

MPRA

Munich Personal RePEc Archive

Temporal disaggregation by dynamic regressions: recent developments in Italian quarterly national accounts

Laura Bisio and Filippo Moauro

ISTAT - Italian National Institute of Statistics

14 July 2017

Online at <https://mpra.ub.uni-muenchen.de/80211/>
MPRA Paper No. 80211, posted 17 July 2017 16:37 UTC

Temporal disaggregation by dynamic regressions: recent developments in Italian quarterly national accounts

Laura Bisio¹ e Filippo Moauro²

Abstract

In this paper we discuss the most recent developments of temporal disaggregation techniques carried out at ISTAT. They concern the extension from static to dynamic autoregressive distributed lag ADL regressions and the change to a state-space framework for the statistical treatment of temporal disaggregation. Beyond the development of a unified procedure for both static and dynamic methods from one side and the treatment of the logarithmic transformation from the other, we provide short guidelines for model selection. From the empirical side we evaluate the new dynamic methods by implementing a large scale temporal disaggregation exercise using ISTAT annual value added data jointly with quarterly industrial production by branch of economic activity over the period 1995-2013. The main finding of this application is that ADL models either in levels and logarithms can reduce the errors due to extrapolating disaggregated data in last quarters before the annual benchmarks become available. When the attention moves to the correlations with the high-frequency indicators the ADL disaggregations are also generally in line with those produced by the static Chow-Lin variants, with problematic outcomes limited to few cases.

Keywords: temporal disaggregation; state-space form; Kalman filter; ADL models; linear Gaussian approximating model; quarterly national accounts.

1. Introduction

Since mid-eighties, when Italian quarterly national accounts releases became systematic, temporal disaggregation methods gathered larger attention as a decisive tool for their production. At that time ISTAT adopted new information technology instruments borrowing most of the work already developed at the Bank of Italy. Temporal disaggregation methods like both the Chow and Lin (1971) solution – see the development by Barbone et al. (1981) – and the approach by Denton (1971) were adopted for estimating quarterly national accounts. A second renovating phase dates back to mid-nineties when a critical analysis of temporal disaggregation methods by Lupi and Parigi (1996) came out and which inspired the development of a sophisticated procedure for diagnostic checking. Such procedure still largely

¹ ISTAT researcher, e-mail: bisio@istat.it.

² ISTAT first researcher, e-mail: moauro@istat.it.

supports the operational phases of the quarterly accounts process since it offers a complete diagnostic report of quarterly disaggregations.

Later on, between 2004 and 2005, an ISTAT study commission was set up with the task of formulating new proposals for temporal disaggregation. The final remarks in Di Fonzo (2005) provided evidence of the ISTAT commitment to modernize both conceptual and technical tools used within quarterly accounts, largely implemented in the following years.

Most recently the literature has proposed further methodological developments mainly due to the initiative by Eurostat. See Frale et al. (2010 e 2011), Grassi et al. (2014) e Moauro (2014), among others. At the same time ISTAT favoured a relatively more pragmatic approach. In particular, the effort was aimed at technically implementing temporal disaggregation methods based on autoregressive distributed lag ADL models according to Proietti (2005).

Such extension, on one hand significantly broadened the range of the available models to be used for temporal disaggregation, on the other, given it is based on the *Kalman filter* (Kalman, 1960), entailed a number of practical benefits for quarterly accounts analysis. Namely, the computation of innovations, the development of diagnostics concerning the extrapolations and the resorting to models based on the logarithmic transformation of the data to be disaggregated.

The present work describes the innovative elements of temporal disaggregation recently introduced at ISTAT concerning the enlargement of the range of models to be selected and the development of statistics for diagnostic checking. These tools are largely used in the empirical section of the paper where we present the results of an extensive temporal disaggregation experiment. We compare the performances of the enlarged class of models, providing some guidelines for model selection and highlighting the critical points. Advantages and disadvantages of alternative solutions are discussed by taking into account the main features of the exercise.

This paper is structured as follows: section 2 describes the analytics of the reference ADL(1,1) model and its linkage to the static regression setup; the main features of the state space representation including the case of logarithmic transformation, diagnostic checking and temporal disaggregation evaluation are discussed in section 3; section 4 presents the results of the empirical experiment and section 5 shortly concludes. In the appendix section, appendix A includes a set of tables describing the features of the dataset and the main results of the application; appendix B is devoted to the graphs of the dataset.

2. Dynamic regression methods

2.1 The ADL(1,1) model

Temporal disaggregation methods here discussed are based upon dynamic regression models. They encompass a linear univariate relationship between the

dependent variable y_t , its lagged values y_{t-1} and a series of regressors x_t in a given time span $t = 1, \dots, T$. The problem is that y_t is available only as a temporal aggregate Y_t over s periods, i.e. only annual values are available over the quarterly time span. In case the annual aggregate value results as the sum of quarters, Y_t is defined as $Y_t = y_t + \dots + y_{t-s+1}$ with $s=4$. Alternatively $Y_t = (y_t + \dots + y_{t-s+1})/s$ reflecting the case of averaged stocks. Hence, Y_t is only observed in periods $t=s, 2s, \dots, [T/s]$, where $[T/s]$ is the biggest whole number of the ratio T/s . On the other hand, the k covariates $x_t = (x_{1t}, \dots, x_{kt})$ are observable at each quarter $t=1, \dots, T$.

A general representation of the relation between the two sets of variables is given by the dynamic autoregressive distributed lag models ADL(1,1) that is specified at the higher frequency as:

$$\Delta^1 y_t = \phi \Delta^1 y_{t-1} + m + gt + \Delta^1 x'_t \beta_0 + \Delta^1 x'_{t-1} \beta_1 + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2). \quad (1)$$

In equation (1) Δ is the difference operator such that $\Delta y_t = y_t - y_{t-1}$; 1 represents the differencing order assuming here either a value of 0, i.e. no difference, or 1 for first order difference; ϕ is the autoregressive term such that $-1 < \phi < 1$; m and gt are the deterministic components i.e., respectively, a constant term and a linear trend; β_0 and β_1 are the regression coefficients vectors at lag 0 and 1, respectively; ϵ_t is the vector of stochastic errors for which a normal distribution with zero mean and constant variance equal to σ^2 is assumed.

The ADL(1,1) model was made popular by Hendry e Mizon (1978) who pointed out that the stability of the model would hold even if the regressors determined a spurious relationship in level and were uncorrelated in differences. ADL(1,1) models in the case of $I=0$ represent a simple reparametrization of error correction models, made popular by the literature on cointegration.

Within the domain of temporal disaggregation Proietti (2005) suggested a methodology based both on the parametrization of ADL(1,1) models in the state space form (SSF) and on the use of the Kalman filter for its statistical treatment. In particular the use of the Kalman filter is intended for: log-likelihood computation, model parameters estimation, high-frequency (e.g. quarterly) distribution or temporal disaggregation of data observed as the sum/average in a lower frequency time span (e.g. annually) and the extension to the non-linear temporal disaggregation in case of logged data. Concerning maximum likelihood estimation of model parameters of equation (1), the most appropriate solution is the generalised least squares (GLS) method. Indeed, all model regression coefficients, i.e. m, g, β_0, β_1 , the variance term σ^2 of ϵ_t residuals can be concentrated out of the log-likelihood function, thereby originating a profile likelihood depending only from the autoregressive parameter ϕ . Its estimation can be conveniently set up as a grid search of ϕ over the interval $(-1, 1)$.

Either regression models with AR(1) residuals - including I(1) models - and ARIMA (1,1,0) models are nested in the ADL(1,1) models of equation (1); within the temporal disaggregation domain these specific cases correspond, respectively, to the Chow-Lin (1971), Fernández (1981) and Litterman (1983) methods.

2.2 From ADL(1,1) to AR(1) Chow-Lin model

Under suitable hypothesis on initial conditions and under proper linear restrictions on regressor parameters, the model ADL(1,1) nests a linear regression model whose residuals follow an AR(1) process. For example, model (1) in levels without deterministic components can be rewritten in terms of the lag polynomial $\beta_0 + \beta_1 L$ such that L is the lag operator for which $Lx_t = x_{t-1}$:

$$y_t = \phi y_{t-1} + x'_t(\beta_0 + \beta_1 L) + \epsilon_t.$$

Under this form and given the condition $\beta_1 = -\phi \beta_0$ with $|\phi| < 1$, it becomes

$$y_t(1 - \phi L) = x'_t \beta_0(1 - \phi L) + \epsilon_t$$

and, therefore

$$\begin{aligned} y_t &= x'_t \beta_0 + \alpha_t \\ \alpha_t &= \phi \alpha_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2) \end{aligned}$$

where the residual term α_t follows a first order autoregressive stationary process.

2.3 From ADL(1,1) in differences to the Litterman (1983) and Fernández (1981) models

Under suitable initial conditions reflecting the non-stationarity of ADL(1,1) model in differences, both Fernández (1981) and Litterman (1983) models also can be derived from model (1). They will result as linear regression models with residual following, for the former, a random walk or I(1) process and, for the latter, an ARIMA (1,1,0) process.

In formulas, when $I=1$ then:

$$\Delta y_t = \phi \Delta y_{t-1} + \Delta x'_t \beta_0 + \Delta x'_{t-1} \beta_1 + \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma^2) \quad (2)$$

That, if $\phi = 0$ and $\beta_1 = 0$, corresponds to the Fernández model, i.e.:

$$\Delta y_t = \Delta x'_t \beta_0 + \epsilon_t,$$

that is:

$$\begin{aligned} y_t &= x'_t \beta_0 + u_t \\ u_t &= u_{t-1} + \epsilon_t \end{aligned}$$

where u_t follows a random walk process.

Under the condition $\beta_1 = -\phi \beta_0$, with $\phi < 1$ model in differences (2), nests the Litterman model such that:

$$\begin{aligned} \Delta y_t &= \phi \Delta y_{t-1} + \Delta x'_t \beta_0 - \Delta x'_{t-1} \phi \beta_0 + \epsilon_t, & \epsilon_t &\sim \text{WN}(0, \sigma^2) \\ y_t &= x'_t \beta_0 + u_t \\ \Delta u_t &= \phi \Delta u_{t-1} + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \sigma^2) \end{aligned}$$

Where u_t follows an ARIMA(1,1,0) process.

3 Statistical treatment and diagnostic checking

3.1. Essentials of the state space representation

In general, the state space representation within the temporal disaggregation domain is defined by two equations: the former defines the time series structure (measurement equation), the latter how the latent structural components evolve from one state to the following one (transition equation). The SSF representation allows to resort to the Kalman filter methodology which in turn allows to compute the optimal estimator of the state variables vector at the time t for $t=1, \dots, T$, given the information available by the same horizon. The Kalman filter is usually associated to a smoothing algorithm which allows to optimally estimate the state vector conditioned to the whole information set.

The application of the SSF approach to temporal disaggregation was formerly introduced by Harvey and Pierce (1984) and then developed by Harvey (1989), Harvey and Chung (2000), Harvey and Koopman (1997) and Moauro and Savio (2002), among others. The peculiarities of the SSF representation applied to the temporal disaggregation techniques have been subsequently treated by Proietti (2005; 2006) whose contributions are the essential basis of this work.

Advantages of the SSF are the following: i) a suitable treatment of the initial conditions in presence of a non-stationary time series; ii) the availability of more effective diagnostics oriented to evaluate the quality of maximum likelihood estimates like the innovations; iii) the chance to easily obtain extrapolations of the series in case of models without covariates. Among disadvantages we find a relatively larger complexity due to the Kalman filter which in some environment implies slower computations.

According to Harvey (1989, sec. 6.3) temporal disaggregation traces back to a “missing observations” problem which is appropriately treated by augmenting the SSF representation of a general model, and therefore the ADL model of equation (1), by a cumulator variable y_t^C observable only at time $t = s, 2s, 3s, \dots$. For quarterly series of flows $s = 4$ and y_t^C is such that:

$$\begin{aligned} y_1^C &= y_1, & y_2^C &= y_1 + y_2, & y_3^C &= y_1 + y_2 + y_3, & y_4^C &= y_1 + y_2 + y_3 + y_4 \\ y_5^C &= y_5, & y_6^C &= y_5 + y_6, & y_7^C &= y_5 + y_6 + y_7, & y_8^C &= y_5 + y_6 + y_7 + y_8 \\ &\dots & & & & & & \end{aligned}$$

or in Markovian terms $y_t^C = \psi_t y_{t-1}^C + y_t$, where ψ_t is such that

$$\psi_t = \begin{cases} 0, & t = 1, 5, \dots \\ 1, & \text{otherwise} \end{cases}$$

As far as the statistical treatment is concerned, the required steps are: the cumulator variable y_t^C is added to the state vector of the SSF of the model defined at the highest frequency of observation; the measurement equation is adjusted so that the Kalman filter could take into account the missing observations of y_t^C ; then, the likelihood function of the given model is computed, its maximization with respect to the unknown parameters vector is carried out and both missing observations and disaggregated data are estimated through the smoothing algorithm. For full details see Proietti (2005).

3.2 The case of log-transformed series

The logarithmic transformation of data is a common practice in time series econometrics, and in particular when the series refer to variables defined as the ratio of flow aggregates. (Proietti, 2005). Applying the logarithmic transformation to the series implies a number of well-known advantages such as the downsizing of the series volatility, or the larger plausibility of the hypothesis underpinning the regression model (model linearity, errors homoscedasticity and normality). Within the logarithmic context, both the temporal aggregation constraints that must hold for the disaggregated (unknown) series and the constraints represented by the cumulator variable are defined in non-linear terms. For the sake of clarity, we can consider the following relationship holding between the logarithmic transformation of the disaggregated series y_t and the correspondent aggregated series Y_τ :

$$Y_\tau = \sum_{j=0}^{s-1} \exp(y_{\tau s - j}), \quad \tau = 1, \dots, \left\lfloor \frac{T}{s} \right\rfloor. \quad (3)$$

with the cumulator variable becoming $y_t^c = \psi_t y_{t-1}^c + \exp(y_t)$.

Disaggregated data \hat{y}_t are computed applying an iterative method converging towards the constrained posterior mode estimate of the unknown solution which satisfies exactly the restrictions of equation (3). Given a trial initial estimate \tilde{y}_t (e.g. a series of ones) of y_t , iterations start from and the first order Taylor approximation of $\exp(y_t)$ around that trial estimate. This allows to expand the SSF defined for the linear disaggregation case to a linear Gaussian approximating model (LGAM) for log-transformed data. In a second step running the Kalman filter and smoother of the LGAM computed at the first step produces a first disaggregated series \hat{y}_t . Then $\tilde{y}_t = \hat{y}_t$ is set and a new LGAM is computed producing a second disaggregation \hat{y}_t . This process is iterated until convergence, which usually requires not more than 6-7 rounds.

3.3 Test statistics and diagnostic checking of temporal disaggregation

Among the main features of the SSF representation and the Kalman filter there is the estimation of forecasting errors or innovations as a by-product of the application of the Kalman algorithm. In particular, the main diagnostic statistics implemented within the new procedures rely upon standardized innovations \hat{v}_t which, in temporal disaggregation problems, assume real values at time $t=s, 2s, \dots, [T/s]$ but are missing otherwise. Innovations v_t are such that $v_t = y_t^c - E(y_t^c | I_{t-1})$ where $I_{t-1} = \{y_{t-1}, x_t\}$ is the information set of the lagged dependent variables y_{t-1} and the exogenous regressor x_t . Therefore standardized innovations are defined as

$$\hat{v}_t = \frac{v_t}{\sqrt{f_t}}$$

where f_t are the estimated variances of v_t for $t = 1, \dots, T$ also computed by the Kalman filter.

Two remarks: first, both for static and dynamic models, the whole set of statistics resulting by the standardized innovations \hat{v}_t are consistent to the standard regression formulas. For an exhaustive and comprehensive discussion on this topic see Harvey (1989, sec.5.4 p. 256 ff.). Second, within temporal disaggregation, the innovations measure the one-step-ahead forecast error over the lowest frequency of observation, as v_t cumulates the errors from the first to the s^{th} sub-period. In other terms the s extrapolations $\hat{y}_{t+1/t}, \dots, \hat{y}_{t+s/t+s-1}$ for $t=s, 2s, \dots, [T/s]$ are such that

$$\begin{aligned} \hat{y}_{t+1/t} &= E(y_{t+1}^c | I_t) \\ \hat{y}_{t+2/t+1} &= E(y_{t+2}^c | I_{t+1}) = E(y_{t+2}^c | I_t) = \hat{y}_{t+2/t} \\ &\dots \\ \hat{y}_{t+s/t+s-1} &= E(y_{t+s}^c | I_{t+s-1}) = E(y_{t+s}^c | I_t) = \hat{y}_{t+s/t} \end{aligned}$$

and therefore, apart the regressors x_t that in this context are exogenous, conditional to the last available low-frequency aggregate Y_t .

A first common statistic is the determination coefficient R^2 and its corrected formula R_c^2 respectively defined as:

$$R^2 = 1 - SSR/SST, \quad (4)$$

$$R_c^2 = \frac{1-k}{[T/s]-1} + \frac{[T/s]-1}{[T/s]-k} R^2 \quad (5)$$

where SSR is the sum of squared residuals that within the state-space framework reads as

$$SSR = [T/s] \cdot \hat{\sigma}^2$$

with $\hat{\sigma}^2$ resulting from the biased maximum likelihood estimation of σ^2 of equation (1) obtained by the Kalman filter and SST is the sum of squared deviations of y_t from its mean. In equation (5) k is the number of covariates x_t of the regression equation (1) possibly including the constant and the linear trend.

However, Harvey (1989, p.268-9) severely criticizes the R^2 coefficient as a model goodness of fit measure within a time series context, given that y_t is often non-stationary. This would imply a R^2 coefficient close to unity. Therefore a recommended solution is to compute SST in terms of first differences of y_t in order to correct the R^2 statistics to cope with non-stationary data.

Furthermore, models goodness of fit statistics are given by the standard error of regression and maximum log likelihood value, respectively, defined as follows:

$$SER = \sqrt{SSR/([T/s] - k)}, \quad (6)$$

$$\hat{L}_c = -0.5 \cdot \{d_{T+1} + [T/s][\ln \hat{\sigma}^2 + \ln(2\pi) + 1]\} \quad (7)$$

where d_{T+1} is the sum of innovations variances taken in logarithms at time $t=s, 2s, \dots, [T/s]$.

The information criteria AIC and BIC employed to compare alternative model specifications can also be recovered as functions of $\hat{\sigma}^2$, respectively:

$$AIC = 2k/[T/s] + \ln \hat{\sigma}^2 \quad (8)$$

$$BIC = k \cdot \frac{\ln[T/s]}{[T/s]} + \ln \hat{\sigma}^2. \quad (9)$$

Durbin-Watson (1950, 1951) test statistic used to detect the presence of first-order autocorrelation in the residuals is defined with respect the standardized innovations \hat{v}_t as follows:

$$d = \frac{\sum_{t=k+2}^{[T/s]} (\hat{v}_t - \hat{v}_{t-1})^2}{\sum_{t=k+2}^{[T/s]} \hat{v}_t^2}.$$

The Jarque-Bera test statistic N for the normality of the residuals are derived from the formulas in Harvey (1989, p.560 eq.5.4.10 and 5.4.11) respectively for the third and the fourth moment of standardized residuals \hat{v}_t . In particular:

$$\sqrt{b_1} = \hat{\sigma}_*^{-3} \sum (\hat{v}_t - \bar{v})^3 / T^*$$

$$b_2 = \hat{\sigma}_*^{-4} \sum (\hat{v}_t - \bar{v})^4 / T^*$$

where the relation between $\hat{\sigma}^2$ and $\hat{\sigma}_*^2$ (unbiased residual variance) is

$$\hat{\sigma}_*^2 / \hat{\sigma}^2 = T / (T - k) \text{ and } T^* = T - k.$$

Hence, the N statistic of residual normality results as:

$$N = \frac{T^*}{6} \cdot b_1 + \frac{T^*}{24} \cdot (b_2 - 3)^2, \quad (10)$$

that under to the null-hypothesis follows a chi-squared distribution with 2 degrees of freedom for large samples.

An option for testing the statistical significance of the first P residual autocorrelations is given by the Ljung-Box Q-test statistic computed as:

$$Q = T^*(T^* + 2) \sum_{\tau=1}^P (T^* - \tau)^{-1} r_v^2(\tau) \quad (11)$$

where $r_v(\tau)$ are the sample autocovariances of standardized innovations. See Harvey (1989) p.259, eq.5.4.7 for detail. Within the context of ADL(1,1) models, we compare the related computed statistic to a chi-squared with degrees of freedom equal to $\sqrt{[T/s]} - 1$.

The last test is the $H(h)$ statistic for checking the heteroscedasticity of residuals given by:

$$H(h) = \sum_{t=T-h+1}^T \hat{v}_t^2 / \sum_{t=k+1}^{k+1+h} \hat{v}_t^2. \quad (12)$$

where h is an integer close to $T^*/3$. In this case the statistic $h \cdot H(h)$ is compared to a chi-squared with h degrees of freedom. See again Harvey (1989) p.259 eq.5.4.9, for a deeper treatment.

3.4 Evaluation of a temporal disaggregation

The diagnostic statistics described so far are important tools for evaluating the temporal disaggregation performance but they are not exhaustive to guarantee the quality of an exercise. Further criteria are required, especially when the analyst copes with incomplete dataset and time constraints.

A first remark of the above mentioned diagnostics is that they are based on residuals computed at low-frequency of observation. Therefore they are not able to provide insight about the quality of disaggregations in terms of high-frequency comovement with the related indicator. In other words - and this point will become clearer in next section - there are situations in which the exercise appears well specified according to the residuals-based statistics but yields disaggregated data either relatively too smooth or too volatile compared to the indicator. It is thus recommended to complement the usual analysis of residual diagnostics with both graphical inspection of disaggregations and a set of statistics of correlation between the indicator and the resulting disaggregation.

A second limit of innovation-based statistics within national accounts concerns their reduced power because the length of time series rarely exceeds 20 annual observations.

Therefore the approach ordinarily adopted within national accounts encompasses more than a single phase: at first, a preliminary analysis of general good fit of the indicator is undertaken, like comparing the pattern between the annualized indicators and target data adopting both graphical and synthetic statistics tools; then model estimation is carried out, followed by diagnostic checking through residual-based statistics complemented by statistics of revisions generated by the disaggregations and correlations between indicators and disaggregated data. In the standard practice these latter statistics are computed in terms of both quarterly and annual growth rates.

4 The empirical application

4.1 Design of the exercise

In this section we present the main results of an exercise of temporal disaggregation based on Italian data and providing evidence of a comparative analysis of the alternative classes of models presented in section 2. The exercise aims at reproducing the current practice of quarterly national accounts both for its

extension and for the nature of implied time series. It shows a selection of quarterly disaggregations based on annual national accounts and short term indicators by ISTAT: annual data are relative to the industrial value added split into 17 branches of economic activity (sections B-E of NACE Rev.2) according to the compilation detail of the Italian practice; quarterly indicators are industrial production indexes at same detail of activity. The sample period covers the interval 1995-2013. The table 1A of appendix A provides a summary description of the data employed in the exercise, while figures in appendix B present four graphs for both indicators and annual data by branch of activity. In these graphs nominal and volume data are presented separately, whereas the two figures 2B and 4B devoted to the indicators show together both raw and seasonal adjusted data.

The exercise has been carried out under a double perspective: the former looks at the performance of temporal disaggregations with respect to the type of data correction, i.e. taking into account the distinction between seasonal adjusted and unadjusted data; the latter at the type of evaluation, i.e. looking at both current price and volume data (chain linked values reference year 2010). In total 68 cases have been investigated. Each case reviews the full set of methods, from both the static regression approaches by Chow and Lin and Fernandez to the ADL class in the two specifications $ADL(1,0)$ and $ADL(1,1)$.³ With the exception of the Fernández approach, the estimations have concerned the unrestricted form of model (1) and the restricted variants without trend and with neither constant nor trend. ADL models have been estimated in both levels and first differences. Finally, using the approaches by both Proietti (2006) and Proietti and Moauro (2006), each form has been treated also in the logarithms. In total 2176 temporal disaggregations have been carried out.

Estimation of a so large variety of models and specifications implies the risk that some models could be not significant. An example is when the estimation leads to values for some regression coefficients either close to zero or to low values for the T statistic. Indeed, the aim of the exercise is indeed to mimic the current practice of quarterly national accounts when it is rather high the risk of not selecting at best the specification, due to lack of either data, or time, or for the presence of organizational constraints.

The exercise allows to appreciate the main advantages of the SSF-based estimate with respect to the regression approach, like the possibility to handle both a wider range of models, and the logarithmic transformation and the availability of a wider set of uniform diagnostic statistics based on standardized innovations.

4.2 Comparison of temporal disaggregation models

³ A similar exercise concerned also the Litterman model. However, these results, available upon request, were generally very problematic for the uncertain estimate of the autoregressive term ϕ occurring in log-likelihood maximization. For full details in this respect see Proietti (2005), pp.104-106.

In the comparative analysis of the performances between alternative model solutions we have considered two statistics: the former given by the mean absolute forecast error (MAE) of annual growth rates over the period 2006-2013 measured as the average difference in absolute terms between growth rates of the annual dependent variable and the sum of four extrapolated quarters over the annual totals of previous year; the latter statistic based on the correlation between quarterly growth rates of the indicator and the disaggregated series over the quarters 2005q1-2013q4. MAE statistics provide a measure of goodness of fit for extrapolated quarters, whereas correlations a synthetic measure on the quality of disaggregations over the last half sample with respect to the available indicator.

Tables 2A-3A of appendix A provide the values of MAEs relative to quarterly disaggregations of value added obtained using the seasonal adjusted version of the indicators. These statistics are presented by class of model, by branch of economic activity and type of evaluation, i.e. both at current prices and in volume. The first evidence is that the values of MAE move over a wide range of values, reflecting therefore both problematic cases –see for instance MAE of branch 7 equal to 16.3 for current price estimates- and virtuous situations – like MAE of branch 6 equal to 1.8.

Despite this set of tables are not fully informative of all the details of model specification and diagnosing checking, they allow to appreciate the performance of the ADL class of models. When the comparison concerns the (minimum) MAE statistic, ADL models outperform static disaggregations in 13 and 6 cases over a total of 17, respectively for nominal and volume data. Hence in 19 times over 34 occurrences (55,9%) dynamic models are relatively more performant than static forms, providing a tool able to reduce revisions of the extrapolations.

In appendix A, tables 6A-7A complete the analysis of results. Here are provided all the details of the best model specifications in terms of MAE relative to each branch: in particular, type of specification, possible log transformation and differentiation, the maximum log-likelihood value and all parameter estimates are presented.

From a joint analysis of tables 6A and 7A emerges that ADL models in difference are more suited to nominal time series (table 6A) than data in volumes (table 7A). This is not in contrast with the evidence that nominal data include the inflative component which usually features more evolutive trends or higher order of integration, both elements properly treated by models in differences. Concerning the logarithmic transformation, its effectiveness emerges in several cases for modelling both nominal and volume data.

In appendix A, tables 4A-5A show the correlations between quarterly growth rates of the indicator and the disaggregated series by branch and class of best model. The same logic of tables 2A-3A is followed. Concerning data at current prices, it emerges that the models where the quality of disaggregations is maximum in terms of fit to the indicator are the static ones. In particular the model by Fernández

guarantees maximum correlation in 12 branches over 17 and the model by Chow-Lin in the remaining 5 cases. Concerning data in volumes, the exercise also shows a clear prevalence of the Fernández approach over the others, notably in 12 of 17 cases. However also ADL(1,1) models are satisfactory, prevailing in the remaining 5 branches.

Tables 8A-9A provide model specifications of best model by branch according to the correlation criterium whose statistics are shown respectively in tables 4A-5A for nominal and volume data. From these tables we learn that the log-transformation is more performant in the majority of cases, notably 26 out of 34 occurrences (76,5%). Therefore, under this criterium log transformation appears superior than the option of no-treatment: , an additional property beside well-known advantages of the log-transformation applied to time series data. Furthermore, the logarithmic transformation ensure that disaggregated data assume only positive values differently from the treatment of data in levels.

4.3 Model selection

In this section we discuss a specific temporal disaggregation example in order to provide standard elements of model selection, identification and diagnosis among the enlarged class of models presented so far. In particular we focus on the quarterly disaggregation of the Italian annual value added relative to manufacture of machinery and equipment n.e.c (NACE Rev.2 A*38 code CK) at current prices, which represents 2.1% of total value added. The quarterly indicator is the industrial production inflated by output prices relative to the same branch of economic activity. The sample period is 1995:q1-2013:q4.

Figures 1 and 2 show, respectively, the seasonal adjusted and unadjusted disaggregated value added of machinery and equipment according to several model specifications: Chow and Lin (including the constant term and therefore denoted as CLc), Fernández (denoted as Fe), ADL(1,0) and ADL(1,1) both in levels and first differences (denoted in figure 2 with the suffix Δ to be distinguished by the corresponding models in levels).

Both figures 1 and 2 show how the alternative models produce similar disaggregated data: their patterns are so similar that it is very difficult to distinguish one disaggregation from the others by graphical inspection. Nevertheless, the pattern produced by the ADL(1,0) model in levels (red line) is relatively smoother than other disaggregations. In the case of figure 2, devoted to unadjusted data, the seasonal component of disaggregated data almost disappears, whereas in the case of seasonal adjusted data the pattern of the ADL(1,0) model tend to interpolate the alternative disaggregations.

Both tables 1 and 2 provide a synthetic comparative view of statistics for six alternative model specifications. Table 1 is for seasonal adjusted data and table 2 for unadjusted data.

A first evidence is that the models identified by the former set of data are very similar to the latter as it results from a comparative view of tables 1 and 2. In fact from both tables emerges that all the autoregressive parameters, the estimated regression coefficients, the statistics relative to MAE, correlations, information criteria AIC and BIC and R squared are almost identical.

Figure 1 - Quarterly disaggregated value added of machinery and equipment and inflated industrial production (seasonal adjusted data at current prices)

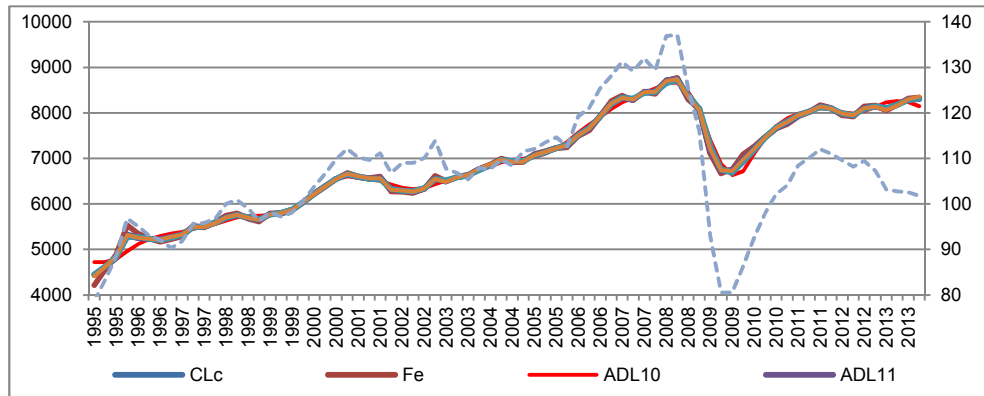
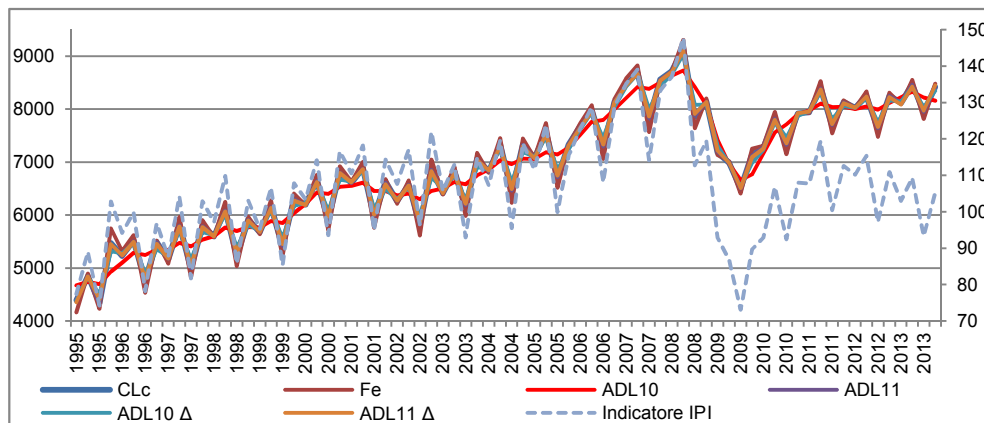


Figure 2 - Quarterly disaggregated value added of machinery and equipment and inflated industrial production (unadjusted data at current prices)



A second result which pops up from both tables 1 and 2 is the sub-optimality of the model ADL(1,0) in levels as emerged by graphical inspection. Indeed from one side the values of MAE of this model are the highest (6.47 and 6.39 for seasonal

adjusted and unadjusted data respectively) and from the other correlations between disaggregated data and the indicators in terms of quarterly growth rates are the lowest.

Table 1 - Estimated parameters and other statistics on quarterly disaggregations of value added for machinery and equipment (seasonal adjusted data at current prices)

	ρ	c	β_0	β_1	MAE	Correlations on quarterly growth rates	Correlations on annual growth rates	Log-lik	AIC	BIC	R^2 adj
CLc	0.996	11.49**	38.47**	--	3.15	0.96	0.93	-155.35	9.8	9.9	0.996
Fe	0	--	51.60**	--	3.30	0.93	0.94	-158.86	10.3	10.3	0.993
ADL(1,0)	0.874	--	8.48**	--	6.47	0.75	0.89	-167.14	11.5	11.5	0.975
ADL(1,1)	0.983	--	37.58**	-36.32**	2.95	0.96	0.93	-153.58	9.7	9.8	0.996
ADL(1,0) Δ	0.224	--	31.49**	--	2.84	0.80	0.68	-152.52	9.2	9.3	0.998
ADL(1,1) Δ	0.668	--	36.24**	-22.57**	2.51	0.80	0.67	-151.88	8.2	8.4	0.999

Table 2 - Estimated parameters and other statistics on quarterly disaggregations of value added for machinery and equipment (unadjusted data at current prices)

	ρ	c	β_0	β_1	MAE	Correlations on quarterly growth rates	Correlations on annual growth rates	Log-lik	AIC	BIC	R^2 adj
CLc	0.996	11.22**	37.66**	--	2.98	0.99	0.93	-155.0	9.8	9.9	0.996
Fe	0	--	51.48**	--	3.19	0.99	0.91	-159.2	10.3	10.4	0.992
ADL(1,0)	0.876	--	8.33**	--	6.39	0.62	0.88	-166.6	11.4	11.5	0.976
ADL(1,1)	0.982	--	36.62**	-35.44**	2.82	0.99	0.93	-153.11	9.7	9.8	0.996
ADL(1,0) Δ	0.235	--	30.43**	--	2.78	0.85	0.76	-152.02	9.2	9.3	0.998
ADL(1,1) Δ	0.685	--	35.12**	-22.44**	2.45	0.87	0.74	-151.01	8.1	8.2	0.999

Both the AIC and BIC information criteria in tables 1 and 2 are consistent with the comparative analysis provided so far since their values are higher for the ADL(1,0) than the other models. Same conclusion holds for the R^2 statistic whose values for the ADL(1,0) form are the lowest.

Concerning the other ADL models their performance are overall in line with the Chow and Lin form. In particular the ADL(1,1) in differences appears the best solution in terms of R^2 and MAE measuring, respectively, the fit of disaggregated values over all the sample period and of extrapolations. Concerning the overall fit, all the log-likelihood, the coefficient of determination R^2 , the AIC and BIC information criteria suggest this model as the best solution. Concerning the quality of extrapolations, the value of MAE of the model ADL(1,1) in differences (around 2.5 in both cases of seasonal adjusted and unadjusted data) is clearly lower than the other specifications (ranging between the values 2.8-6.5).

4.4 Quarterly disaggregation of raw data

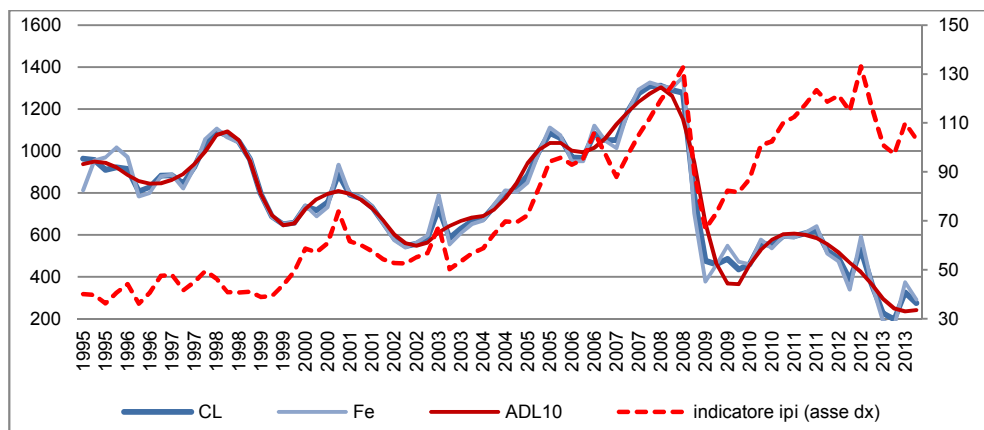
Main statistics on the overall fit of temporal disaggregation models as well as

parameter estimates are usually invariant with respect to the use of adjusted or unadjusted data. The results of the experiment of tables 1 and 2 are in line with this consideration since all the regression coefficients, the autoregressive parameters and the log-likelihood almost coincide in the two cases. Nevertheless, the ADL(1,0) model in levels is an example for which the fit over unadjusted data implies an imperfect transfer of the seasonal pattern from the indicator to the disaggregations.

The remarkable smoothness of disaggregated data produced by the ADL(1,0) model in levels emerges from the analysis of both figures 3 and 4 relative to the disaggregated value added of coke and refined petroleum products, respectively, at current prices and chain linked values. Here the pair of disaggregations of the ADL(1,0) form, in red lines, are accompanied by those by Chow and Lin (in blue) and Fernández (in gray), which both reproduce the seasonal pattern of the indicator (in dashed red lines).

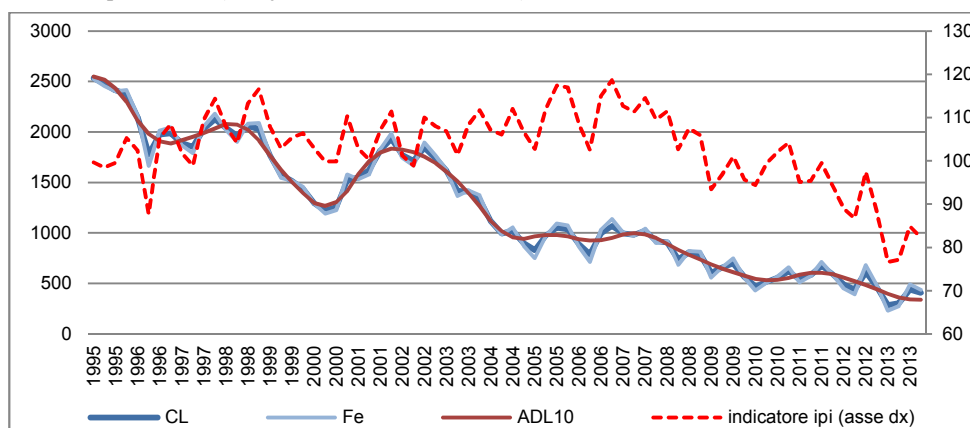
To complement the graphical inspection of figures 3-4, table 10A of Appendix A shows the correlations between the indicator and disaggregated data in terms of both quarterly and annual growth rates by main class of model⁴. In table 10A correlations of ADL(1,0) models are lower than other forms in the 82% of cases when the comparison concerns quarterly growth rates. In this comparison we have considered only the models with correlations higher than 0.2. Concerning correlations in terms of annual growth rates the distance between ADL(1,0) and other models is mitigated, as these statistics provide information more appropriate to measure the model goodness of fit.

Figure 3 – Quarterly disaggregated value added of coke and refined petroleum products and inflated industrial production (unadjusted data at current prices)



⁴ It is the average of correlations of the 3 options adopted in model specification, i.e. the unrestricted and the 2 restricted forms without trend and without both constant and trend. Note that quarterly growth rates of raw data, albeit without apparent economic sense, are particularly useful in this context since correlations are more discriminant for the selection of the best specification when the criterium is the fit to the pattern of the indicator.

Figure 4 – Quarterly disaggregated value added of coke and refined petroleum products and industrial production (unadjusted chain linked values)



4.5 Diagnosis of disaggregations: selection criteria and admissible results

There is a trade-off between enlargement to a wider class of temporal disaggregation methods and model selection. The solution suggested by the literature is adopting a general-to-specific strategy of model selection following given rules which allow to move from a general unrestricted form to restricted and more parsimonious models. Within dynamic regressions refer to Castle et al. (2011).

In the context of ADL regressions, the general-to-specific strategy moves in several respects: getting rid of the deterministic components from the unrestricted form including constant and trend; reducing the order of differentiation to apply to the data (from 1 to 0); restricting the lag order of the ADL model. A further dimension is the choice between modelling the data in levels or in logarithms.

The typical criteria adopted in model selection within general-to-specific strategies should be slightly adapted in the context of temporal disaggregation. In our application we have evaluated the following aspects: i) the statistical significance (at least 5%) of estimated model parameters; ii) positive value of the estimated autoregressive parameter to avoid volatility of estimated disaggregations (higher than -0.2 for ADL in differences); iii) positive and reasonable high correlations between the indicator and disaggregated data in both quarterly and annual growth rates; iv) general comovement between disaggregated data and the indicator by graphical inspection.

Percentage shares of admissible temporal disaggregation models are shown in table 3 where the shares are presented on the total of cases and according to type of

adjustment applied to the data, type of evaluation and transformation. Taking into account the 4 criteria listed above, admissible disaggregations are 49% out of the entire 2176 cases. The share slightly increases to 51% for current price data and decreases to 48% for chain-linked values.

From table 3 also emerges that the log-transformation provides a higher share of admissible results (51.8%) compared to estimations in levels (only 47%), meaning a larger stability of models in logarithms. A similar evidence (not shown in table 3) is found also looking separately at raw and seasonal adjusted data.

Looking at the class of models, the shares of table 3 are 100% for Fernández (but we do not considered all the variants), 71.6% for Chow-Lin, 61% for the ADL(1,0) form in levels, 52.2% for ADL(1,1). Lower shares have been found for the models ADL in differences for which we have obtained 32.8% and 11.5% of admissible cases respectively for ADL(1,0) and ADL(1,1) forms. Not surprisingly a higher complexity of model specification is accompanied by a lower chance of obtaining significative coefficients from estimation.

Table 3 – Percentage shares of admissible disaggregations in total and by class of model

	Seasonal adjustment		Type of evaluation		Type of transformation		Total shares
	Raw data	Seasonal adjusted data	Current prices	Chain linked values	Models in Levels	Models in logarithms	
Total	49.2	49.6	50.9	47.9	47.0	51.8	49.4
CL	71.1	72.1	70.1	73.0	64.2	78.9	71.6
FE	100	100	100	100	100	100	100
ADL(1,0)	58.8	63.2	58.8	63.2	61.3	60.8	61.0
ADL(1,1)	52.0	52.5	56.9	47.5	49.0	55.4	52.2
ADL(1,0) Δ	33.3	32.4	36.3	29.4	30.4	35.3	32.8
ADL(1,1) Δ	12.3	10.8	14.7	8.3	10.3	12.7	11.5

5. Conclusions

The paper has presented the most recent developments carried out at ISTAT within temporal disaggregation. The description of the new methodologies have been followed by a discussion of the results of a large scale experimental exercise based on ISTAT data with the aim of evaluating their performances.

Main contributions concern: the enlargement of disaggregation methods based on regressions from the static to the dynamic class of ADL models; the adoption of the

state space approach for model estimation, computation of disaggregated data and diagnosing checking; the introduction of the non-linear disaggregation for the treatment of the log-transformed data; the full integration of the dynamic setup within the standard procedures currently used in the quarterly national accounts process.

Concerning the empirical application, the results highlight a general good performance of dynamic models especially in terms of their predictive capacity. By contrast, traditional static disaggregation methods maintain their superiority with respect to the fit of disaggregated results to the indicator overall the sample. Then any hierarchy among models emerges from the exercise. Such not clear-cut outcomes in terms of models superiority according to the two criteria (MAE and quarterly growth rates correlations) reinforce the presumption that a sound evaluation requires a twofold analytic dimension. Indeed, checking both model robustness and the dynamic properties of the disaggregated series is a recommendable practice.

The non-linear treatment of data in logarithms has been found very effective: very often such specifications outperform linear ones. We conclude that temporal disaggregation of data in logarithms is an evolution of remarkable significance.

Among the critical features of dynamic models, we have detected a lack of transfer of the indicator seasonal pattern to the disaggregated data for the ADL(1,0) model.

In conclusion the enlargement to both the dynamic class of models and the non-linear treatment of logged data present undeniable advantages as emerges from the application. Nevertheless, the highlighted remarks, though limited to few cases, call for a deeper critical inspection of the results by the analyst.

References

- Astolfi, R. and M. Marini. 2005. Procedure di Disaggregazione Temporale utilizzate dall'ISTAT per la Stima dei Conti Economici Trimestrali in: Rapporto finale della Commissione di studio sul trattamento dei dati ai fini dell'analisi congiunturale, ISTAT, Ottobre 2005.
- Barbone L., G. Bodo and I. Visco. 1981. Costi e profitti in senso stretto: un'analisi su serie trimestrali, 1970-1980. *Bollettino della Banca d'Italia*, 36, numero unico.
- Bruno G. and G. Marra. 2005. New features for time series temporal disaggregation in the Modeleasy+ environment, in: Rapporto finale della Commissione di studio sul trattamento dei dati ai fini dell'analisi congiunturale, ISTAT, Ottobre 2005.
- Castle J.L., Doornik J. A., Hendry D. F. (2011) Evaluating automatic model selection, *Journal of Time Series econometrics*, 3(1), DOI: 10.2202/1941-1928.1097.

- Chow, G. C. and A. Lin. 1971. Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53, 372–5.
- De Jong P. 1991. The Diffuse Kalman Filter. *Annals of statistics*, 19: 1073-1083.
- 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. 2005. Il lavoro svolto e i risultati ottenuti, in: Rapporto finale della Commissione di studio sul trattamento dei dati ai fini dell'analisi congiunturale, ISTAT, Ottobre 2005.
- Durbin, J., and G. S. Watson. 1951a. Testing for Serial Correlation in Least Squares Regression. *Biometrika* 37: 409-428.
- Durbin, J., and G. S. Watson. 1951b. Testing for Serial Correlation in Least Squares Regression. *Biometrika* 38: 159-179.
- Fernández, R. B. 1981. A methodological note on the estimation of time series. *The Review of Economics and Statistics* 63, 471–6.
- Frale C, Marcellino M, Mazzi G., Proietti T. 2010. Survey data as coincident or leading indicators, *Journal of Forecasting*, 29: 109–131.
- Frale C, Marcellino M, Mazzi G.L., Proietti T. 2011. Euromind: a monthly indicator of the euro area economic conditions. *Journal of the Royal Statistical Society*, A174(2): 439–470.
- Grassi S., Proietti T., Frale C., Marcellino M. and Mazzi G. 2014. EuroMInd-C: a Disaggregate Monthly Indicator of Economic Activity for the Euro Area and member countries. *International Journal of Forecasting*, accepted for publication.
- Harvey, A. C., Pierce R. G. 1984. Estimating missing observations in economic time series. *Journal of the American statistical Association*, 79: 125-131.
- Harvey, A. C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey A. and C.H. Chung. 2000. Estimating the underlying change in unemployment in the UK. *Journal of the Royal Statistical Society, A*, 163: 303-328.
- Harvey A.C. and S.J. Koopman. 1997. Multivariate Structural Time Series Models (with comments), in C. Heij, J.M. Shumacher, B. Hanzon and C. Praagman (eds.), *System Dynamics in Economic and Financial Models*, Wiley, New York: 269-298.
- Hendry, D.F., and Mizon, G.E. 1978. Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal*, 88: 549-563.

- ISTAT. 2005. Rapporto finale della Commissione di studio sul trattamento dei dati ai fini dell'analisi congiunturale.
- ISTAT. 2015. I Conti Economici Trimestrali - Principali elementi informativi. Nota Informativa pubblicata il 2 settembre 2015. <http://www.istat.it/it/archivio/167411>
- Jarque, C. M. and A. K. Bera. 1987. A Test for Normality of Observations and Regression Residuals. *International Statistical Review*, 55.: 163–172.
- R.E. Kalman. 1960. A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME. *Journal of Basic Engineering*, 1960, 82, (Series D): 35-45.
- Koopman, S. J. 1997. Exact initial Kalman filtering and smoothing for non-stationary time series models. *Journal of the American Statistical Association*, 9: 1630–8.
- Litterman R.B. 1983. A random walk, Markov model for the distribution of time series. *Journal of Business and Economic Statistics*, 1: 169-173.
- Ljung, G. M. and G.E.P. Box. 1978. On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, 65,2: 297–303.
- Lupi C. and G. Parigi G. 1996. La disaggregazione temporale di serie economiche: un approccio econometrico. *ISTAT, Quaderni di Ricerca*, 3.
- Moauo F. 2014. Monthly Employment Indicators of the Euro Area and Larger Member States: Real-Time Analysis of Indirect Estimates. *Journal of forecasting*, 33, 5: 339-49.
- Moauo F. and G. Savio. 2005. Temporal disaggregation using multivariate structural time series models. *Econometrics Journal*, 8: 214-234.
- Proietti T. 2005. Temporal disaggregation by state space methods: dynamic regression methods revisited, in: “Rapporto finale della Commissione di studio sul trattamento dei dati ai fini dell'analisi congiunturale”, ISTAT, Ottobre 2005.
- Proietti T. 2006. On the Estimation of Nonlinearly Aggregated Mixed Models. *Journal of Computational and Graphical Statistics*, 15,1: 18-38.
- Proietti T. and F. Moauo. 2006. Dynamic factor analysis with non-linear temporal aggregation constraints. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55,2: 1467-9876.
- Santos Silva J.M.C. and F. N. Cardoso. 2001. The Chow-Lin method using dynamic models. *Economic Modelling*, 18 : 269-280.
- Sanz, R. 1981. Metodos de desagregacion temporal de series economicas. *Banco Espana. Servicio Estudios Economicos*, 22.

Appendix A – Tables

Table 1A - Time series used in the exercise listed by branch of economic activity, evaluation and sample period

Branch	ISIC rev.4 NACE rev.2	NACE Divisions	Evaluation	Period
3 Mining and quarrying	B	05 - 09	Current prices / Chain-linked values	1995-2013
4 Manufacture of food products, beverages and tobacco products	CA	10 - 12	Current prices / Chain-linked values	1995-2013
5 Manufacture of textiles, apparel, leather and related products	CB	13-15	Current prices / Chain-linked values	1995-2013
6 Manufacture of wood and paper products, and printing	CC	16-18	Current prices / Chain-linked values	1995-2013
7 Manufacture of coke, and refined petroleum products	CD	19	Current prices / Chain-linked values	1995-2013
8 Manufacture of chemicals and chemical products	CE	20	Current prices / Chain-linked values	1995-2013
9 Manufacture of pharmaceuticals, medicinal chemical and botanical products	CF	21	Current prices / Chain-linked values	1995-2013
10 Manufacture of rubber and plastics products, and other non-metallic mineral products	CG	22-23	Current prices / Chain-linked values	1995-2013
11 Manufacture of basic metals and fabricated metal products, except machinery and equipment	CH	24-25	Current prices / Chain-linked values	1995-2013
12 Manufacture of computer, electronic and optical products	CI	26	Current prices / Chain-linked values	1995-2013
13 Manufacture of electrical equipment	CJ	27	Current prices / Chain-linked values	1995-2013
14 Manufacture of machinery and equipment n.e.c.	CK	28	Current prices / Chain-linked values	1995-2013
15 Manufacture of motor vehicles, trailers and semi-trailers	CL	29	Current prices / Chain-linked values	1995-2013
16 Manufacture of other transport equipment		30	Current prices / Chain-linked values	1995-2013
17 Other manufacturing, and repair and installation of machinery and equipment	CM	31-33	Current prices / Chain-linked values	1995-2013
18 Electricity, gas, steam and air-conditioning supply	D	35	Current prices / Chain-linked values	1995-2013
19 Water supply, sewerage, waste management and remediation	E	36-39	Current prices / Chain-linked values	1995-2013

Table 2A – Mean absolute errors (MAE) of annualized extrapolations in growth rates (seasonal adjusted data at current prices) (a) (b) (c)

Model	branch 3	branch 4	branch 5	branch 6	branch 7	branch 8	branch 9	branch 10	branch 11	branch 12	branch 13	branch 14	branch 15	branch 16	branch 17	branch 18	branch 19
CL	6.64	2.44	2.98	1.70	14.61	3.72	2.48	2.47	1.89	4.23	2.98	2.15	4.12	2.19	2.17	4.63	11.65
FE	6.98	3.18	3.27	2.20	20.35	4.46	2.40	4.18	2.71	3.36	2.30	3.30	4.84	2.69	4.13	7.45	6.44
ADL(1,0)	6.53	1.81	3.02	1.47	15.59	4.02	2.35	2.45	3.25	3.39	2.29	1.49	2.26	2.89	2.24	3.70	2.99
ADL(1,1)	7.52	1.83	3.09	1.85	14.62	3.53	2.40	2.29	2.28	3.53	2.37	1.59	1.88	2.53	1.91	3.56	3.08

(a) In bold the minimum MAE by branch.

(b) For each branch and class of models the value refers to the model specification with minimum MAE.

(c) Each class of models estimated in both levels and logs. The Chow-Lin, ADL(1,0) and ADL(1,1) classes estimated in the unrestricted form of equation (1) and the restricted forms without trend and both constant and trend. The ADL classes include also the models in differences.

Table 3A – Mean absolute errors (MAE) of annualized extrapolations in growth rates (seasonal adjusted data of chain linked values) (a) (b) (c)

Model	branch 3	branch 4	branch 5	branch 6	branch 7	branch 8	branch 9	branch 10	branch 11	branch 12	branch 13	branch 14	branch 15	branch 16	branch 17	branch 18	branch 19
CL	7.58	2.19	3.90	1.85	7.24	5.70	3.20	1.60	2.66	4.18	3.63	2.22	2.90	2.55	3.41	3.31	4.49
FE	8.36	1.97	4.65	2.22	10.17	7.67	4.27	1.86	2.36	3.46	3.23	2.72	3.12	2.03	3.98	3.61	7.54
ADL(1,0)	8.67	1.99	4.42	2.00	7.96	5.57	3.09	2.39	2.82	4.15	3.48	1.71	5.39	2.36	3.18	3.54	3.57
ADL(1,1)	7.65	2.27	3.99	1.87	8.26	5.63	3.27	2.51	3.13	3.45	3.60	1.73	3.13	2.56	3.13	3.56	4.23

(a) In bold the minimum MAE by branch.

(b) For each branch and class of models the value refers to the model specification with minimum MAE.

(c) Each class of models estimated in both levels and logs. The Chow-Lin, ADL(1,0) and ADL(1,1) classes estimated in the unrestricted form of equation (1) and the restricted forms without trend and both constant and trend. The ADL classes include also the models in differences.

Table 4A – Correlations of quarterly growth rates between indicator and disaggregated data (seasonal adjusted data at current prices) (a) (b) (c)

Model	branch 3	branch 4	branch 5	branch 6	branch 7	branch 8	branch 9	branch 10	branch 11	branch 12	branch 13	branch 14	branch 15	branch 16	branch 17	branch 18	branch 19
CL	0.91	0.85	0.93	0.80	0.91	0.88	0.97	0.82	0.97	0.92	0.92	0.96	0.96	0.94	0.94	0.72	0.63
FE	0.91	0.85	0.94	0.83	0.92	0.88	0.97	0.84	0.95	0.92	0.91	0.93	0.96	0.94	0.94	0.73	0.63
ADL(1,0)	0.81	0.45	0.89	0.71	0.81	0.62	0.91	0.76	0.97	0.67	0.92	0.95	0.90	0.91	0.83	0.18	0.10
ADL(1,1)	0.88	0.26	0.92	0.79	0.91	0.87	0.90	0.75	0.97	0.84	0.92	0.96	0.96	0.92	0.94	0.16	0.31

(a) In bold the maximum correlation by branch.

(b) For each branch and class of models the value refers to the model specification with maximum correlation.

(c) Each class of models estimated in both levels and logs. The Chow-Lin, ADL(1,0) and ADL(1,1) classes estimated in the unrestricted form of equation (1) and the restricted forms without trend and both constant and trend. The ADL classes include also the models in differences.

Table 5A – Correlations of quarterly growth rates between indicator and disaggregated data (seasonal adjusted data of chain linked values) (a) (b) (c)

Model	branch 3	branch 4	branch 5	branch 6	branch 7	branch 8	branch 9	branch 10	branch 11	branch 12	branch 13	branch 14	branch 15	branch 16	branch 17	branch 18	branch 19
CL	0.87	0.90	0.94	0.87	0.90	0.89	0.96	0.91	0.97	0.91	0.95	0.97	0.99	0.94	0.94	0.89	0.64
FE	0.87	0.90	0.94	0.89	0.85	0.89	0.96	0.92	0.97	0.91	0.94	0.96	0.98	0.94	0.94	0.89	0.65
ADL(1,0)	0.78	0.49	0.90	0.80	0.71	0.85	0.89	0.89	0.94	0.67	0.95	0.97	0.99	0.90	0.56	0.58	0.22
ADL(1,1)	0.77	0.77	0.90	0.88	0.91	0.84	0.91	0.90	0.97	0.79	0.95	0.97	0.99	0.93	0.93	0.66	0.26

(a) In bold the maximum correlation by branch.

(b) For each branch and class of models the value refers to the model specification with maximum correlation.

(c) Each class of models estimated in both levels and logs. The Chow-Lin, ADL(1,0) and ADL(1,1) classes estimated in the unrestricted form of equation (1) and the restricted forms without trend and both constant and trend. The ADL classes include also the models in differences.

Table 6A – Model specifications and estimated parameters of estimations for which the minimum MAE by branch is obtained (seasonal adjusted data at current prices) (a)

Branch	Model	Specification	Log-lik	ϕ	c	g	β_0	β_1
3	ADL(1,0)	Δ	-142.73	0.31			8.90 (5.52)**	
4	ADL(1,0)	--	-150.69	0.98			1.70 (16.07)**	
5	Chow-Lin	--	-152.76	0.81			60.15 (71.77)**	
6	ADL(1,0)	Δ - log	-140.88	0.54			0.31 (5.59)**	
7	Chow-Lin	--	-150.09	0.98			9.37 (4.52)**	
8	ADL(1,1)	--	-142.01	0.96			16.00 (4.19)**	-14.85 (-3.88)**
9	ADL(1,0)	Δ	-131.93	0.51			6.24 (3.32)**	
10	ADL(1,1)	log	-153.43	0.99			0.38 (3.43)**	-0.37 (-3.26)**
11	Chow-Lin	--	-156.38	0.99			53.39 (15.35)**	
12	ADL(1,1)t	log	-132.67	0.93	0.05 (0.34)	0.00 (4.41)**	0.57 (3.56)**	-0.48 (-3.04)**
13	ADL(1,0)t	Δ -log	-132.52	0.03	0.01 (3.30)**	0.00 (-1.61)	0.39 (7.51)**	
14	ADL(1,0)c	Δ	-147.34	0.20	0.00 (4.04)**		0.44 (13.30)**	
15	ADL(1,1)t	--	-132.46	0.37	333.74 (3.26)**	3.59 (6.58)**	34.35 (9.85)**	-22.76 (-6.83)**
16	Chow-Lin	--	-136.01	0.75			16.30 (51.104)**	
17	ADL(1,1)	Δ	-139.15	0.83			24.93 (10.39)**	-19.45 (-7.57)**
18	ADL(1,1)c	--	-152.74	0.53	1113.98 (12.69)**		-35.90 (-2.21)**	54.86 (3.38)**
19	ADL(1,0)t	--	-130.59	0.81	353.58 (7.30)**	5.42 (9.96)**	-1.26 (-1.69)*	

(a) T-statistics in parenthesis: * p-value \leq 0.001; ** pvalue \leq 0.01; ***pvalue \leq 0.05.

Table 7A – Model specifications and estimated parameters of estimations for which the minimum MAE by branch is obtained (seasonal adjusted data of chain linked values) (a)

Branch	Model	Specification	Log-lik	φ	c	g	β_0	β_1
3	Chow-Lin c	--	-144.9	0.89	0.54 (3.99)*		0.51 (1.95)*	
4	Fernández	--	-156.41	0.00			76.91 (34.46)**	
5	Chow-Lin	--	-158.05	0.80			58.58 (60.05)**	
6	Chow-Lin c	log	-137.16	0.87	0.75 (16.76)**		0.55 (7.55)**	
7	Chow-Lin t	log	-144.12	0.70	0.39 (0.61)	-0.01 (-11.29)**	1.40 (3.05)**	
8	ADL(1,0)	--	-145.41	0.25			18.68 (82.29)**	
9	ADL(1,0)t	--	-131.18	0.63	-59.89 (-36)	3.23 (5.49)**	6.53 (3.39)**	
10	Chow-Lin	--	-152.81	0.999			31.67 (6.09)**	
11	Fernández	--	-158.95	0.00			69.94 (41.13)**	
12	ADL(1,1)	--	-142.49	0.98			6.84 (1.87)*	-6.60 (-1.81)*
13	Fernández	--	-144.17	0.00			14.83 (24.81)**	
14	ADL(1,0)c	Δ - log	-147.63	0.00	0.00 (3.09)**		0.64 (15.08)**	
15	Chow-Lin t	log	-136.92	0.91	0.34 (13.64)**	0.00 (4.43)**	0.85 (15.22)**	
16	Fernández	--	-144.02	0.00			21.70 (25.78)*	
17	ADL(1,1)	--	-153.12	0.94			37.55 (5.143)**	-34.42 (-4.72)**
18	Chow-Lin	--	-159.57	0.92			66.89 (28.26)**	
19	ADL(1,0)	log	-145.38	0.96			0.07 (132.33)**	

(a) T-statistics in parenthesis: * p-value \leq 0.001; ** pvalue \leq 0.01; ***pvalue \leq 0.05.

Table 8A – Model specifications and estimated parameters of estimations for which the maximum correlation by branch is obtained (seasonal adjusted data at current prices) (a)

Branch	Model	Specification	Log-lik	φ	c	g	β_0
3	Chow-Lin	log	-149.00	0.94			1.57 (62.78)**
4	Fernández	log	-163.86	0.00			2.00 (169.21)**
5	Fernández	log	-166.48	0.00			1.89 (170.37)**
6	Fernández	--	-151.28	0.00			39.49 (25.59)**
7	Fernández	log	-152.08	0.00			1.85 (34.35)**
8	Fernández	--	-145.73	0.00			26.72 (23.68)**
9	Fernández	log	-141.93	0.00			1.63 (148.65)**
10	Fernández	--	-160.44	0.00			48.21 (20.67)**
11	Chow-Lin t	log	-150.37	0.81	1.14 (29.63)**	0.00 (6.62)**	0.64 (14.49)**
12	Fernández	log	-147.42	0.00			1.44 (118.08)**
13	Chow-Lin t	log	-131.94	0.83	0.93 (17.78)**	0.00 (12.15)**	0.43 (6.97)**
14	Chow-Lin t	log	-144.04	0.68	1.92 (31.15)**	0.00 (21.21)**	0.56 (13.39)**
15	Fernández	log	-163.64	0.00			1.67 (75.57)**
16	Fernández	log	-143.2	0.00			1.64 (119.58)**
17	Chow-Lin c	--	-144.18	0.97	69.63 (5.22)**		28.23 (6.82)**
18	Fernández	log	-181.61	0.00			2.18 (61.05)**
19	Fernández	log	-160.8	0.00			1.76 (68.16)**

(a) T-statistics in parenthesis: * p-value \leq 0.001; ** pvalue \leq 0.01; ***pvalue \leq 0.05.

Table 9A – Model specifications and estimated parameters of estimations for which the maximum correlation by branch is obtained (seasonal adjusted data of chain linked values) (a)

Branch	Model	Specification	Log-lik	ϕ	c	g	β_0	β_1
3	Fernández	log	-150.59	0.00			1.54 (83.63)**	
4	Fernández	log	-159.15	0.00			1.98 (260.92)**	
5	Fernández	log	-166.86	0.00			1.83 (184.2)**	
6	Fernández	--	-145.66	0.00			38.46 (38.71)**	
7	ADL(1,1)c	log	-152.33	0.99	0.12 (0.15)		2.35 (2.68)**	-2.36 (-2.49)**
8	Fernández	log	-152.32	0.00			1.68 (134.43)**	
9	Fernández	log	-143.71	0.00			1.63 (133.25)**	
10	Fernández	--	-152.31	0.00			42.89 (35.13)**	
11	ADL(1,1)t	log	-150.45	0.58	2.29 (19.97)**	0.00 (12.73)**	0.81 (8.61)**	-0.50 (-5.14)**
12	Fernández	log	-149.26	0.00			1.50 (129.31)**	
13	ADL(1,1)	Δ - log	-140.50	0.57			0.73 (7.86)**	-0.49 (-4.73)**
14	ADL(1,1)c	Δ - log	-148.08	0.40	0.00 (2.84)**		0.66 (10.50)**	-0.28 (-3.77)**
15	ADL(1,1)t	--	-136.01	0.89	124.41 (2.06)**	0.86 (2.69)**	23.48 (14.61)**	-21.88 (-13.38)**
16	Fernández	log	-146.54	0.00			1.65 (138.97)**	
17	Fernández	--	-156.22	0.00			54.52 (29.45)**	
18	Fernández	log	-168.65	0.00			1.99 (153.7)**	
19	Fernández	log	-163.90	0.00			1.68 (85.60)**	

(a) T-statistics in parenthesis: * p-value \leq 0.001; ** pvalue \leq 0.01; ***pvalue \leq 0.05.

Table 10A – Correlations between disaggregated series and quarterly indicator in terms of quarterly (Δq) and annual (Δy) growth rates (a)

		Current prices data						Chain-linked data					
		CL	Fe	ADL10	ADL11	ADL10 Δ	ADL11 Δ	CL	Fe	ADL10	ADL11	ADL10 Δ	ADL11 Δ
br 3	Δq	0.89	0.91	0.39	0.88	0.51	0.58	0.76	0.91	0.23	0.37	0.76	0.83
	Δy	0.67	0.68	0.58	0.67	0.63	0.38	0.40	0.47	0.27	0.36	0.41	0.37
br 4	Δq	0.94	0.98	0.40	0.21	0.53	-0.82	0.77	0.99	0.25	-0.29	-0.74	-0.82
	Δy	0.45	0.58	0.26	0.22	0.36	-0.21	0.37	0.58	0.22	0.12	0.02	-0.43
br 5	Δq	0.99	0.99	0.80	0.99	0.97	-0.25	0.99	0.99	0.80	0.98	0.98	0.98
	Δy	0.81	0.82	0.70	0.79	0.78	0.41	0.75	0.75	0.66	0.73	0.76	0.77
br 6	Δq	0.98	0.98	0.64	0.97	0.85	-0.28	0.99	0.99	0.69	0.99	0.94	-0.68
	Δy	0.74	0.75	0.71	0.74	0.72	0.70	0.81	0.82	0.78	0.81	0.80	0.61
br 7	Δq	0.76	0.70	0.41	0.76	0.77	0.76	0.48	0.82	-0.01	0.57	0.69	0.67
	Δy	0.72	0.73	0.63	0.73	0.72	0.72	0.36	0.64	0.25	0.36	0.55	0.55
br 8	Δq	0.59	0.97	0.36	0.96	0.57	0.92	0.67	0.96	0.86	0.76	0.96	-0.14
	Δy	0.54	0.72	0.49	0.70	0.56	0.67	0.57	0.70	0.69	0.59	0.72	0.65
br 9	Δq	0.89	0.99	0.70	0.72	0.97	0.97	0.99	0.99	0.64	0.61	0.96	0.95
	Δy	0.68	0.77	0.55	0.46	0.69	0.70	0.55	0.56	0.29	0.31	0.48	0.53
br 10	Δq	0.48	0.98	0.86	0.40	0.84	0.84	0.83	0.99	0.61	0.92	0.97	0.90
	Δy	0.57	0.70	0.68	0.55	0.65	0.65	0.76	0.86	0.65	0.73	0.85	0.84
br 11	Δq	0.99	0.99	0.88	0.99	0.99	0.98	0.99	0.99	0.71	0.99	0.93	0.94
	Δy	0.95	0.94	0.92	0.95	0.94	0.94	0.94	0.94	0.88	0.94	0.92	0.91
br 12	Δq	0.93	0.95	0.20	0.62	0.80	0.31	0.94	0.96	0.21	0.95	0.85	-0.69
	Δy	0.67	0.70	0.56	0.44	0.62	0.55	0.56	0.61	0.46	0.56	0.53	0.23
br 13	Δq	0.97	0.97	0.46	0.97	0.96	0.82	0.98	0.98	0.45	0.98	0.98	0.89
	Δy	0.84	0.84	0.73	0.84	0.84	0.83	0.89	0.89	0.79	0.89	0.89	0.89
br 14	Δq	0.99	0.99	0.75	0.99	0.98	0.91	0.99	0.99	0.29	0.99	0.99	0.99
	Δy	0.93	0.91	0.90	0.93	0.94	0.93	0.95	0.95	0.65	0.95	0.95	0.95
br 15	Δq	0.76	0.99	0.64	0.97	0.03	0.93	0.99	0.99	0.78	0.99	0.98	0.99
	Δy	0.81	0.90	0.64	0.87	0.04	0.85	0.96	0.96	0.88	0.96	0.96	0.96
br 16	Δq	0.99	0.99	0.89	0.99	0.98	0.08	0.99	0.99	0.80	0.99	0.96	-0.03
	Δy	0.79	0.79	0.77	0.79	0.79	0.65	0.78	0.79	0.72	0.78	0.77	0.63
br 17	Δq	0.99	0.99	0.76	0.99	0.94	0.99	0.99	0.99	0.62	0.99	0.89	0.89
	Δy	0.79	0.78	0.68	0.79	0.73	0.77	0.76	0.77	0.65	0.76	0.58	0.55
br 18	Δq	0.90	0.96	0.45	-0.34	0.59	-0.72	0.98	0.99	0.58	0.91	0.97	-0.12
	Δy	0.15	0.24	0.09	0.01	0.08	-0.08	0.56	0.64	0.33	0.45	0.49	0.32
br 19	Δq	0.15	0.96	0.03	-0.13	-0.27	0.83	0.91	0.97	0.41	0.60	0.77	-0.65
	Δy	-0.06	0.17	-0.06	-0.07	-0.06	0.02	0.25	0.29	0.17	0.18	0.19	0.12

(a) Average correlations over the alternative specifications (standard, with constant, with constant and trend) by each class of model in levels are presented.

Appendix B – Graphs of annual and quarterly data used in the application

Figure 1B – Value added by branch: annual chain-linked values reference year 2010 over the years 1995-2003

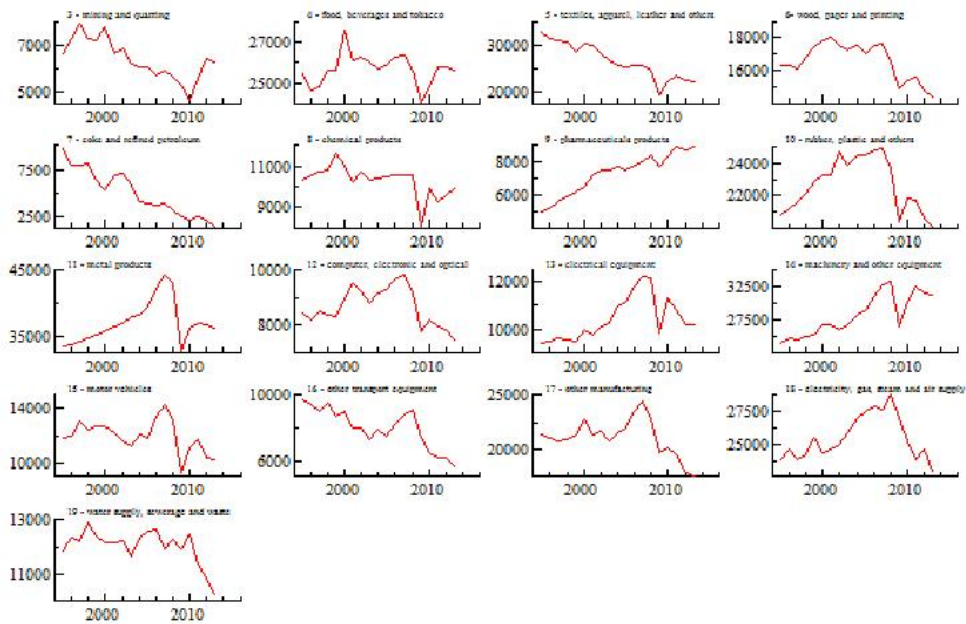


Figure 2B - Industrial production by branch: seasonal adjusted (in red) and raw data (in blue) over the quarters 1995q1-2003q4

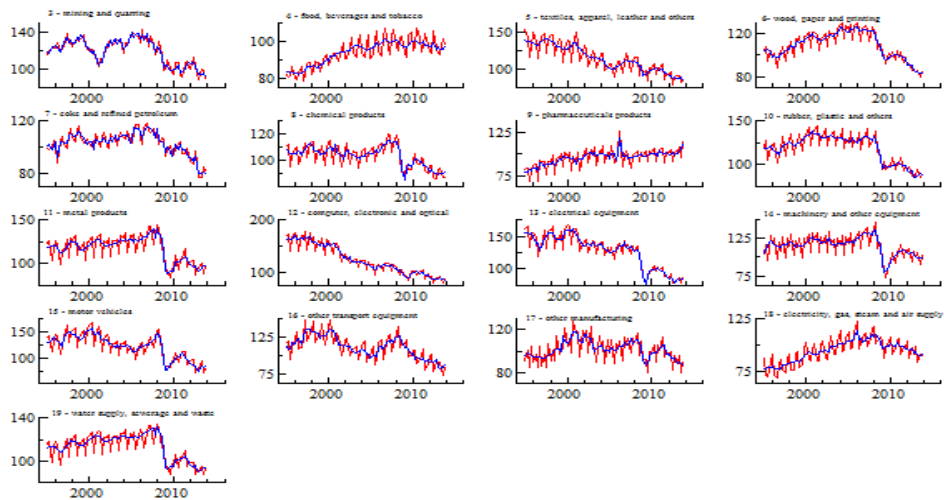


Figure 3B – Value added by branch: annual values at current prices over the years 1995-2003

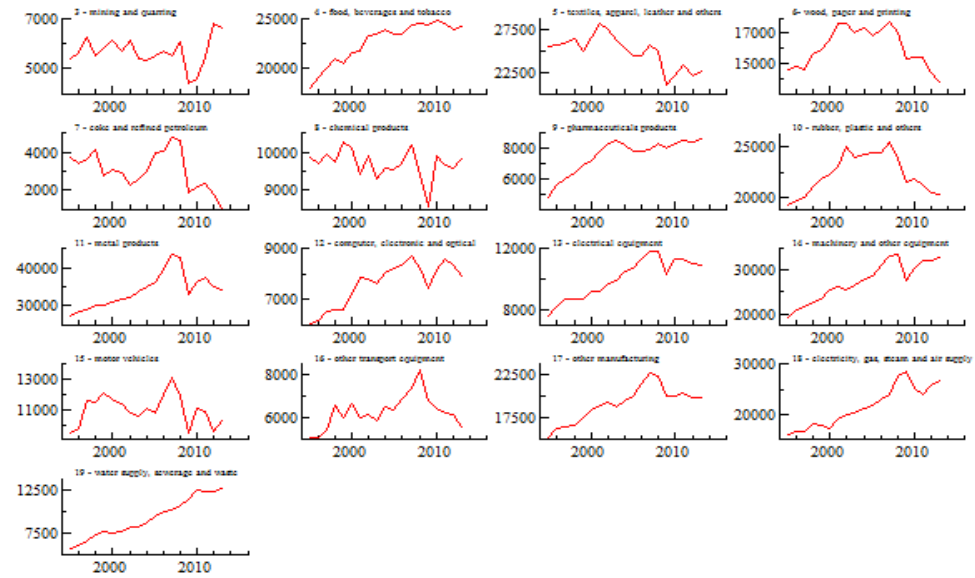


Figure 4B – Inflated industrial production by branch: seasonal adjusted (in red) and raw data (in blue) over the quarters 1995q1-2003q4

