator of the true value of the autoregressive parameter, as long as the series are cointegrated.

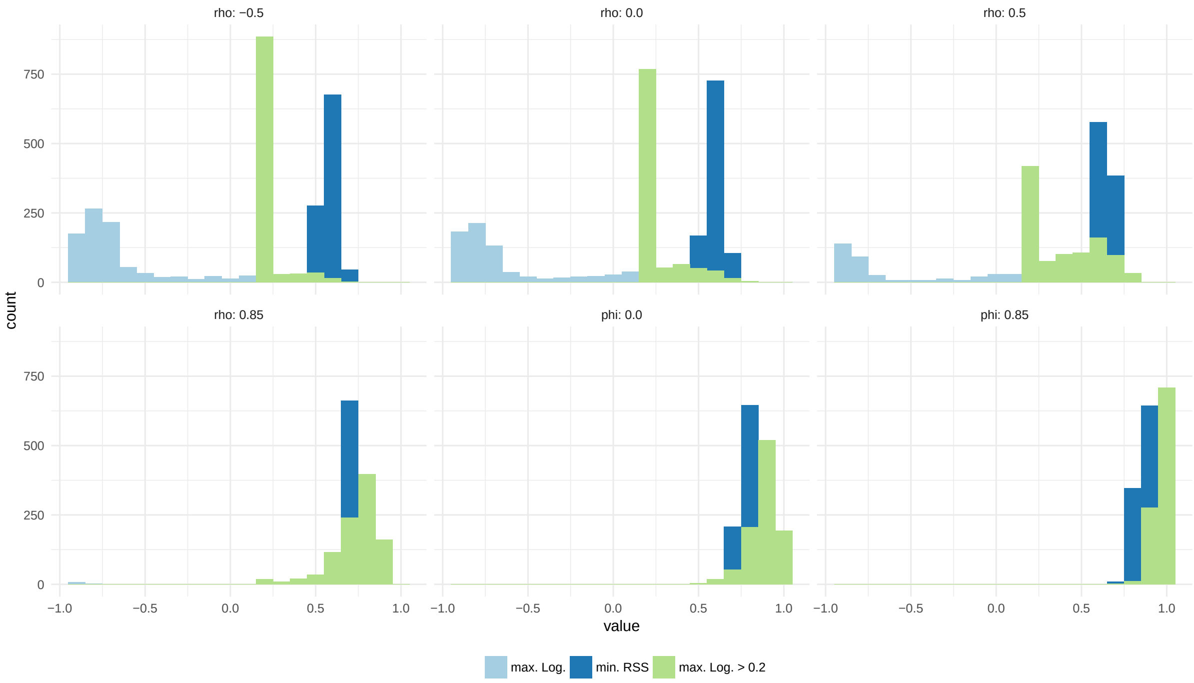

Figure 1: Frequency of estimated ρ for different AR1/RW processes

If the series are non-cointegrated, the parameter is estimated to be close to 1, and the results will be not far away from Fernandez. Thus, even with non-cointegrated series, the Chow-Lin max-log method performs relatively well. This is consistent with the findings of Silva and Cardoso (2001), who state that although the two methods are asymptotically equivalent, the maximum likelihood method is often preferred since it tends to lead to more reasonable estimates of the parameters.

As mentioned by Ciammola *et al.* (2005), a low true ρ causes the max-log method to occasionally generate negative estimations, which worsen the in-sample estimation accuracy substantially. However, with truncation, the values of ρ are limited to be at least 0.2. With this modification, With truncation, the max-log method out-performs the other methods most of the time.

5 Real-world SARIMA Processes

This section discusses the results of a Monte Carlos experiment, where the simulated series are *not* consistent with the theoretical assumptions of the methods. As with the simpler processes, it turns out that there is no trade-off between quarterly and annual forecast performance. The methods with the highest annual forecast accuracy also have the highest quarterly and in-sample accuracy. I will refer to the annual forecast accuracy, quarterly and in-sample results are found in the appendix.

Again, depending on whether the series are cointegrated or not, either Chow-Lin or Litterman performs best. Differences among the Litterman methods are relatively small, as long as the autoregressive parameter, ϕ, is not close to one. Among the Chow-Lin methods, a higher ρ works better if the series are non-cointegrated. If ρ is close to 1, the resulting series converges to the results of Fernandez.

Table 5: Annual forecast accuracy (RMSE) for real-world SARIMA processes

method	SAL	IMP	ENE	CON	SER
Chow-Lin, $\rho : -0.5$	24.30	2292.75	36.41	1229.45	10.63
Chow-Lin, $\rho : 0.0$	23.36	2146.53	33.56	1103.74	9.59
Chow-Lin, $\rho : 0.5$	**22.50**	1949.71	29.87	925.54	8.06
Chow-Lin, $\rho : 0.85$	24.50	**1820.46**	**26.36**	584.57	5.43
Dynamic, $\rho : -0.5$	24.40	2323.29	37.31	1208.60	10.52
Dynamic, $\rho : 0.0$	23.36	2146.53	33.56	1103.74	9.59
Dynamic, $\rho : 0.5$	23.60	2176.87	36.01	903.93	8.03
Dynamic, $\rho : 0.85$	30.85	3113.97	59.86	590.77	5.76
Fernandez, $\phi : 0.0$	28.73	2057.10	29.05	**450.09**	5.13
Litterman, $\phi : 0.5$	32.03	2227.98	32.33	463.78	**5.04**
Litterman, $\phi : 0.85$	44.78	2974.90	45.44	569.69	5.64
Chow-Lin, max. Log.	23.26	1868.95	27.25	488.65	**5.10**
Chow-Lin, min. RSS.	**22.58**	**1832.00**	**26.96**	599.78	5.73
Chow-Lin, max. Log. (≥ 0.2)	22.86	1863.54	27.17	488.65	5.10
Dynamic, max. Log.	23.65	2173.39	33.29	562.74	5.84
Dynamic, min. RSS.	30.20	2674.79	39.00	543.09	5.94
Dynamic, max. Log. (≥ 0.2)	23.18	2134.29	32.84	562.26	5.83
Litterman, max. Log.	27.86	2040.38	28.42	**457.37**	5.18
Litterman, min. RSS.	49.30	3230.50	48.52	645.53	6.48
Litterman, max. Log. (≥ 0.2)	29.42	2100.71	29.79	459.50	5.17

Of the automatic methods, Chow-Lin with a truncated max-log algorithm does a good job for cointegrated series, although it is slightly out-performed by the min-RSS method. With the chosen cointegrated processes, the relatively fixed ρ values that result from min-RSS may works to its advantage. However, fixing ρ at 0.85 would lead to even better results.

However, as soon as the series are non-cointegrated, the max-log method clearly out-performs min-RSS. The Litterman methods are slightly better for the non-cointegrated series, but they are substantially worse for cointegrated series.

Differences between the standard max-log and the truncated max-log methods are very small, because the residuals show a high degree of auto-correlation. Therefore, negative estimations of the ρ parameter are very rare, and truncation is barely applied.

Again, if there is no a priori knowledge on the degree of cointegration, the best method depends on the frequency of the occurrence of each type of series combinations. If the frequency is around half-half, the Chow-Lin method with a truncated max-log approach generally out-performs the other automated methods.

From the practical experience with Swiss quarterly GDP estimation, the ratio of cointegrated and non-cointegrated series leans towards cointegrated series. Based on economic judgment, visual and econometric residual analysis and out-of-sample forecast experiments, 14 indicator-series combinations of the production account have been classified as cointegrated, 6 as non-cointegrated. Formal cointegration tests are not very powerful with such a low number of observation, so the division in in cointegrated and non-cointegrated combinations of series is somewhat arbitrary. But even with a lower ratio of cointegrated to non-cointegrated series, the max-log Chow-Lin method is expected to out-perform Fernandez and Litterman.

6 Conclusions

Using Monte Carlos Simulations, this paper evaluates different methods for temporal disaggregation. First, time series are simulated that are consistent with the simplifying theoretical assumptions of the methods. The methods are evaluated by their annual and sub-annual forecast accuracy, by their in-sample accuracy, and by their coefficient estimation accuracy. Second, several real-world SARIMA processes are simulated and used for evaluation, both with cointegrated and non-cointegrated time series.

It is found that, first, Litterman and Fernandez perform best with non-cointegrated series (i.e. with non-stationary residuals), Chow-Lin with cointegrated series (i.e. with stationary residuals). This finding does not depend on whether the residuals follow the AR1 structure that is implied by the model or whether real-world SARIMA processes are considered.

Second, the max-log algorithm of determining the autoregressive parameter (Bournay and Laroque, 1979) is the only consistent estimation method. The alternative min-RSS approach (Barbone *et al.*, 1981) in its ECOTRIM implementation is not a consistent estimator. Rather it always leads to an estimation of the parameter of around 0.6 to 0.8.

Third, the max-log algorithm has the undesirable feature of sometimes leading to a strong negative estimation of the parameter (close to -1), even when the true parameter is positive. This leads to a small reduction in annual forecast accuracy and a substantial reduction in the sub-annual forecast and in-sample accuracy. Negative estimates are more frequent if the true parameter is close to 0. However, a simple truncation to positive values avoids the problem.

Forth, if there is no a priori knowledge on the degree of cointegration, the Chow-Lin method with the max-log algorithm for finding the autoregressive parameter leads to good results. If series are non-cointegrated, the method usually estimates a high autoregressive parameter, which makes the resulting series close to the Fernandez result.

Fifth, if there is a priori knowledge on the true autocorrelation of the residuals, the use of this knowledge will increase the quality of the estimation. Thus, economic knowledge, a visual and econometric analysis of the residuals and out-of-sample forecast evaluation may be used to adjust the estimation of the max-log algorithm.

Sixth, the dynamic Chow-Lin method under-performs the static methods in all settings. This may be due to fact that the simulation design favors the static methods, or that there are implementation issues in the R-package *tempdisagg* (Sax and Steiner, 2014). At the moment, the static methods, especially Chow-Lin in conjunction with a truncated max-log algorithm, seem to be be preferrable.

A Additional Tables: Results

A.1 AR1 Processes

Table 6: β_1 estimation accuracy (RMSE) for AR1/RW processes

Methods	$\rho : -0.5$	$\rho : 0.0$	$\rho : 0.5$	$\rho : 0.85$	$\phi : 0.0$	$\phi : 0.5$	$\phi : 0.85$
Chow-Lin, $\rho : -0.5$	**0.001204**	0.001853	0.003591	0.010183	0.04960	0.09957	0.31412
Chow-Lin, $\rho : 0.0$	0.001206	**0.001842**	0.003578	0.010016	0.04919	0.09877	0.31188
Chow-Lin, $\rho : 0.5$	0.001225	0.001845	**0.003574**	0.009751	0.04856	0.09754	0.30837
Chow-Lin, $\rho : 0.85$	0.001721	0.002309	0.004144	**0.009045**	0.04506	0.09063	0.28703
Dynamic, $\rho : -0.5$	0.249583	0.249643	0.249428	0.250886	0.26258	0.29080	0.54457
Dynamic, $\rho : 0.0$	0.001206	0.001842	0.003578	0.010016	0.04919	0.09877	0.31188
Dynamic, $\rho : 0.5$	0.249707	0.249692	0.249777	0.249509	0.25045	0.25541	0.29644
Dynamic, $\rho : 0.85$	0.424500	0.424492	0.424517	0.424422	0.42440	0.42502	0.42716
Fernandez, $\phi : 0.0$	0.009288	0.011526	0.016889	0.021488	**0.03164**	0.06072	0.16043
Litterman, $\phi : 0.5$	0.013383	0.016740	0.023183	0.026825	0.03278	**0.05849**	0.13885
Litterman, $\phi : 0.85$	0.029380	0.037065	0.048702	0.052369	0.05073	0.07380	**0.09609**
Chow-Lin, max. Log.	**0.001212**	0.001850	0.003607	0.009372	0.04089	0.07991	0.2049
Chow-Lin, min. RSS.	0.001244	0.001860	0.003605	**0.009355**	0.04611	0.09193	0.2830
Chow-Lin, max. Log. (≥ 0.2)	0.001213	**0.001844**	**0.003600**	0.009370	0.04089	0.07991	0.2049
Dynamic, max. Log.	0.040781	0.049809	0.085448	0.149233	0.22317	0.35423	0.4547
Dynamic, min. RSS.	0.046699	0.069895	0.145564	0.291548	0.37760	0.45043	0.4833
Dynamic, max. Log. (≥ 0.2)	0.099911	0.099961	0.102074	0.131481	0.21900	0.35137	0.4547
Litterman, max. Log.	0.008175	0.010185	0.015615	0.022157	**0.03313**	0.06205	0.1010
Litterman, min. RSS.	0.032685	0.041418	0.054873	0.060267	0.05849	0.08543	0.1024
Litterman, max. Log. (≥ 0.2)	0.010072	0.012556	0.018340	0.023619	0.03315	**0.06168**	**0.1009**

Table 7: Quarterly forcast accuracy (RMSE) for AR1/RW processes

Methods	$\rho : -0.5$	$\rho : 0.0$	$\rho : 0.5$	$\rho : 0.85$	$\phi : 0.0$	$\phi : 0.5$	$\phi : 0.85$
Chow-Lin, $\rho : -0.5$	**1.1489**	1.0720	1.2804	2.2480	4.4900	8.7369	26.4385
Chow-Lin, $\rho : 0.0$	1.1541	**1.0554**	1.2023	1.9949	3.8485	7.4408	22.4496
Chow-Lin, $\rho : 0.5$	1.1643	1.0650	**1.1806**	1.8173	3.2771	6.2879	18.8190
Chow-Lin, $\rho : 0.85$	1.2061	1.1379	1.2684	**1.6473**	2.1458	3.9164	10.7014
Dynamic, $\rho : -0.5$	1.8195	1.7971	1.9457	2.6844	4.6014	8.7104	25.4729
Dynamic, $\rho : 0.0$	1.1541	1.0554	1.2023	1.9949	3.8485	7.4408	22.4496
Dynamic, $\rho : 0.5$	1.8615	1.8062	1.8703	2.3278	3.5235	6.2820	18.1966
Dynamic, $\rho : 0.85$	3.5353	3.5347	3.5683	3.7190	3.8924	4.9513	10.3798
Fernandez, $\phi : 0.0$	1.2565	1.2325	1.4320	1.7572	**1.7713**	2.9994	6.2139
Litterman, $\phi : 0.5$	1.3069	1.3175	1.5506	1.8338	1.8030	**2.9302**	5.6642
Litterman, $\phi : 0.85$	1.5309	1.6568	2.0476	2.2909	2.1873	3.2614	**4.9214**
Chow-Lin, max. Log.	1.1801	1.1192	1.2890	1.6882	1.9328	3.3217	6.6522
Chow-Lin, min. RSS.	1.1686	1.0732	**1.1917**	**1.6817**	2.2785	4.0735	9.7101
Chow-Lin, max. Log. (≥ 0.2)	**1.1579**	**1.0589**	1.1995	1.6849	1.9317	3.3215	6.6522
Dynamic, max. Log.	1.1715	1.0970	1.3246	11488.6691	14.6053	5.7802	8.2246
Dynamic, min. RSS.	1.1825	1.1252	1.4470	2.5994	3.6324	5.1249	7.7819
Dynamic, max. Log. (≥ 0.2)	1.2715	1.1785	1.3028	2.0397	3.4385	5.2171	8.2231
Litterman, max. Log.	1.2425	1.2084	1.4026	1.7658	1.8076	3.0078	5.0981
Litterman, min. RSS.	1.5800	1.7246	2.2017	2.5442	2.4953	3.8021	5.5616
Litterman, max. Log. (≥ 0.2)	1.2665	1.2496	1.4566	1.7796	**1.8073**	**2.9939**	**5.0978**

Table 8: In-sample forcast accuracy (RMSE) for AR1/RW processes

Methods	$\rho : -0.5$	$\rho : 0.0$	$\rho : 0.5$	$\rho : 0.85$	$\phi : 0.0$	$\phi : 0.5$	$\phi : 0.85$
Chow-Lin, $\rho : -0.5$	**1.0699**	0.9324	0.9472	1.1995	1.6854	3.1173	8.5811
Chow-Lin, $\rho : 0.0$	1.0862	**0.8972**	0.7993	0.7981	0.8338	1.2100	2.4200
Chow-Lin, $\rho : 0.5$	1.0992	0.9068	**0.7873**	0.7616	0.8528	1.2720	3.0822
Chow-Lin, $\rho : 0.85$	1.1039	0.9167	0.7945	**0.7454**	0.7515	0.9714	1.9436
Dynamic, $\rho : -0.5$	1.6953	1.6264	1.6312	1.7657	2.1659	3.3550	8.3926
Dynamic, $\rho : 0.0$	1.0862	0.8972	0.7993	0.7981	0.8338	1.2100	2.4200
Dynamic, $\rho : 0.5$	1.4887	1.3413	1.2693	1.2599	1.2653	1.5932	3.0519
Dynamic, $\rho : 0.85$	1.9368	1.8189	1.7680	1.7506	1.7114	1.8269	2.2296
Fernandez, $\phi : 0.0$	1.1057	0.9229	0.8063	0.7534	**0.7156**	0.8558	1.2141
Litterman, $\phi : 0.5$	1.1134	0.9369	0.8215	0.7657	0.7204	**0.8415**	1.0879
Litterman, $\phi : 0.85$	1.1288	0.9692	0.8650	0.8169	0.7625	0.8813	**0.9596**
Chow-Lin, max. Log.	1.1025	1.0167	0.9931	7.796e-01	0.7365	0.9143	1.3282
Chow-Lin, min. RSS.	1.1000	0.9087	**0.7883**	**7.491e-01**	0.7639	0.9980	1.8171
Chow-Lin, max. Log. (≥ 0.2)	1.0934	**0.9012**	0.7901	7.506e-01	0.7361	0.9077	1.3282
Dynamic, max. Log.	1.1077	0.9178	0.9626	1.153e+04	15.4080	3.5073	2.0383
Dynamic, min. RSS.	1.1108	0.9430	1.0013	1.363e+00	1.5637	1.8689	2.0350
Dynamic, max. Log. (≥ 0.2)	1.1755	0.9939	0.9024	9.587e-01	1.1467	1.6741	2.0383
Litterman, max. Log.	**1.0914**	0.9116	0.8034	7.609e-01	0.7244	0.8583	**0.9807**
Litterman, min. RSS.	1.1308	0.9737	0.8743	8.366e-01	0.7868	0.9226	1.0081
Litterman, max. Log. (≥ 0.2)	1.1080	0.9266	0.8100	7.559e-01	**0.7192**	**0.8472**	0.9807

A.2 SARIMA Processes

Table 9: Quarterly forcast accuracy (RMSE) for real-world SARIMA processes

method	SAL	IMP	ENE	CON	SER
Chow-Lin, $\rho : -0.5$	7.6454	696.5721	10.6360	650.8159	2.9524
Chow-Lin, $\rho : 0.0$	6.9988	623.7803	9.4334	628.4102	2.5103
Chow-Lin, $\rho : 0.5$	**6.7371**	577.4597	8.6224	611.1416	2.1537
Chow-Lin, $\rho : 0.85$	7.1800	**550.4638**	**7.8246**	581.1396	1.5432
Dynamic, $\rho : -0.5$	11.9260	764.2745	11.9401	675.4285	2.9333
Dynamic, $\rho : 0.0$	6.9988	623.7803	9.4334	628.4102	2.5103
Dynamic, $\rho : 0.5$	9.5115	657.5837	11.3655	615.6806	2.1496
Dynamic, $\rho : 0.85$	12.7887	905.1427	17.8657	599.7195	1.6345
Fernandez, $\phi : 0.0$	8.2684	605.8718	8.4199	**568.1381**	1.4724
Litterman, $\phi : 0.5$	9.1177	644.6135	9.1589	569.0699	**1.4513**
Litterman, $\phi : 0.85$	12.7404	827.1712	12.3046	578.4774	1.5956
Chow-Lin, max. Log.	7.4224	570.9771	8.1176	571.9314	**1.4670**
Chow-Lin, min. RSS.	**6.7491**	**552.3826**	**7.9838**	582.1777	1.6142
Chow-Lin, max. Log. (≥ 0.2)	6.8289	559.2191	8.0196	571.9314	1.4670
Dynamic, max. Log.	17342.1078	1375529.3641	147.3488	602.7083	1.6514
Dynamic, min. RSS.	12.4223	793.3507	12.2928	598.4845	1.6841
Dynamic, max. Log. (≥ 0.2)	8.1463	631.7043	9.6640	595.5564	1.6499
Litterman, max. Log.	8.0576	602.3060	8.2805	**568.1722**	1.4853
Litterman, min. RSS.	14.1209	892.2990	13.0784	585.5663	1.8081
Litterman, max. Log. (≥ 0.2)	8.4407	615.5106	8.5836	568.4113	1.4827

Table 10: In-sample forcast accuracy (RMSE) for real-world SARIMA processes

method	SAL	IMP	ENE	CON	SER
Chow-Lin, $\rho : -0.5$	5.0000	419.3455	5.9172	562.4595	1.3547
Chow-Lin, $\rho : 0.0$	3.8311	311.1561	4.0920	551.5605	0.7346
Chow-Lin, $\rho : 0.5$	**3.6774**	300.0365	3.9564	551.5800	0.7132
Chow-Lin, $\rho : 0.85$	3.6888	**296.3829**	**3.9490**	548.1893	0.6551
Dynamic, $\rho : -0.5$	10.5435	507.8430	7.9119	594.6425	1.3744
Dynamic, $\rho : 0.0$	3.8311	311.1561	4.0920	551.5605	0.7346
Dynamic, $\rho : 0.5$	7.3543	357.1736	6.8910	559.3013	0.7114
Dynamic, $\rho : 0.85$	10.0849	424.2590	9.9077	567.9794	0.6747
Fernandez, $\phi : 0.0$	3.9803	299.6846	4.0589	**544.0352**	0.6441
Litterman, $\phi : 0.5$	4.2625	303.7124	4.2098	544.3829	**0.6404**
Litterman, $\phi : 0.85$	5.7556	320.8228	4.5959	548.3197	0.6491
Chow-Lin, max. Log.	4.9903	335.7785	4.2621	545.1921	0.6471
Chow-Lin, min. RSS.	**3.6642**	**297.5838**	**3.9445**	548.2795	0.6624
Chow-Lin, max. Log. (≥ 0.2)	3.7155	299.6613	3.9735	545.1923	0.6471
Dynamic, max. Log.	17412.5194	1381045.3978	152.9054	573.3726	0.6755
Dynamic, min. RSS.	9.7650	397.7578	7.3482	569.9889	0.6762
Dynamic, max. Log. (≥ 0.2)	5.6779	328.0218	4.8367	565.7545	0.6752
Litterman, max. Log.	3.9580	303.1685	4.0551	**543.6532**	0.6430
Litterman, min. RSS.	6.4221	325.5364	4.6525	550.4806	0.6578
Litterman, max. Log. (≥ 0.2)	4.0317	300.5810	4.0973	543.9688	**0.6419**

B Additional Tables: Simulation Set Up

Table 11: OLS Regression in Levels (x_t)

	SAL	IMP	ENE	CON	SER
(Intercept)	12.80***	−3550.04***	64.66***	−1111.74**	39.09***
	(0.91)	(219.90)	(3.11)	(385.75)	(3.01)
Exports of Chemicals and Pharma	13.31***				
	(0.10)				
Total Imports		0.28***			
		(0.01)			
Consumer Prices Oil			2.23***		
			(0.04)		
Construction Employment				11.15***	
				(1.31)	
Consumer Prices Total					0.62***
					(0.03)
R^2	0.99	0.96	0.97	0.44	0.79
Adj. R^2	0.99	0.96	0.97	0.44	0.79
Num. obs.	145	94	94	93	94

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 12: OLS Regression in seasonal Log-Differences ($\log(x_t) - \log(x_{t-4})$)

	SAL	IMP	ENE	CON	SER
(Intercept)	0.00	0.04***	0.01*	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)
Exports of Chemicals and Pharma	0.77***				
	(0.06)				
Total Imports		0.72***			
		(0.09)			
Consumer Prices Oil			0.58***		
			(0.02)		
Construction Employment				1.66***	
				(0.25)	
Consumer Prices Total					0.51***
					(0.11)
R^2	0.53	0.43	0.88	0.33	0.20
Adj. R^2	0.52	0.42	0.88	0.33	0.19
Num. obs.	141	90	90	89	90

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

References

Barbone L, Bodo G, Visco I (1981). "Costi e profitti nell'industria in senso stretto: un'analisi su serie trimestrali, 1970-1980." *Bolletino della Banca d'Italia*, pp. 467–510.

Bournay J, Laroque G (1979). "Réflexions sur la méthode d'élaboration des comptes trimestriels." *Annales de l'INSÉÉ*, (36), 3–30.

Chow GC, Lin AL (1971). "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series." *The Review of Economics and Statistics*, **53**(4), 372–375.

Ciammola A, Di Palma F, Marini M (2005). "Temporal disaggregation techniques of time series by related series: A comparison by a Monte Carlo experiment." *Technical report*, Eurostat, Working Paper.

Fernández RB (1981). "A Methodological Note on the Estimation of Time Series." *The Review of Economics and Statistics*, **63**(3), 471–476.

Hyndman RJ, Athanasopoulos G, Razbash S, Schmidt D, Zhou Z, Khan Y, Bergmeir C, Wang E (2014). *forecast: Forecasting functions for time series and linear models*. R package version 5.4, URL `http://CRAN.R-project.org/package=forecast`.

Kwiatkowski D, Phillips PC, Schmidt P, Shin Y (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of econometrics*, **54**(1), 159–178.

Litterman RB (1983). "A Random Walk, Markov Model for the Distribution of Time Series." *Journal of Business Economic Statistics*, **1**(2), 169–173.

Miralles JMP, Lladosa LEV, Vallés RE (2003). "On the performance of the Chow-Lin procedure for quarterly interpolation of annual data: Some Monte-Carlo analysis." *Spanish Economic Review*, **5**(4), 291–305.

Osborn DR, Chui AP, Smith JP, Birchenhall CR (1988). "Seasonality and the order of integration for consumption." *Oxford Bulletin of Economics and Statistics*, **50**(4), 361–377. URL `http://www.est.uc3m.es/esp/nueva_docencia/comp_col_get/lade/tecnicas_prediccion/OCSB_OxBull1988.pdf`.

Proietti T (2006). "Temporal disaggregation by state space methods: Dynamic regression methods revisited." *Econometrics Journal*, **9**(3), 357–372. URL `http://ideas.repec.org/a/ect/emjrnl/v9y2006i3p357-372.html`.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Sax C (2017). *tdmc: Reproducible Monte-Carlo Evaluation of Regression-Based Temporal Disaggregation Methods*. R package version 0.0.1, URL `https://github.com/christophsax/tdmc`.

Sax C, Steiner P (2013). "Temporal Disaggregation of Time Series." *The R Journal*, **5**(2), 80–88. URL `http://journal.r-project.org/archive/2013-2/sax-steiner.pdf`.

Sax C, Steiner P (2014). *tempdisagg: Methods for Temporal Disaggregation and Interpolation of Time Series*. R package version 0.24.9, URL `http://journal.r-project.org/archive/2013-2/sax-steiner.pdf`.

Silva JS, Cardoso F (2001). "The Chow-Lin method using dynamic models." *Economic Modelling*, **18**, 269–280.