Temporal Disaggregation by State Space Methods: Dynamic Regression Methods Revisited

Tommaso Proietti
Dipartimento di Scienze Statistiche
Università di Udine

Abstract

The paper documents and illustrates state space methods that implement time series disaggregation by regression methods, with dynamics that depend on a single autoregressive parameter. The most popular techniques for the distribution of economic flow variables, such as Chow-Lin, Fernández and Litterman, are encompassed by this unifying framework.

The state space methodology offers the generality that is required to address a variety of inferential issues, such as the role of initial conditions, which are relevant for the properties of the maximum likelihood estimates and for the the derivation of encompassing representations that nest exactly the traditional disaggregation models, and the definition of a suitable set of real time diagnostics on the quality of the disaggregation and revision histories that support model selection.

The exact treatment of temporal disaggregation by dynamic regression models, when the latter are formulated in the logarithms, rather than the levels, of an economic variable, is also provided.

The properties of the profile and marginal likelihood are investigated and the problems with estimating the Litterman model are illustrated. In the light of the nonstationary nature of the economic time series usually entertained in practice, the suggested strategy is to fit an autoregressive distribute lag model, which, under a reparameterisation and suitable initial conditions, nests both the Chow-Lin and the Fernández model, thereby incorporating our uncertainty about the presence of cointegration between the aggregated series and the indicators.

Keywords: Autoregressive Distributed Lag Models, COMFAC, Augmented Kalman filter and smoother, Marginal Likelihood, Logarithmic Transformation.

1 Introduction

Temporal disaggregation methods play an important role for the estimation of short term economic indicators. This is certainly the case for a number of European countries, including France, Italy, Spain, Belgium and Portugal, whose national statistical institutes make extensive use of those methods, among which the Chow-Lin (Chow and Lin, 1973) procedure stands out prominently, for constructing the quarterly national economic accounts from annual figures, using a set of indicators available at the quarterly frequency. As a consequence, a large share of the Euro area quarterly gross domestic product is actually estimated by disaggregation techniques.

Interpolation and distribution are the two facets of the disaggregation problem. The former deals with the estimation of the missing values of a stock variable that at points in time that have been systematically skipped by the observation process; the latter arises when measurements over flow variables is in the form of a linear aggregate, typically the total or the average over s consecutive periods. We shall consider as the leading case of interest the situation when the aggregate series, concerning the annual totals of an economic variable, have to be distributed across the quarters, using related series that are available for the shorter subperiod.

This paper concentrates on a set of dynamic regression methods that depend on a single autoregressive parameter and a regression kernel aiming at capturing the role of related indicators, and that encompass the most popular techniques such as Chow-Lin (1981), Fernández (1981) and Litterman (1983), which are are based on a regression model with autocorrelated errors generated respectively by a first order autoregressive (AR) process, a random walk and an ARIMA(1,1,0) process. In addition, we consider autoregressive distributed lag models (ADL) that nest the traditional models. See Di Fonzo (2003) for a review and additional references. Discussion will be limited to the disaggregation problem; interpolation poses slightly different issues and will not be considered further.

The class of models investigated is very restricted and its choice requires some motivation. First and foremost, the investigation focuses on popular methods that have widespread application. During the process of implementing them, for comparison with more sophisticated methods, we felt that certain aspects were relatively unexplored, despite the large literature on the topic. For instance, from the empirical standpoint, what seemed to be a plausible, stripped to the bone, representation, in the lack of cointegration (Engle and Granger, 1987) between the series and the indicator variable, i.e. the Litterman model, could not be estimated reliably, which called for further analysis. The role of initial conditions and deterministic components was another unsettled issue.

Secondly, we adhere to the idea that the disaggregated model should be kept relatively simple: as it will become clear in the course of the discussion, with particular reference to the Litterman model, parsimonious modelling is particularly compelling here, since the aggregate data may not be very informative on the parameters of the disaggregate model.

Finally, there are excellent papers covering other methodologies, among which we mention Harvey and Chung (2000) and Moauro and Savio (2002). These references implement a genuinely multivariate time series approach to the disaggregation problem, which overcomes some of the limitations of the regression based methods, namely the assumption of exogeneity of the indicators, see also Harvey (1989, sec 8.7.1), and the assumption that the indicators are measured free of measurement error.

The unifying framework of the paper is the state space representation and the associated methods. The statistical treatment follows Harvey (1989, ch. 6, sec. 3): starting from the the state space representation for the disaggregated model, the derived state space model handling the aggregated observations is derived, augmenting the state vector for a cumulated variable that is only partially observed. This converts the disaggregation problem onto a missing values problem, that can be addressed by skipping certain updating operations in the filtering and smoothing equations.

Particular attention is devoted to the exact initialisation of the models, and the treatment of regression effects and initial conditions as fixed or diffuse. Diffuseness embodies uncertainty about initial conditions or parameters and model nonstationarity. The issue is of fundamental importance for the properties of the maximum likelihood estimates and for nesting exactly the regression methods within more general ADL models. As far as the latter are concerned, the illustrative examples presented in the paper provide further support for the modelling strategy set outlined in Hendry and Mizon (1978), which leads to entertain the Chow-Lin or the Fernández model provided certain common factor restriction proves valid.

The augmented Kalman filter and smoother proposed by de Jong (1991) is the key algorithm for evaluating the likelihood, and for defining the set of time series innovations, that can be used for diagnostic checking and addressing the issue of revision of the real time estimates when the aggregate information accrues. The relevance of diagnostic checking is usually neglected in the literature, one reason being that the innovations are not automatically available from the implementation of these methods in a classical regression framework. The role of revision histories for model selection is illustrated using a real life example.

The paper also provides an exact treatment of temporal disaggregation by dynamic regression models when the disaggregated model is formulated in the logarithms, rather than the levels, of an economic variable. This is usually the case for flows measured on a ratio scale, such as production and income; the logarithm provides the natural transformation under which the usual assumptions concerning the linear model (linearity, homoscedasticity and normality of errors) are plausible.

The paper is structured as follows: section 2 introduces the main disaggregated models and their state space representation. Section 3 discusses how the latter is modified as a consequence of temporal aggregation. The statistical treatment for linear disaggregation methods is the topic of section 4, dealing with evaluation of the likelihood, marginalisation of regression effects, diagnostic checking, filtering and smoothing. Section 5 addresses the problem of distribution of flows that are modelled on a logarithmic scale.

The reader that is more interested in the applications may skip some of the more technical parts in sections 3-5, which nevertheless form integral part of the paper, and move to section 6, which contains five illustrations concerning the role of different assumptions on initial conditions, the virtues of nesting the traditional procedures in more general autoregressive distributed lag models, the problems with the estimation of the Litterman model, and the use of nonlinear disaggregation techniques for ensuring that the estimated series can take only admissible values.

Finally, a separate appendix documents a set of functions, written in Ox (see Doornik, 2001), implementing the procedures discussed in this paper.

2 Disaggregated Time Series Models

The disaggregated time series models considered in this paper admit the following state space representation:

$$\begin{array}{rcll} y_t &=& z'\alpha_t + x_t'\beta, & t = 1, \ldots, n, \\ \alpha_t &=& T\alpha_{t-1} + W_t\beta + H\epsilon_t, & t = 2, \ldots, n \\ \alpha_1 &=& a_1 + W_1\beta + H_1\epsilon_1, & \epsilon_t \sim \text{NID}(0, \sigma^2), \quad \beta \sim \text{N}(b, \sigma^2 V) \end{array} \tag{1}$$

It should be noticed that the measurement equation does not feature a measurement error, and that the system matrices are time invariant. Both restrictions can be easily relaxed.

The vectors x_t and the matrices W_t contain exogenous regressors that enter respectively the measurement equation and the transition equation and zero elements corresponding to effects that are absent from one or the other equations. They are usually termed ”indicators” or ”related variables” in the literature.

The initial state vector, α_1, is expressed as a function of fixed and known effects (a_1), random stationary effects ($H_1\epsilon_1$, where the notation stresses that H_1 may differ from H), and regression effects, $W_1\beta$.

Two assumptions can be made concerning β: (i) β is considered as a fixed, but unknown vector ($V \to 0$); this is suitable if it is deemed that the transition process governing the states has started at time $t = 1$; (ii) β is a diffuse random vector, i.e. it has an improper distribution with a mean of zero ($b = 0$) and an arbitrarily large variance matrix ($V^{-1} \to 0$). This is suitable if the process has started in the indefinite past.

The first case has been considered by Rosenberg (1973), who showed that β can be concentrated out of the likelihood function, whereas (ii) is considered in de Jong (1991). Diffuseness expresses parameter uncertainty or the nonstationarity of a particular state component and entails marginalising the likelihood with respect to the parameter vector β.

The representation is sufficiently rich to accommodate the traditional linear disaggregation techniques proposed in the literature. The rest of this section introduces the main models currently in use, discussing its state space representation and the initialisation issue, which turns out to be crucial for the comparison of the different specifications.

2.1 The Chow-Lin model

The Chow-Lin (Chow and Lin, 1971, CL henceforth) disaggregation method is based on the assumption that y_t can be represented by a linear regression model with first order autoregressive errors:

$$y_t = \alpha_t + x_t'\beta, \quad \alpha_t = \phi\alpha_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{NID}(0, \sigma^2), \tag{2}$$

with $|\phi| < 1$ and $\alpha_1 \sim \mathrm{N}(0, \sigma^2/(1-\phi^2))$.

The model is thus a particular case of (1), with scalar α_t, system matrices $z = 1,\ T = \phi,\ H = 1$. As far the initial conditions are concerned, as α_t is a stationary zero mean AR(1) process, assuming that the process applies since time immemorial, $\alpha_1 \sim N(0, \sigma^2/(1-\phi^2)$, which amounts to setting: $a_1 = 0,\ W_1 = 0$, and $H_1 = (1-\phi^2)^{-1/2}$.

If some element of x_t are nonstationary, the CL model postulates full cointegration between them and the series y_t.

Deterministic components (e.g. a linear trend) are handled by including appropriate regressors in the set x_t, e.g. by setting $x_t = [1, t, x_{3t}, \dots, x_{kt}]'$, and writing $y_t = \mu + \gamma t + \sum_j \beta_j x_{jt} + \alpha_t$, with the first two elements of β being denoted μ and γ.

Alternatively, they can be accommodated in the transition equation, which becomes $\alpha_t = \phi\alpha_{t-1} + m + gt + \epsilon_t$. The state space form corresponding to this case features $W_t = [1, t, 0']$, for $t > 1$, whereas $W_1 = [(1-\phi)^{-1}, (1-2\phi)/(1-\phi)^2,\ 0']$. The first two elements of the vector of exogenous regressors x_t are zero, since m and g do not enter the measurement equation.

In fact, if it is assumed that the new transition model has applied since time immemorial,

$$\alpha_1 = \frac{m}{1-\phi} + g\left(1 - \phi\sum_{j=1}^{\infty} j\phi^j\right) + \frac{\epsilon_1}{1-\phi L} = \frac{m}{1-\phi} + \frac{1-2\phi}{(1-\phi)^2}g + \frac{\epsilon_1}{1-\phi L},$$

recalling $\sum_{j=0}^{\infty} j\phi^j = \phi/(1-\phi)^2$; L is the lag operator, such that $L^j y_t = y_{t-j}$.

Under fixed regression coefficients, the two alternative representations are exactly equivalent, with

$$\mu = \frac{1}{1-\phi}m - \frac{\phi}{(1-\phi)^2}g, \quad \gamma = \frac{1}{1-\phi}g. \tag{3}$$

A difference arise with respect to the definition of the marginal likelihood when these coefficients are diffuse, as we shall see in section 4.1.

First of all, it ought to be noticed that first order differences $\Delta y_t = y_t - y_{t-1}$ eliminate the constant term, whereas second order differences are required, $\Delta^2 y_t$, so as to eliminate dependence on the coefficient of the linear trend. This is true of both parameterisations, with one notable exception that arises for the second one, when $\phi = 1$, in which case $\Delta^2 y_t = g + \Delta\epsilon_t$.

Denoting $\sigma^2_\gamma = \mathrm{Var}(\gamma)$ and $\sigma^2_m = \mathrm{Var}(g)$, from (3) we find that, for instance, $\sigma^2_\gamma = \sigma^2_g/(1-\phi)^2$, so that a diffuse γ, $\sigma^2_\gamma \to \infty$, arises both for $\sigma^2_m \to \infty$ and $\phi \to 1$.

2.2 The Litterman and Fernandez models

According to the Litterman (1983) model, the disaggregated process is a regression model with ARIMA(1,1,0) disturbances:

$$y_t = x_t'\beta + u_t, \quad \Delta u_t = \phi\Delta u_{t-1} + \epsilon_t. \tag{4}$$

Litterman explicitly assumes that the u_t process has started off at time $t = 0$ with $u_0 = \Delta u_0 = 0$ (Litterman, 1983, last paragraph of page 170). This is usually inadequate, unless the set of indicators includes a constant (which would capture the effect of the initial value); the inclusion of a linear trend amounts to allowing for non zero drift in the ARIMA(1,1,0) process.

The Fernández (1981) model arises in the particular case when $\phi = 0$ and thus u_t is a random walk.

The state space representation of (4) is obtained by defining the state vector and system matrices as follows:

$$\alpha_t = \begin{bmatrix} u_{t-1} \\ \Delta u_t \end{bmatrix}, \quad z' = [1, 1], \quad T = \begin{bmatrix} 1 & 1 \\ 0 & \phi \end{bmatrix}, \quad H = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The Litterman initialisation implies $u_1 = u_0 + \phi \Delta u_0 + \epsilon_1 = \epsilon_1$, which is implemented casting:

$$a_1 = 0, \; W_1 = 0, \; H_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Alternatively, including u_{-1} in the vector β as its first element, in which case x_t features a zero element in first position, and assuming that the stationary process has started in the indefinite past, the initial conditions are:

$$W_1 = \begin{bmatrix} 1 & 0' \\ 0 & 0' \end{bmatrix}, \; H_1 = \frac{1}{\sqrt{1-\phi^2}} \begin{bmatrix} 1 \\ \phi \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

This follows from writing

$$\begin{bmatrix} u_0 \\ \Delta u_1 \end{bmatrix} = \begin{bmatrix} u_{-1} \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ \phi \end{bmatrix} \Delta u_0 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \epsilon_1,$$

and taking $\Delta u_0 \sim \mathrm{N}(0, \sigma^2/(1-\phi^2)), \epsilon_1 \sim \mathrm{N}(0, \sigma^2)$. The diffuse nature of u_{-1} arises from the nonstationarity of the model.

It should be noticed that in this second setup we cannot include a constant in x_t, since this effect is captured by u_{-1}.

Finally, the ARIMA(1,1,0) process can be extended to include a constant and a trend in $\Delta u_t = \phi \Delta u_{t-1} + m + gt + \epsilon_t$; the parameters m and g are incorporated in the vector β and the matrices W_1 and W_t are easily extended; for instance, if $\beta = [u_{-1}, m, g, \beta_2']'$ where β_2 corresponds to the regression effects affecting only the measurement equation,

$$W_1 = \begin{bmatrix} 1 & 0 & 0 & 0' \\ 0 & \frac{1}{1-\phi} & \frac{1-2\phi}{(1-\phi)^2} & 0' \end{bmatrix}, W_t = \begin{bmatrix} 0 & 0 & 0 & 0' \\ 0 & 1 & t & 0' \end{bmatrix}, x_t = [0, 0, 0, x_{2t}']'.$$

The alternative is to include a trend in the measurement equation; however, the remarks concerning the inclusion of a constant when the starting value is already incorporated in the vector β continue to hold.

2.3 Autoregressive Distributed Lags Models

Following Hendry and Mizon (1978), it is well known that both the CL and Litterman models can be nested within a more general dynamic regression model.

Consider the Autoregressive Distributed Lag model known as ADL(1,1), which takes the form:

$$y_t = \phi y_{t-1} + m + gt + x_t'\beta_0 + x_{t-1}'\beta_1 + \epsilon_t, \qquad \epsilon_t \sim \text{NID}(0, \sigma^2). \tag{5}$$

Under suitable assumptions about initial conditions, the ALD(1,1) model nests the CL regression model with AR(1) errors: in particular if

$$\beta_1 = -\phi\beta_0$$

the AR polynomial and the distributed lag polynomial $(\beta_0 + \beta_1 L)$ share a ”common factor”, and the model (5) can be rewritten as $y_t = x_t\beta_0 + \alpha_t$ where α_t is the AR(1) process given in (2). The ADL(1,0) arises instead if $\beta_1 = 0$.

The benefits of this representation versus a simple regression with AR(1) errors are thoroughly discussed in Hendry and Mizon (1978), especially as a more effective modelling framework for avoiding the insurgence of spurious regressions between economic time series. See also Banerjee *et al.* (1993, chapter 2). Section 6.2 illustrates some other interesting features relating more specifically to the disaggregation problem.

The state representation is

$$\begin{aligned} y_t &= \alpha_t \\ \alpha_t &= \phi\alpha_{t-1} + W_t\beta + \epsilon_t \end{aligned} \tag{6}$$

with system matrices $z' = 1, T = \phi, H = 1, W_t = [1, t, x_t', x_{t-1}']$; notice that differently from the CL model the regression effects are all included in the transition equation. The β vector has elements $\beta = [m, g, \beta_0, \beta_1]$.

As for initial conditions, assuming that the process started in the indefinite past, exact nesting of the CL model occurs if one posits:

$$y_1 = \alpha_1 = \frac{1}{1-\phi}m + \frac{1-2\phi}{(1-\phi)^2}g + \frac{1}{1-\phi}x_t'(\beta_0 + \beta_1) + \frac{1}{1-\phi L}\epsilon_1, \tag{7}$$

which corresponds to setting

$$a_1 = 0, \;\; W_1 = \frac{1}{1-\phi}\left[1, \frac{1-2\phi}{1-\phi}, x_1', x_1'\right], H_1 = \frac{1}{\sqrt{1-\phi^2}}.$$

Other initialisations are possible:

- y_1 can be considered as a fixed value, which would amount to assuming that the state space model (6) holds from $t = 2$ onwards, with $\alpha_2 \sim \text{N}(\phi y_1 + m + 2g + x_2'\beta_0 + x_1'\beta_1, \sigma^2)$, (in terms of the state space representation $a_2 = \phi y_1, W_2 = [1, 2, x_2', x_1'], H_2 = 1$. This solution has no practical relevance in our case since, due to temporal aggregation, y_1 is not available.

- $y_1 \sim N(c + x_1'\beta, \sigma^2)$: y_1 is random and the process is supposed to have started at time $t = 0$ with a value which is fixed, but unknown; c is an additional parameter that is included in the vector β, e.g. as the first element, so that $\beta = [c, m, g, \beta_0', \beta_1']$.

  The corresponding state space form has $a_1 = 0, W_1 = [1, 1, 1, x_t', 0'], H_1 = 1$, and $W_t = [0, 1, t, x_t', x_{t-1}'], t > 1$.

- Notice that when y_1 is random and the process is supposed to have started in the indefinite past:

$$y_1 = \frac{1}{1-\phi}m + \frac{1-2\phi}{(1-\phi)^2}g + \sum_{j=1}^{\infty}\phi^{j-1}(x_{-j}'\beta_0 + x_{-j-1}'\beta_1) + \frac{1}{1-\phi L}\epsilon_1,$$

  This poses the issue either of making x_t available before the initial time and of truncating the infinite sum on the right hand side, or, in general of back-casting the regressors. Were x_t a random walk, the initialisation would be as in (7).

2.3.1 Encompassing the nonstationary case

The treatment has hitherto focused on the stationary case. Formally, the ADL(1,1) with $\phi = 1$ and $\beta_1 = -\beta_0$ nests the Fernández model using Δx_t as an indicator. Nevertheless, to allow for diffuse initialisation, due to nonstationarity, it is preferable to use a different parameterisation, which is obtained by replacing m and g from (3):

$$y_t \;=\; \phi(y_{t-1} + \gamma) + (1-\phi)(\mu + \gamma t) + \beta_0' x_t + \beta_1' x_{t-1} + \epsilon_t. \tag{8}$$

The process can be initialised with $y_1 = \mu + \gamma + (1-\phi)^{-1}x_t'(\beta_0 + \beta_1) + (1-\phi^2)^{-1/2}\epsilon_1$; treating μ as diffuse allows to incorporate the uncertainty about the cointegrating relationship between the aggregate series and the indicators, and effectively amounts to estimating the ADL model in first differences. The drift parameter γ can also be marginalised out of the likelihood (see Shephard, 1993), in which case the marginal likelihood is based on the second order differences of the observations; a problem arise instead with the marginalisation of the effects associated to the indicators; as a matter of fact, unless the common factor restriction is enforced, the data transformation which marginalises the parameters β_0, β_1 depends on ϕ.

Since later on we shall argue that the further flexibility of the Litterman model is more apparent than real, due to the problem one faces in its estimation, fitting (8) is the suggested strategy, if it is desired that the inferential process, concerning parameter estimation and disaggregation, embodies the uncertainty on the time series characterisation of the dynamic relationship between the variables.

The ADL(1,1) model can also be formulated in the first differences of y_t:

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

when x_t is the first difference of the variables entering the Litterman model, the latter is nested under the common factor restriction, $\beta_0 = -\phi\beta_1$. The Fernández model arises in two circumstances: when x_t is in levels, when $\phi = 0$ and $\beta_0 + \beta_1 = 0$; otherwise, if x_t is in the changes, when $\phi = 0$ and $\beta_1 = 0$.

The treatment of initial conditions proceeds along similar lines to the Litterman model. The state space form features two state elements: $\alpha_t = [y_{t-1}, \Delta y_t]'$, with $y_t = [1, 1]\alpha_t$, and the transition equation is the same as for Litterman model (see section 2.2), the only difference being the presence of regression effects in the transition equation.

3 Temporal aggregation and State space form

Assume now that y_t is not observed, but the temporally aggregated series is available, $\sum_{j=0}^{s-1} y_{\tau s-j}$ at times $\tau = 1, 2, \ldots [n/s]$, where $[k]$ denotes the integral part of k.

Following the approach proposed by Harvey (1989, sec. 6.3) the state space representation of the disaggregated model is adapted to the observational constraint by defining the cumulator variable

$$y_t^c = \psi_t y_{t-1}^c + y_t, \quad \psi_t = \begin{cases} 0, & t = s(\tau - 1) + 1, \tau = 1, \ldots, [n/s] \\ 1, & \text{otherwise} \end{cases} \tag{9}$$

In the case of quarterly flows whose annual total is observed ($s = 4$) the cumulator variable ($s = 4$) takes the following values:

$$\begin{array}{llll} y_1^c = y_1, & y_2^c = y_1 + y_2, & y_3^c = y_1 + y_2 + y_3, & y_4^c = y_1 + y_2 + y_3 + y_4 \\ y_5^c = y_5, & y_6^c = y_5 + y_6, & y_7^c = y_5 + y_6 + y_7, & y_8^c = y_5 + y_6 + y_7 + y_8, \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

Only a systematic sample of every s-th value of y_t^c process is observed, $y_{\tau s}^c, \tau = 1, \ldots [n/s]$, so that all the other values are missing.

The state space representation for y_t^c is derived as follows: first, replacing $y_t = z'\alpha_t$ in (9) substituting from the transition equation, rewrite

$$y_t^c = \psi_t y_{t-1}^c + z'T\alpha_{t-1} + (z'W_t + x_t')\beta + z'H\epsilon_t;$$

then, appending y_t^c to the original state vector, and defining $\alpha_t^* = [\alpha_t', y_t^c]'$

$$\begin{array}{rcl} y_t^c & = & z^{*'}\alpha_t^*, \\ \alpha_t^* & = & T_t^*\alpha_{t-1}^* + W_t^*\beta + H^*\epsilon_t, \quad \alpha_1^* = a_1^* + W_1^*\beta + H_1^*\epsilon_1, \end{array} \tag{10}$$

with $z^{*'} = [0' \; 1]$, and

$$T_t^* = \begin{bmatrix} T & 0 \\ z'T & \psi_t \end{bmatrix}, \; W_t^* = \begin{bmatrix} W_t \\ z'W_t + x_t' \end{bmatrix}, H^* = \begin{bmatrix} H \\ z'H \end{bmatrix},$$

$$a_1^* = \begin{bmatrix} a_1 \\ z'a_1 \end{bmatrix}, \; W_1^* = \begin{bmatrix} W_1 \\ z'W_1 + x_1' \end{bmatrix}, \; H_1^* = \begin{bmatrix} H_1 \\ z'H_1 \end{bmatrix},$$

Notice that the regression effects are all contained in the transition equation, and the measurement equation has no measurement noise. Moreover, the transition matrix T is time varying.

For the CL model (2), $\alpha_t^* = [\alpha_t, y_t^c]'$,

$$T_t^* = \begin{bmatrix} \phi & 0 \\ \phi & \psi_t \end{bmatrix}, W_t^* = \begin{bmatrix} 0 \\ x_t' \end{bmatrix}, H_t^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, a_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, W_t^* = \begin{bmatrix} 0 \\ x_1' \end{bmatrix}, H_1^* = \frac{1}{\sqrt{1-\phi^2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

4 Statistical treatment

The statistical treatment of model (10) is based upon the augmented Kalman filter (KF) due to de Jong (1991, see also de Jong and Chu-Chun-Lin, 1994), suitably modified to take into account the presence of missing values, which is easily accomplished by skipping certain updating operations.

The algorithm enables exact inferences in the presence of fixed and diffuse regression effects; the parameters β can be concentrated out of the likelihood function, whereas the diffuse case is accommodated by simple modification of the likelihood. See Koopman (1997) and Durbin and Koopman (2001) for an alternative exact approach to the initialisation under diffuseness and a comparison of the approaches.

The usual KF equations are augmented by additional recursions which apply the same univariate KF to k series of zero values, where k is the dimension of the vector β, with different regression effect in the state equation, provided by the elements of W_t.

Using the initial conditions in (10), and further defining $A_1^* = W_1^*$, $P_1^* = H_1^* H_1^{*'}$, $q_1 = 0, s_1 = 0, S_1 = 0$, the augmented Kalman filter consists of the following equations and recursions: for $t = 1, \dots, n$, and $t = \tau s, \tau = 1, \dots, [n/s]$ (y_t^c is available):

$$\begin{array}{lllllll}
v_t & = & y_t^c - z^{*'} a_t^*, & & V_t' & = & -z^{*'} A_t^*, \\
f_t & = & z^{*'} P_t^* z^*, & & K_t & = & T_t^* P_t^* z^* / f_t \\
a_{t+1}^* & = & T_{t+1}^* a_t^* + K_t v_t, & & A_{t+1}^* & = & W_{t+1}^* + A_t^* T_{t+1}^* + K_t V_t' \\
P_{t+1}^* & = & T_{t+1}^* P_t^* T_{t+1}^{*'} + H_t^* H_t^{*'} - K_t K_t' f_t, & & & & \\
q_{t+1} & = & q_t + v_t^2 / f_t, & & s_{t+1} & = & s_t + V_t v_t / f_t \\
S_{t+1} & = & S_t + V_t V_t' / f_t & & d_{t+1} & = & d_t + \ln f_t
\end{array} \qquad (11)$$

Else, for $t \neq \tau s$ (y_t^c is missing),

$$\begin{array}{lll}
a_{t+1}^* & = & T_{t+1}^* a_t^*, \quad A_{t+1}^* = W_{t+1}^* + A_t^* T_{t+1}^*, \quad P_{t+1}^* = T_{t+1}^* P_t^* T_{t+1}^{*'} + H_t^* H_t^{*'}, \\
q_{t+1} & = & q_t, \quad s_{t+1} = s_t, \quad S_{t+1} = S_t;
\end{array} \qquad (12)$$

The symbol V_t' denotes a row vector with k elements. The quantities q_t, S_t, s_t accumulate weighted sum of squares and cross-products that will serve the estimation of β via generalised regression.

It should be noticed that the quantities f_t, K_t (Kalman gain) and P_t do not depend on the observations and that the first two are not computed when y_t^c is missing. Missing values imply that updating operations, related to the new information available, are skipped.

The augmented KF computes all the quantities that are necessary for the evaluation of the likelihood function.

4.1 Maximum Likelihood Estimation

Under the fixed effects model, as shown in Rosenberg (1973), maximising the likelihood with respect to β and σ^2 yields:

$$\hat{\beta} = -S_{n+1}^{-1} s_{n+1}, \text{Var}(\hat{\beta}) = S_{n+1}^{-1}, \quad \hat{\sigma}^2 = \frac{q_{n+1} - s_{n+1}' S_{n+1}^{-1} s_{n+1}}{[n/s]}, \tag{13}$$

The profile likelihood is

$$\mathcal{L}_c = -0.5 \left[d_{n+1} + [n/s] \left(\ln \hat{\sigma}^2 + \ln(2\pi) + 1 \right) \right]; \tag{14}$$

it is a function of the parameter ϕ alone. Thus maximum likelihood estimation can be carried out via a grid search over the interval (-1,1).

When β is diffuse (de Jong, 1991), the maximum likelihood estimate of the scale parameter is

$$\hat{\sigma}^2 = \frac{q_{n+1} - s_{n+1}' S_{n+1}^{-1} s_{n+1}}{[n/s] - k},$$

and the diffuse profile likelihood, denoted $\mathcal{L}_\infty$, takes the expression:

$$\mathcal{L}_\infty = -0.5 \left[d_{n+1} + ([n/s] - k) \left(\ln \hat{\sigma}^2 + \ln(2\pi) + 1 \right) + \ln |S_{n+1}| \right]. \tag{15}$$

The diffuse likelihood is based on a linear transformation (e.g. first order differences) that makes likelihood of the transformed data invariant to β. This yields estimators of ϕ with better small sample properties, as it will be illustrated in section 6.1, in agreement with Tunnicliffe-Wilson (1986) and Shephard and Harvey (1990).

Some care about the parameterisation chosen should be exercised when ϕ is close to 1: this point can be illustrated with respect to the simple CL model with a constant term, which admits two representations: (A) $y_t = \mu + \epsilon_t/(1 - \phi L)$ and (B) $y_t = \phi y_{t-1} + m + \epsilon_t$; for both (A) and (B) assume $y_1 = \mu + (1 - \phi^2)^{-1/2} \epsilon_1$.

As hinted in section 2.1, taking first differences yields a strictly noninvertible ARMA(1,1) process that does not depend on μ in case (A) and on m in case (B), except when $\phi = 1$, for which $\Delta y_t = m + \epsilon_t$ and further differencing would be required to get rid of the nuisance parameter m. Hence, setting aside for a moment the temporal aggregation problem, $\mathcal{L}(\Delta y_2, \ldots, \Delta y_n)$ cannot be made independent of m when $\phi = 1$.

Denoting by $\mathcal{L}_\mu$ and $\mathcal{L}_m$ the two diffuse likelihood under (A) and (B), it can be shown (proof available from the author) that $\mathcal{L}_\mu = \mathcal{L}_m + 2\ln(1-\phi)$, which has relevant implications if $\phi \to 1$ (see section 6.1); all other inferences are the same.

If it is suspected that $\phi = 1$ is a likely occurrence, due to the nonstationarity of the series and the lack of cointegration with the indicators, only diffuseness under case (A) provides the correct solution.

Hence, a difference arises in the definition of the diffuse likelihood according to the parameterisation of the deterministic kernel. The ambiguity arises from the fact that the transformation adopted (differencing) is dependent on the parameter ϕ. See Severini (2000, sec. 8.3.4) for a general treatment of marginal likelihoods based on parameter dependent transformations.

The ambiguity is resolved anyway if (B) is reparameterised as $y_t = \phi y_{t-1} + (1-\phi)\mu + \epsilon_t, t > 1$, in which case the two diffuse likelihood can be shown to be equivalent.

4.2 Diagnostic checking and disaggregated estimates

Diagnostics and goodness of fit are based on the innovations, that are given by $\tilde{v}_t = v_t - V_t' S_t^{-1} s_t$, with variance $\tilde{f}_t = f_t + V_t' S_t^{-1} V_t$. As illustrated in section 6.3, the standardised innovations, $\tilde{v}_t/\sqrt{\tilde{f}_t}$ can be used to check for residual autocorrelation and departure from the normality assumption. They also are a good indicator of the process of revision of the disaggregated estimates.

The filtered, or real time, estimates of the state vector and their estimation error matrix are computed as follows: $\tilde{\alpha}^*_{t|t} = a^*_t - A^*_t S_t^{-1} s_t + P^*_t z^* \tilde{v}_t / f_t$, $P^*_{t|t} = P^*_t + A^*_t S_t^{-1} A^{*\prime}_t - P^*_t z^* z' P^*_t / f_t$.

The smoothed estimates are obtained from the augmented smoothing algorithm proposed by de Jong (1988), appropriately adapted to hand missing values. Defining $r_n = 0, R_n = 0, N_n = 0$, for $t = n, \ldots, 1$, and $t = \tau s, \tau = 1, \ldots, [n/s]$ (y_t^c is available):

$$\begin{array}{rcl} r_{t-1} & = & z^* v_t / f_t + (T_{t+1} - K_t z^{*\prime}) r_t, \quad R_{t-1} = z^* V_t' / f_t + (T_{t+1} - K_t z^{*\prime}) R_t, \\ N_{t-1} & = & z^* z^{*\prime} / f_t + (T_{t+1} - K_t z^{*\prime}) N_t (T_{t+1} - K_t z^{*\prime})' \end{array}$$

Else, for $t \neq \tau s$ (y_t^c is missing),

$$r_{t-1} = T_{t+1} r_t, \quad R_{t-1} = T_{t+1} R_t, \quad N_{t-1} = T_{t+1} N_t T_{t+1}'.$$

The smoothed estimates are obtained as

$$\begin{array}{rcl} \tilde{\alpha}^*_{t|n} & = & a^*_t + A^*_t \tilde{\beta} + P^*_t (r_{t-1} + R_{t-1} \tilde{\beta}) \\ P^*_{t|n} & = & P^*_t + A^*_t S_{n+1}^{-1} A^{*\prime}_t - P^*_t N_{t-1} P^*_t \end{array}$$

The provides disaggregated estimates of the cumulator $\tilde{y}_t^c$ (corresponding to the last component of the state vector); the latter can be decumulated by inverting (9) so as to provide estimates of the disaggregated series; however, its estimation error variance is not available.

To remedy this situation, the strategy is to augment the state vector by appending y_t to it; the new transition equation is derived straightforwardly writing:

$$y_t = z' \alpha_t + x_t \beta = z' T \alpha_{t-1} + (x_t + z' W_t) \beta + z H \epsilon_t.$$

5 Logarithmic transformation and nonlinear disaggregation

Time series models for flow variables measured on a ratio scale (such as production, turnover, value added, revenues, etc.) are usually formulated in terms of the logarithmic transformation of the disaggregated values.

The assumptions underlying the disaggregated model (additivity of effects, normality and homoscedasticity of errors) appear more suitable for the logarithms, rather then the levels of economic flows. See also Banerjee *et al.* (1993), section 6.3, for further arguments and discussion concerning the modelling of the logarithms versus the levels of an economic time series.

Assuming that y_t denotes the logarithm of the disaggregated values, for which the dynamic regression model $y_t = z'\alpha_t + x_t'\beta$ of section 2 applies, the aggregate value of a flow can be expressed in terms of the y's as follows:

$$Y_\tau = \sum_{j=0}^{s-1} \exp(y_{\tau s-j}), \tau = 1, \dots, [n/s]; \tag{16}$$

For instance, when $s = 4$, the linear time series model is formulated for the logarithms of the quarterly values, y_t, but the available data are only annual, originating from the sum of the levels of the flow variable over the four quarters making up the year.

It should be noticed that (16) is a nonlinear aggregation constraint which is responsible eventually for the nonlinearity of the aggregated model. The general statistical treatment of disaggregation under the class of Box-Cox (1964) transformations is provided in Proietti (2004); Proietti and Moauro (2003) present an application to the problem of estimation an index of coincident indicators and a monthly GDP series using data with different frequency of observations within the Stock and Watson (1991) dynamic factor model.

Statistical inference is carried out by an iterative method that at convergence provides a disaggregated series $\{\hat{y}_t, t = 1, \dots, n\}$, that is a constrained posterior mode estimate of the unknown $\{y_t\}$, which satisfies the nonlinear observational constraint (16) exactly.

Defining the cumulator on the original scale of measurement,

$$Y_t^c = \psi_t Y_{t-1}^c + \exp(y_t), \tag{17}$$

and denoting by $\tilde{y}_t = z'\tilde{\alpha} + x_t'\tilde{\beta}, t = 1, \dots, n,$ a trial disaggregated series (which needs not satisfy the constraint (16)), let us consider the first order Taylor approximation of $\exp(y_t) = \exp(z'\alpha + x_t'\beta)$ around the trial values $[\tilde{\alpha}', \tilde{\beta}']'$, giving:

$$\begin{array}{rcl} Y_t^c & = & \psi_t Y_{t-1}^c + \tilde{z}_t'\alpha_t + \tilde{x}_t'\beta + \tilde{d}_t, \\ \tilde{z}_t' & = & \exp(\tilde{y}_t)z', \quad \tilde{x}_t' = \exp(\tilde{y}_t)x_t', \quad \tilde{d}_t = (1-\tilde{y}_t)\exp(\tilde{y}_t) \end{array} \tag{18}$$

Conditional on $\tilde{y}_t$, equation (18) is linear and, replacing α_t with the right hand side of the transition equation, it is used to augment the state space representation (1). The only relevant change with respect to the linear case is the inclusion of the sequence d_t and the time varying nature of the system matrices.

Hence, conditional on $\tilde{y}_t$ the linear Gaussian approximating model (LGAM) is formulated for the state vector $\alpha_t^* = [\alpha_t', Y_t^c]'$ as follows:

$$\begin{array}{rcl} Y_t^c & = & [0' \; 1]\alpha_t^*, \\ \alpha_t^* & = & \tilde{T}_t^* \alpha_{t-1}^* + [0', \tilde{d}_t]' + \tilde{W}_t^* \beta + \tilde{H}^* \epsilon_t, \quad \alpha_1^* = a_1^* + [0', \tilde{d}_1]' + \tilde{W}_1^* \beta + \tilde{H}_1^* \epsilon_1 \end{array} \tag{19}$$

where

$$T_t^* = \begin{bmatrix} T & 0 \\ \tilde{z}_t' T & \psi_t \end{bmatrix}, \; W_t^* = \begin{bmatrix} W_t \\ z' W_t + \tilde{x}_t' \end{bmatrix}, H^* = \begin{bmatrix} H \\ \tilde{z}_t' H \end{bmatrix},$$

$$a_1^* = \begin{bmatrix} a_1 \\ \tilde{z}_1' a_1 \end{bmatrix}, \; W_1^* = \begin{bmatrix} W_1 \\ \tilde{z}_1' W_1 + \tilde{x}_1' \end{bmatrix}, \; H_1^* = \begin{bmatrix} H_1 \\ \tilde{z}_1' H_1 \end{bmatrix},$$

Applying the augmented KF and smoother of section 4 a new disaggregated series $\{\hat{y}_t\}$ is computed. Setting $\tilde{y}_t = \hat{y}_t$ and replacing the latter into (19) a new LGAM is obtained; iterating the process until convergence provides the required solution. The likelihood of the nonlinear model is approximated by that of the LGAM at convergence. An illustration is provided in section 6.5.

As shown in Proietti (2004), the iterative algorithm outlined above is a *sequential linear constrained optimisation method,* see Gill et al. (1989, sec. 7), that rests upon the linearisation of the constraint around a trial value, which does not need to be feasible. It can be viewed as a particular case of the recursive conditional mode by extended Kalman filtering and smoothing proposed on Durbin and Koopman (1992, 2001) and Fahrmeir (1992).

We stress that the iterative method differs from the technique known as extended Kalman filter (see Anderson and Moore, 1979, Harvey and Pierse, 1984, sec. 5), which is the non-iterative method that replaces the unknown states affecting the system matrices by the conditional expectation of α_t given the past observations up to time $t-1$ in the Taylor expansion.

As a matter of fact, straight application of the extended Kalman filter and smoother to the linearised model would not lead a feasible estimate of α, and a subsequent adjustment would be required to enforce the aggregation constraint. This approach was adopted by Proietti (1998), who distributed the approximation errors according to the Denton method (Denton, 1971). Mitchell *et al.* (2004) provide an alternative non iterative solution which yields an approximate solution.

According to our approach, on the other hand, the evaluation of the likelihood is based on a linearised model that has the same posterior mode for the states α, and thus for the missing observations, as the true nonlinear model.

6 Illustrations

This section provides five illustrations. The first two concern the CL disaggregation method and deal with (i) the role of marginal likelihood diffuse initial conditions for the estimation of the autoregressive parameter; (ii) the benefits of nesting CL within a more general ADL model.

The third uses a real time experiment to set up revision histories that help in choosing between different methods and/or deterministic kernels.

The fourth concerns the properties of the maximum likelihood estimator of the AR parameter in the Litterman model, in particular, when the true ϕ is equal to zero (Fernandez, 1981).

Our final illustration deals with the nonlinear disaggregation on the logarithmic scale of a time series characterised by the presence of small positive values, for which linear disaggregation methods fail to provide admissible estimates.

The real life applications refer to a dataset, consisting of a set of annual series and the corresponding quarterly indicators, made available by the Italian National Statistical Institute, Istat, consists of a set of annual series and the corresponding quarterly indicatorwho has recently established a commission aiming at the assessment of its current temporal disaggregation methodology and practices, that are based primarily, if not exclusively, on the CL procedure.

A set of Ox[1] functions implementing the methods presented in this paper are available from the author upon request and is documented briefly in the appendix.

6.1 The CL model with diffuse regression effects

We report the results of two small, but representative, Monte Carlo experiment dealing with the estimation of the AR parameter, ϕ, under the two alternative assumptions on the regression effects, β, fixed or diffuse.

The first consists of simulating $M = 1000$ series of length $n = 120$ from the true model $y_t = \mu + x_t + u_t$, with $\mu = 0.5$, x_t is the random walk process $\Delta x_t = 0.5 + \zeta_t$, $\zeta_t \sim \text{NID}(0, 1)$, and u_t is a stationary AR(1) process with $\phi = 0.75$: $u_t = 0.75u_{t-1} + \epsilon_t$, $\epsilon_t \sim \text{NID}(0, 0.8)$.

The generated series are then aggregated with aggregation period $s = 4$, and the CL model with a constant term and regression effects is estimated treating β respectively as a fixed unknown vector and a random diffuse vector. We refer to section 4.1 for the choice of the parameterisation of the constant term.

The distributions of the estimated ϕ coefficients are presented in the first panel of figure 1. Both estimators suffer from a downward bias, but this is larger for the fixed case (-0.17, in the diffuse case being equal to -0.05). The estimation mean square errors are respectively 0.206 and 0.052 for the fixed and diffuse case, implying that assuming diffuse effects greatly improves the estimates.

In the second Monte Carlo experiment the disaggregated data are generated by the random walk with drift: $y_t = y_{t-1} + 0.5 + \epsilon_t$, $\epsilon_t \sim \text{NID}(0, 0.5)$. the other design parameter remain the same and there are no exogenous effects.

The generating model has $\phi = 1$ and is thus nonstationary. Again the treatment of μ as a diffuse effect in the estimation provides a substantive improvement for the properties of the estimates, as highlighted by the comparison of the two distributions of the ϕ estimates over the M replications, represented in the second panel of figure 1.

[1] Ox is a matrix programming language developed by J.A. Doornik (2001).

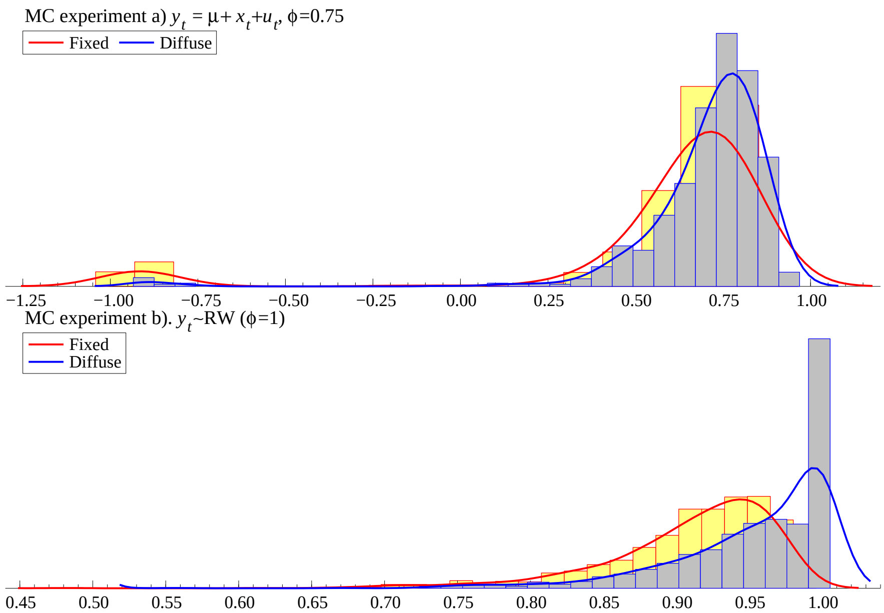

Figure 1: Histograms and nonparametric density estimates of the estimated ϕ coefficient of the CL model under fixed and diffuse effects.

The bias and the estimation mean square error amount respectively to to -0.09 and 0.012 for the fixed case and to -0.05 and 0.005 for the diffuse case.

6.2 Common factor analysis: the ADL(1,1) model versus Chow-Lin

The estimation of the quarterly total production series for the Italian insurance sector from the annual series, carried out by Istat using the CL method, is considered problematic due to the scarce adequacy of the quarterly indicator. The latter is reproduced, along with the the annual figures to be distributed, in the first panel of figure 2. The plot reveals that the indicator is highly volatile towards the end of the sample, being more prone to outliers and level shifts.

In a situation like the present one, a crucial concern is the "volatility transfer" from the indicators to the disaggregated estimates, that originates from the application of the CL procedure.

Before turning to this issue, let us present the solutions presented by the ADL(1,1) and Chow-Lin with regression effects represented by a deterministic trend and the indicator x_t. It should be recalled from section 2.3 that the former nests the latter.

As far as the ADL(1,1) model is concerned, the maximised profile likelihood (plotted in the second panel of figure 3) is -204.30, which corresponds to value $\phi = 0.72$. Moreover, $\hat{\beta}_0 = -0.04$ (standard error 0.08), $\hat{\beta}_0 = 0.17$ (standard error 0.09), $\hat{\sigma}^2 = 11607$. This would suggest dropping x_t, while retaining x_{t-1} (due to the nonstationarity of x_t collinearity, and the reliability of standard errors, is an issue here).

For the Chow-Lin model the estimated AR parameter (restricting the grid search in the positive range) is 0.43, with the maximised profile likelihood taking the value -212.64; also, $\hat{\beta} = 0.40,$ (standard error 0.03), $\hat{\sigma}^2 = 44712$.

Hence, the likelihood ratio statistic of the restriction $\beta_1 = -\phi\beta_0$ in the ADL model takes the value 16.69 and clearly does lead to a rejection. This implies that the CL specification is not supported in this example.

The plot of the disaggregate series (figure 2) reveals interesting interpretive features of the two methods. In particular, the CL estimates are affected to a greater extent by the volatility of the indicator, whereas the sharp movements in the indicator appear more smoothed in the ADL estimates.

Now, recall that for scalar x_t the CL disaggregation is based on $y_t = \beta x_t + u_t$ with u_t ∼AR(1) (possibly with a non zero mean), whereas for the ADL(1,1) case,

$$y_t = \beta_0 \sum_j \phi^j x_{t-j} + \beta_1 \sum_j \phi^j x_{t-j-1} + u_t.$$

The first two terms on the right hand side can be interpreted as a distributed lag function of a filtered version of x_t, where the current and past values of x_t receive weights that decline geometrically over time.

If the indicator is affected by measurement error, so that it can be written $x_t = \bar{x}_t + \eta_t, \eta \sim$ NID$(0, \sigma^2_\eta)$, and the signal $\bar{x}_t$ is generated by a random walk with disturbance variance $\sigma^2_x = (1-\phi)^2\sigma^2_\eta/\phi$, then it can be shown that $E(\bar{x}_t|x_t, x_{t-1}, \dots)$ is proportional to $x_t + \phi x_{t-1} + \phi^2 x_{t-2} + \cdots = (1-\phi L)^{-1}x_t$.

Our second example supports instead the CL restriction and deals with the total production of the Wholesale and Retail Trade sector (Istat B4 series), whose annual values are plotted in the first panel of figure 3, along with the quarterly indicator constructed by Istat.

As far as the ADL(1,1) model is concerned, the maximised profile likelihood (plotted in the second panel of figure 3) is -241.52, which corresponds to value $\phi = 0.72$. Moreover, $\hat{\beta}_0 = 0.55, \hat{\beta}_0 = -0.43$ (standard error 0.12), $\hat{\sigma}^2 = 1.8583 \times 10^5$.

For the Chow-Lin model the estimated AR parameter is the essentially the same (0.72), with the maximised profile likelihood taking the value -241.76; also, $\hat{\beta} = 0.49,$ (standard error 0.08), $\hat{\sigma}^2 = 1.8671 \times 10^5$.

Thus the LR statistic of the restriction $\beta_1 = -\phi\beta_0$ is 0.47 and clearly does not lead reject the null. Figure 3 makes clear that the two models produce the same disaggregated series.

6.3 Diagnostics and reliability of estimates

The European System of National and Regional Accounts (par. 12.04) envisages the following selection criterion:

> *The choice between the different indirect procedures must above all take into account the minimisation of the forecast error for the current year, in order that the provisional annual estimates correspond as closely as possible to the final figures.*

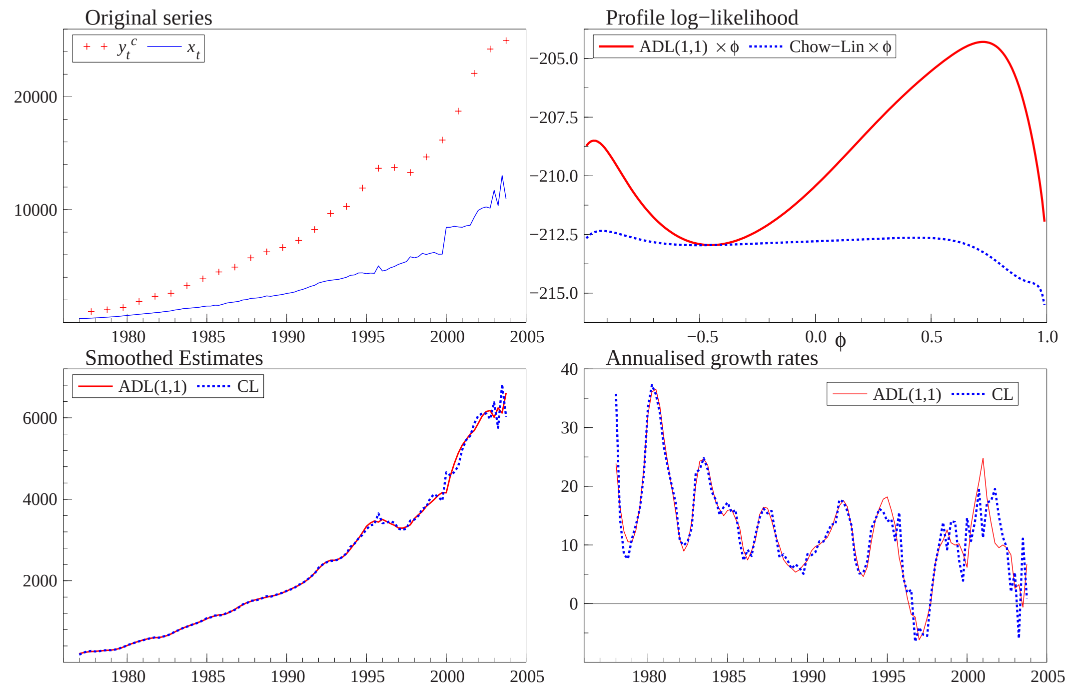

Figure 2: Total Production Insurance sector (Istat A2 series). Comparison between the ADL(1,0) model and Chow-Lin.

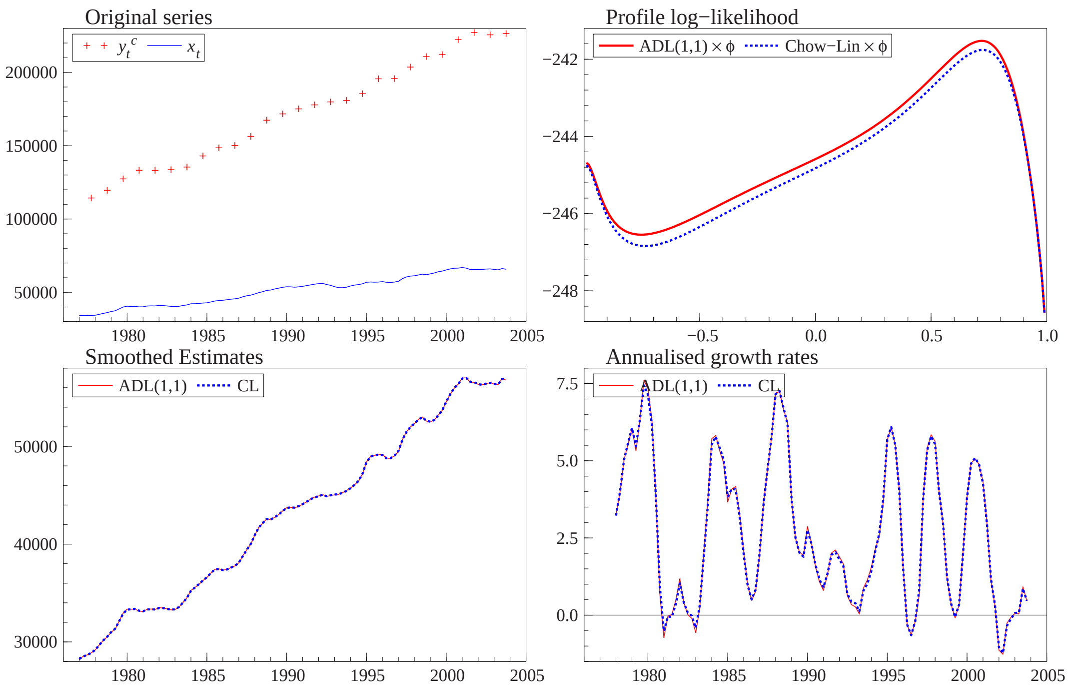

Figure 3: Total Production Wholesale and Retail Trade (Istat B4 series).Comparison between the ADL(1,0) model and Chow-Lin.

This criterion, which refers to the discrepancy between the estimates not using the last aggregate data and those incorporating it, is ambiguous in that it considers the estimates incorporating the yearly figure as final, whereas any disaggregation filter will use also future aggregate observations, the CL with $\phi = 0$ being the only notable exception.

On the other hand, it clearly points out that the decision between alternative methods should be based on a careful assessment of the revision of the estimates as the new total, sometimes referred to as benchmark, becomes available.

In this illustration we compare the revision histories of three methods of disaggregating the total production of the Metals sector (Istat B2 series), available for the years 1977-2003 and plotted in the first panel of figure 4 along with its indicator: CL with regression effects represented by a constant and the indicator; CL with a linear trend and the indicator; the Fernández model with a constant and the indicator.

The filtered and smoothed estimates arising from the Fernández are reproduced in figure 4, whereas figure 5 plots the standardised innovations, their estimated density and correlogram and compares the revision history with the CL model with a trend.

The revision histories are generated as follows: starting from 1992 we perform a rolling forecast experiment such that at the beginning of the year we make predictions for the four quarters using the information available up to the beginning of the year and revise the estimates concerning the four quarters of the previous year.

This assumes that the annual aggregate for the year τ accrues between the end of the quarter τs and the beginning of quarter $\tau s + 1$. At the end of the experiment 12 sets of predictions are available for four horizons (one quarter to four quarters); these are compared with the revised estimates, which incorporate the annual aggregate information. The models are re-estimated at each new annual observations.

Table 1 presents summary statistics pertaining to the revision histories at the four horizons.

The first rather obvious piece of evidence is that the performance of the methods deteriorates with the horizons, except for the CL with a trend. Secondly, the random walk model (Fernández) specifications outperforms according to all the measures presented.

The plot of the percentage revision errors in the last panel of figure 5 points out that the extent of the revision can be anticipated from the standardised innovations; it also reveals that the performance of the CL with trend model is strongly influenced by the inability to predict the 1996 expansion.

6.4 Litterman model: small sample properties of ϕ estimates

The previous illustration showed that the Fernández model outperformed the Chow-Lin specification in terms of the revision errors; as the former is also a particular case of the Litterman model arising when the AR parameter is zero, one would expect that, when the series is not cointegrated at the long run frequency with the indicator, the correct strategy would be to fit the Litterman model in the first place and to entertain the Fernández model only when ϕ is not significantly different from zero.

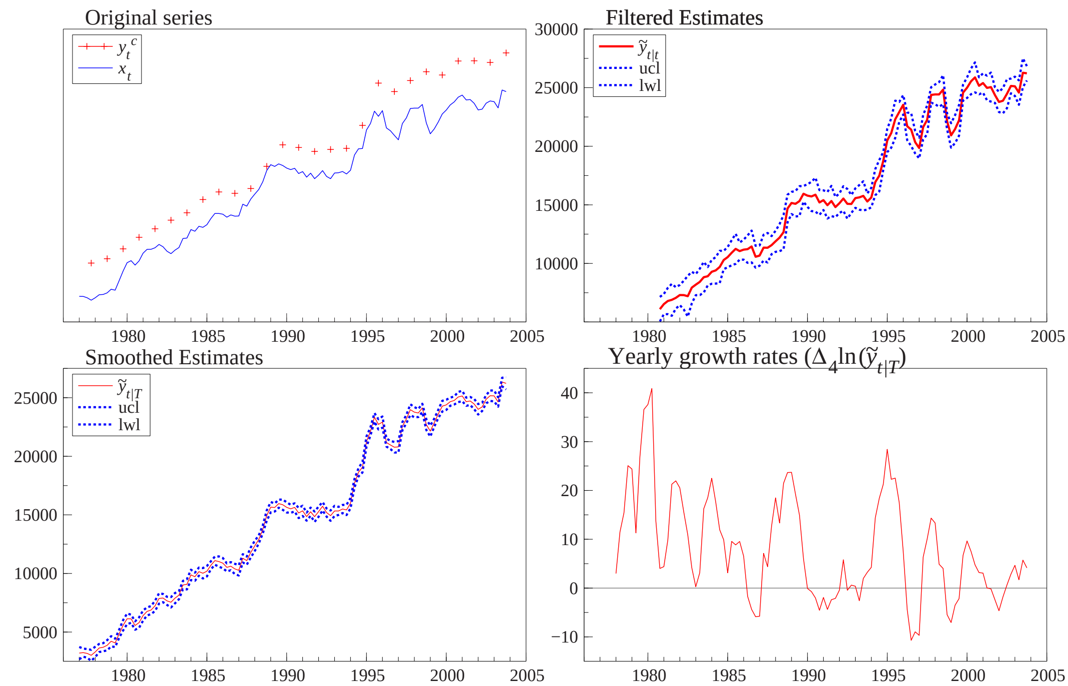

Figure 4: Total Production Metals (Istat B2 series). Filtered and smoothed estimates of the disaggregated series obtained by the Fernandez model.

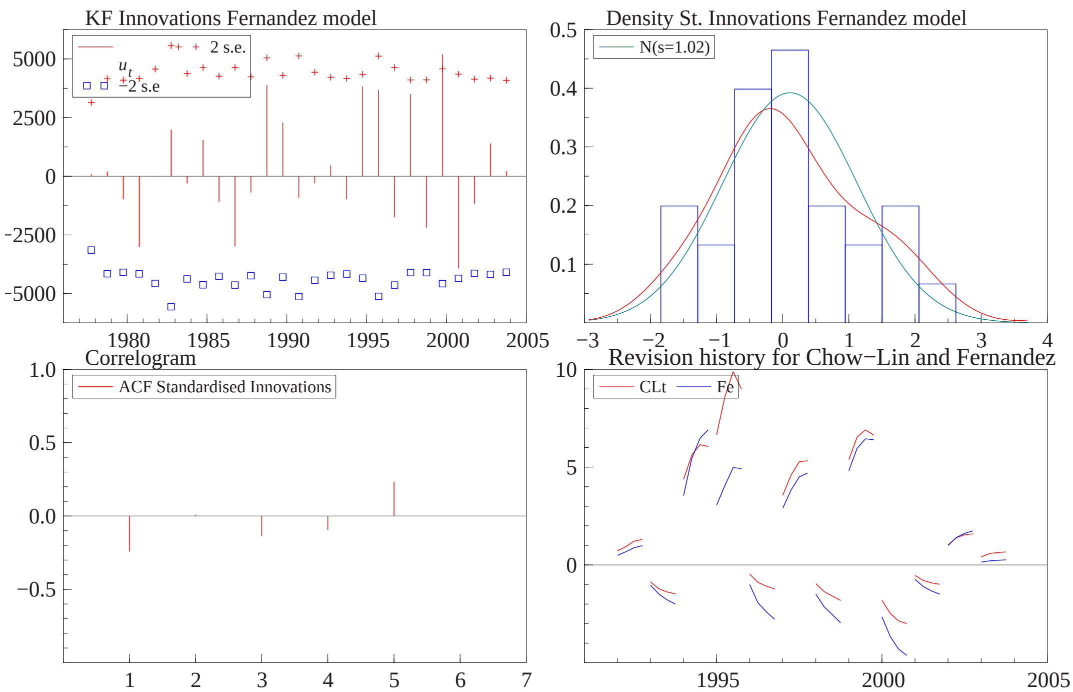

Figure 5: Total Production Metals (Istat B2 series). Standardised innovations for the Fernandez model and revision histories for the Chow-Lin and Fernandez models

This section investigates the properties of the maximum likelihood estimates of the ϕ parameters of the Litterman model, and concludes that these are fairly problematic at the very least.

For the series considered in the above illustration (Istat B2 series) the estimate of ϕ resulted -0.73, even though this is not significantly different from zero, which can be considered as a fairly emblematic outcome, as confirmed by our empirical experience with the Istat dataset: out of 16 series investigated 10 ϕ estimates were in the range (-1, -0.5), 3 in (-0.5, 0), only one in (0,0.5) and three were greater than 0.5. In spite of their polarisation, the likelihood was very flat and the estimates were not significantly different from zero.

The reason is that for $s = 4$, despite the ϕ parameter being theoretically identified, the data are not very informative about the AR parameter, even in large samples, unless the true value is indeed close to the extremes of the stationary range.

The behaviour of the large sample likelihood function of the Litterman model can be investigated using the approach by Palm and Nijman (1984).

Letting z_t denote the sum of s consecutive observations, $z_t = S(L)y_t$, with $S(L) = 1 + L + L^2 + \cdots L^{s-1}$, originating from the ARIMA(1,1,0) process $\Delta y_t = \phi \Delta y_{t-1} + \epsilon_t$; it immediately follows that

$$\Delta_s z_t = \phi^s \Delta_s z_{t-s} + S(L)^2(1 + \phi L + \phi^2 L^2 + \cdots + \phi^{s-1} L^{s-1})\epsilon_t,$$

where $\Delta_s = 1 - L^s$.

The first differences of the aggregated process arise as a systematic sample with step s of $\Delta_s z_t$, and are the ARMA(1,1) process: $\Delta Z_\tau = \phi^s Z_{\tau-1} + (1 + \theta L)\xi_\tau$, where the lag operator applies to the τ index, that is $\Delta Z_\tau = Z_\tau - Z_{\tau-1}$, with AR parameter ϕ^s and moving average parameter, θ, provided by the invertible root of the quadratic equation $\frac{\theta}{1+\theta^2} = \rho(s)$, where $\rho(s) = \gamma(s)/\gamma(0)$ and $\gamma(j)$ is the lag j autocovariance of the process $S(L)^2(1 + \phi L + \phi^2 L^2 + \cdots + \phi^{s-1} L^{s-1})\epsilon_t$; $\sigma^2_\xi = \gamma(0)/(1 + \theta^2)$.

Letting $g_\phi(\omega) = (1 + \theta^2 + 2\theta \cos\omega)/(1 + \phi^{2s} - 2\phi^s \cos\omega)$, so that $g_\phi(\omega)\sigma^2_\xi$ denotes the spectral generating function of ΔZ_τ the large sample log-likelihood evaluated at $\tilde{\phi}$ is

$$\mathcal{L}(\tilde{\phi}) \approx -\frac{1}{2}\left[\frac{n}{s}\right]\left[1 + \ln \int_0^\pi \frac{g_\phi(\omega)}{g_{\tilde{\phi}}(\omega)} \sigma^2_\xi d\omega\right] \tag{20}$$

The integral represents the variance of the residual $\tilde{\xi}_t = (1 + \tilde{\theta} L)^{-1}(1 - \tilde{\phi L})\Delta Z_\tau$ for the Litterman model with $\tilde{\phi}$.

When $\phi = 0$ the true disaggregated data are generated by the Fernández model (a random walk), and the aggregate model for Z_t is IMA(1,1).

In figure 6 we use the effective device, adopted by Palm and Nijman (1984), of plotting $\mathcal{L}(\tilde{\phi})$ for the values of the argument that are not significantly different from the true value $\phi = 0$, which occurs when $2[\mathcal{L}(0) - \mathcal{L}(\tilde{\phi})] < \chi^2_{0.95}(1)$ (obviously $\mathcal{L}(0)$ is a global maximum), with $\chi^2_{0.95}(1)$ representing the 95-th percentile of the chisquare distribution with 1 degree of freedom (3.84), and the fixed value $\mathcal{L}(0) - 0.5\chi^2_{0.95}(1)$ otherwise.

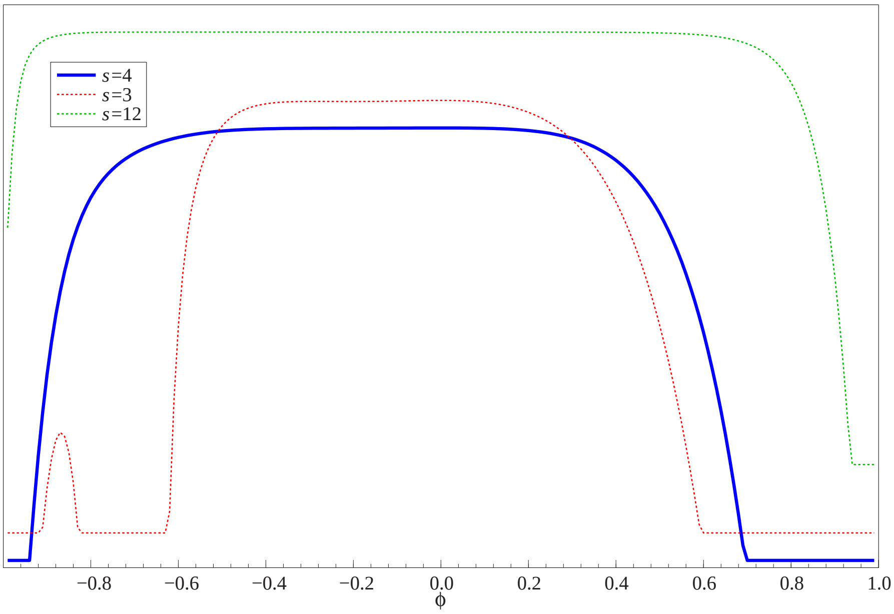

Figure 6: Large sample profile log-likelihood for ϕ coefficients when the true disaggregated model is a random walk (Fernández model).

The plot immediately reveals that the likelihood is very flat, and that the range of values not significantly different from zero covers a large part of the stationary region of the parameter. When the aggregation interval is $s = 3$ a secondary mode of the likelihood appears, and the situation worsens as s increases.

The previous analysis rests upon a large sample approximation. The small sample properties of the MLE of ϕ in the Litterman model when the true value is $\phi = 0$ and the aggregation interval is equal to $s = 4$ are investigated by a MC experiment, designed as follows: $M = 1000$ series of length $n = 120$ are simulated from the process $y_t = x_t + u_t$, where u_t is a random walk $\Delta u_t = \epsilon_t$, $\epsilon_t \sim$ NID$(0, 0.5)$, and $\Delta x_t = 0.5 + \zeta_t$, $\zeta_t \sim$ NID$(0, 1)$.

The series are aggregated into the sum of four consecutive values ($s = 4$ giving 40 yearly observations) and the Litterman model with a constant term (assuming unknown initial conditions) is fitted; for comparison we also estimated two encompassing dynamic regression models, namely the ADL(1,1) model in differences using the levels of x_t as an indicator $\Delta y_t = \phi \Delta y_{t-1} + m + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$, and the ADL(1,1) using first differences of x_t as an explanatory variable: $\Delta y_t = \phi \Delta y_{t-1} + m + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \epsilon_t$,

The density of the estimates of ϕ are displayed in figure 7, which reveals the distribution of the Litterman's estimates is inherently bimodal, so that the bulk of the estimates is far away from the true value. The true value is close to a minimum of the density.

The situation improves substantially when the ADL model using x_t is adopted; the estimates are less biased and the mode is close to the true value; moreover, although not reported here.

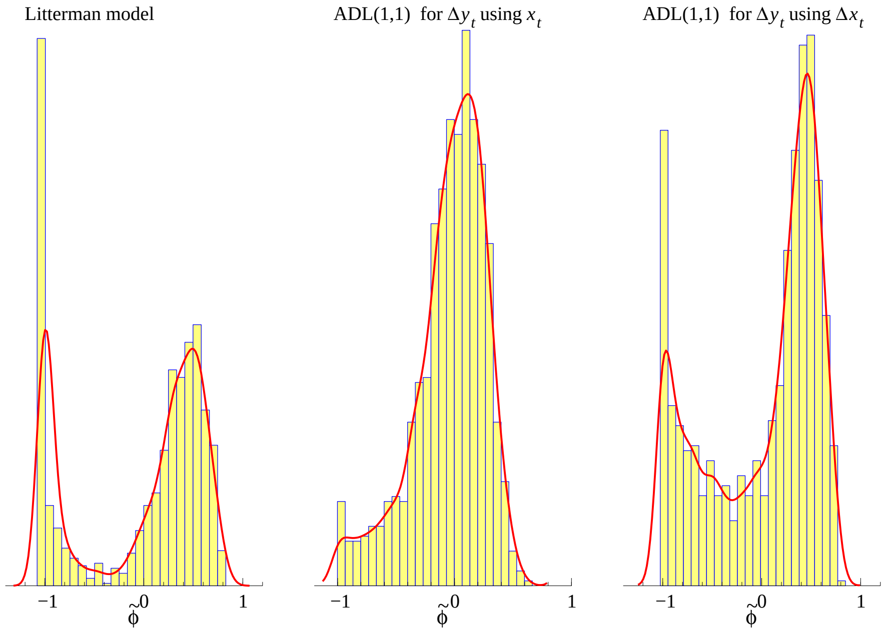

Figure 7: Histograms and nonparametric density estimates of the estimated ϕ coefficient for disaggregated data generated according to $y_t = x_t + u_t$, where u_t is a random walk, $\Delta u_t = \epsilon_t$, $\epsilon_t \sim \text{NID}(0, 0.5)$, and $\Delta x_t = 0.5 + \zeta_t$, $\zeta_t \sim \text{NID}(0, 1)$.

the estimates of the regression coefficient satisfy $\beta_0 + \beta_1 = 1$. Finally, the ADL(1,1) using Δx_t, improves somewhat, but not decisively.

In conclusion, the maximum likelihood estimates of the ϕ parameter characterising the Litterman model should be taken with great care; it is usually fruitful to plot the profile likelihood, highlighting the set of values not significantly different from the maximiser, to perceive the reliability of the point estimates. It is also important to compare results with the ADL model in first differences.

6.5 Nonlinear Disaggregation

Our last illustration concerns the annual tax revenues on methane gas (Istat series A1), which is displayed in the first panel of figure 8; the series is disaggregated by Istat using the Chow-Lin method using as indicator measuring the quantity of methane gas sold (seasonally unadjsted, also reproduced in figure 8).

We leave aside the issue posed by the presence of seasonality. It suffices to say that Istat produces a ”seasonally unadjusted” version of the quarterly accounts that postulates that the seasonality in the aggregated series is proportional to that in the indicator, where the factor of proportionality is the regression coefficient β; if the indicator is integrated at the seasonal frequencies, the underlying assumption is that, the disaggregated series and the indicator will be seasonally cointegrated at those frequencies.

The disaggregation of the annual revenue totals provides an interesting testbed since, due to the small numbers involved, the estimated quarterly series can take on negative values, which are outside the admissible range of the series. As a matter of fact, standard application of the Chow-Lin procedure with a constant term ($\hat{\phi} = 0.36$ and $\beta = 0.98$, with standard error 0.04) yields the disaggregated series plotted in the second panel of figure 8, which indeed becomes negative in coincidence with a few seasonal troughs.

There are two arguments in favour of the nonlinear disaggregation procedure outlined in section 5. First and foremost, working on the logarithmic scale guarantees that the levels of the estimated quarterly series take only positive values, within the admissible range; moreover, the logarithmic transformation mitigates if not eliminates the heteroscedasticity of the series. At least the logarithms of x_t do not display the increase in the amplitude of the seasonal fluctuations which characterises the levels; see the last panel of figure 8.

When the nonlinear CL model with a constant term is fitted to the series the estimates the parameters result $\hat{\phi} = 0.27$, $m = -0.53$, and $\beta = 1.09$. The disaggregated estimates are obtained by the iterative algorithm described in section 5; they are displayed in the middle panel of 8, whereas their logarithm is shown in the last graph.

7 Conclusions

The paper has revisited the problem of disaggregating economic flows by dynamic regression models using a state space framework.

It has discussed the state space formulation of the traditional disaggregation methods, with special attention to the initialisation issue. The latter is crucial for a correct implementation of the methods and for their nesting within more general dynamic specifications.

The associated filtering and smoothing algorithm, suitably modified to allow for the treatment of missing values and fixed and diffuse regression effects, provide the unifying tools for the statistical treatment: likelihood evaluation, diagnostic checking and ultimately disaggregation.

The empirical findings based on real life case studies and Monte Carlo experimentation can be summarised as follows:

- The use of the marginal likelihood has been found to be beneficial for the Chow-Lin model when the AR parameter is close to unity.
- Likelihood inference on the AR coefficient of the Litterman model proves to be very unreliable.
- Nesting the traditional disaggregation models within more general dynamic specifications, such as the ADL in levels and first differences, is a good modelling strategy.

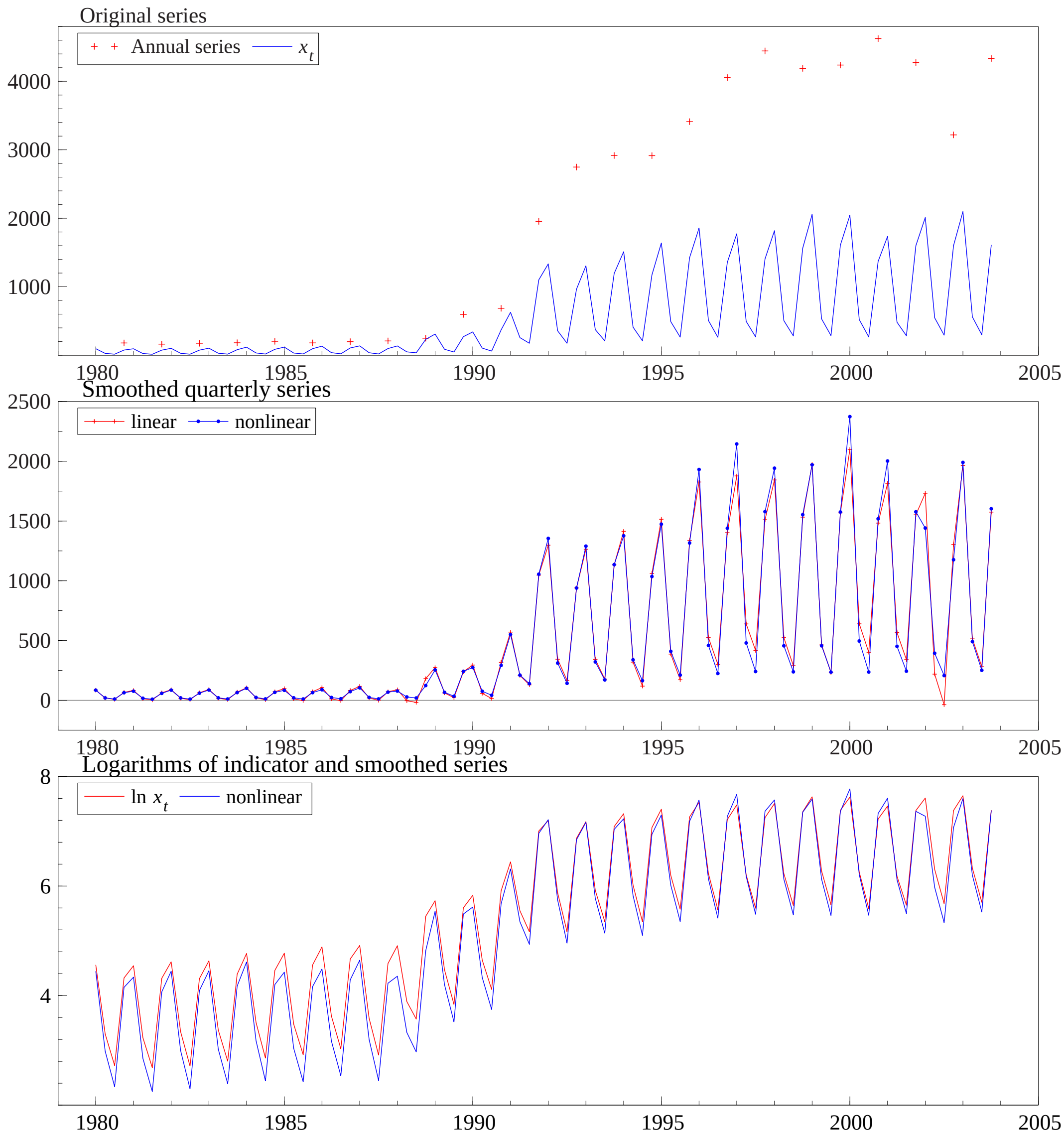

Figure 8: Tax Revenues Methane Gas (Istat A1 series). Comparison between linear and non-linear Chow-Lin disaggregation. The linear quarterly series yields negative unadmissible estimates.

- The exact nonlinear disaggregation problem arising when the disaggregated model is formulated in the logarithms of the series is feasible and computationally attractive, requiring only the iteration of routine linear smoothing operations.

Given the previous discussion on the properties of the profile and marginal likelihood and on the difficulty in estimating the Litterman model, and considering the nonstationary nature of the economic time series usually entertained in practice, the suggested strategy is to fit the ADL(1,1) model, which, under a reparameterisation and suitable initial conditions, nests both the Chow-Lin and the Fernández model, thereby incorporating our uncertainty about the presence of cointegration between the aggregated series and the indicators.

Acknowledgements

This work was carried out for the Commission on temporal disaggregation set up by Istat. I greatly benefited from discussion with Filippo Moauro. I also thank Giulia Genovino and Elisa Stoffino for assistance.

References

Anderson, B.D.O., and Moore, J.B. (1979). *Optimal Filtering*, Englewood Cliffs NJ: Prentice-Hall.

Banerjee, A., Dolado, J., Galbraith, J.W., and Hendry, D.F. (1993). *Co-integration, error-correction, and the econometric analysis of non-stationary data*, Advanced Texts in Econometrics, Oxford University Press, Oxford, UK.

Box, G.E.P., and Cox, D.R. (1964). An analysis of transformations (with discussion), *Journal of the Royal Statistical Society*, B 26, 211-246.

Chow, G., and Lin, A. L. (1971). Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series, *The Review of Economics and Statistics*, 53, 4, pp. 372-375.

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

de Jong, P. (1989). Smoothing and interpolation with the state space model, *Journal of the American Statistical Association*, 84, 1085-1088.

de Jong, P. (1991). The diffuse Kalman filter, *Annals of Statistics*, 19, 1073-1083.

de Jong, P., and Chu-Chun-Lin, S. (1994). Fast Likelihood Evaluation and Prediction for Non-stationary State Space Models, *Biometrika*, 81, 133-142.

Doornik, J.A. (2001). *Ox 3.0 - An Object-Oriented Matrix Programming Language*, Timberlake Consultants Ltd: London.

Durbin, J., and Koopman, S.J. (1992). Filtering, smoothing and estimation for time series models when the observations come from exponential family distributions. *Mimeo*, Department of Statistics, London School of Economics.

Durbin J., and Koopman, S.J. (2001). *Time Series Analysis by State Space Methods*, Oxford University Press: New York.

Engle, R.F., and Granger, C.W.J. (1987). Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55, 251-276.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalised linear models. *Journal of the American Statistical Association* 97, 501-509.

Fernández, P. E. B. (1981). A methodological note on the estimation of time series, *The Review of Economics and Statistics*, 63, 3, pp. 471-478.

Gill, P.E., Murray, W., Saunders, M.A., and Wright, M.H. (1989). Constrained nonlinear programming, in G.L. Nemhauser, A.H.G. Rinnooy Kan, M.J. Todd (eds.), *Handbooks in Operations Research and Management Science*, Vol. 1, Optimization, pp. 171-210, Elsevier, Amsterdam, 1989.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press: Cambridge.

Harvey, A.C. and Chung, C.H. (2000) Estimating the underlying change in unemployment in the UK, *Journal of the Royal Statistics Society, Series A*, 163, 303-339.

Harvey, A.C., and Pierse R.G. (1984). Estimating Missing Observations in Economic Time Series. *Journal of the American Statistical Association*, 79, 125-131.

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.

Koopman, S.J. (1997). Exact initial Kalman filtering and smoothing for non-stationary time series models, *Journal of the American Statistical Association*, 92, 1630-1638.

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

Moauro F. and Savio G. (2002). Temporal Disaggregation Using Multivariate Structural Time Series Models. Mimeo.

Palm, F.C. and Nijman, T.E. (1984). Missing observations in the dynamic regression model, *Econometrica*, 52, 1415-1436.

Proietti T. (1998). Distribution and interpolation revisited: a structural approach, *Statistica*, 58, 411- 432.

Proietti, T. (2004). On the estimation of nonlinearly aggregated mixed models. Working paper, Department of Statistics, University of Udine.

Proietti T. and Moauro F. (2003). Dynamic Factor Analysis with Nonlinear Temporal Aggregation Constraints. Mimeo.

Rosenberg, B. (1973). Random coefficient models: the analysis of a cross-section of time series by stochastically convergent parameter regression, *Annals of Economic and Social Measurement*, 2, 399-428.

Severini, Thomas A. (2000). *Likelihood methods in statistics*, Oxford statistical science series, 22, Oxford University Press.

Shephard, N.G., and Harvey, A.C. (1990). On the probability of estimating a deterministic component in the local level model, *Journal of Time Series Analysis*, 11, 339-347.

Stock, J.H., and Watson M.W. (1991). A probability model of the coincident economic indicators. In *Leading Economic Indicators*, Lahiri K, Moore GH (eds); Cambridge University Press: New York.

Tunnicliffe-Wilson, G. (1989). On the use of marginal likelihood in time series model estimation, *Journal of the Royal Statistical Society, Series B*, 51, 15-27.

A Description of the Ox Programme

The Ox programme `TDSsm.ox` contains functions for the statistical treatment of disaggregation using dynamic regression models.

The models implemented are identified by an appropriate string, assigned to the variable `sSel`, that result from the collation of three components: the Model type, the Deterministic type and the Diffuse type.

The options supported for the Model type are listed below:

Deterministic components may enter the model via the matrix of the regressors (see below) or are automatically included in the transition equation by appending a `c` to the model string for a constant term is or a `t` for a linear trend.

For instance, `sSel = "CLt"` selects the Chow-Lin model with a linear trend; `sSel = "ADL11Dc"` selects the ADL(1,1) model in first differences with a constant term, $\Delta y_t = \phi \Delta y_{t-1} + m + \epsilon_t$.

Notice that `"L0"` can only be used in its original specification, so that if it desired to include a deterministic trend to the Litterman original specification, this has to be done by including $[1, t]$ in the set of regressors x_t.

If further the elements of β are diffuse a `d` is appended to the model string. For instance, `sSel = "Ltd"` identifies the Litterman model with a linear trend and diffuse regression effects.

The series under investigation, y_t^c, is subject to missing values; for instance in the quarterly case it features the following elements:

$$\{\texttt{"."}, \texttt{"."}, \texttt{"."}, y_4^c, \texttt{"."}, \texttt{"."}, \texttt{"."}, y_8^c, \ \texttt{"."}, \ \texttt{"."}, \ \texttt{"."}, y_{12}^c, \ldots\}$$

where "." is a missing value and $y_{\tau s}^c$ is the annual total for the τ-th year.

Along with the string selecting a model, the key ingredients are listed below:

Ox Variable	Symbol
`cn`	n
`cs`	s
`vyf`	$\{y_t^c, t = 1, \ldots, n\}$ ($1 \times n$ vector)
`dphi`	ϕ
`mx`	$[x_1, x_2, \ldots, x_n]$ ($k \times n$ matrix)

When no indicator variables are available `mx =<>` (an empty matrix). Notice also that if the `mx` matrix has $x_t = [1, t, x_t^{\dagger'}]'$ as its column elements, selecting `sSel = "CL"` yields the same results as `sSel = "CLt"` using only $x_t^{\dagger'}$ as regressors. As stressed in section 4.1, a difference arise under diffuse effects, that is `sSel = "CLd"` using $x_t' = [1, t, x_t^{\dagger'}]$ is no longer the same as `sSel = "CLtd"` with $x_t^{\dagger'}$ as regressors, the first being preferable when it is suspected that ϕ is close to unity.

A.1 Linear disaggregation

The linear disaggregation methods described in the paper are implemented by the following set of Ox functions the are described below. Each function declaration lists in parenthesis its arguments and is followed by a brief comment.

IndicatorVariable(const cn, const cs) Generates a row vector of length n with elements $\psi_t, t = 1, \ldots, n$.

Cumulator(const mY, const cs) Generates the row vector with the cumulator values $y_t^c = \psi_t y_{t-1}^c + y_t$ from the series y_t in argument as row vector.

DeCumulator(const mCY, const cs) Generates the row vector of the decumulated sequence y_t from y_t^c.

SelectModel(const sSel) Set the global model selection parameters from the identifying string.

SetStateSpaceForm(const vyf, const dphi, const mx, const cs, const sSel) Builds the system matrices z^*, T^*, H^* and the elements of a_1, W_1^*, H_1^*, of the state space representation for the model identified by `sSel` and returns the matrix $[W_2, W_3, \ldots, W_{n+1}]$

SsfLogLikc(const vyf, const dphi, const mx, const cs, const sSel) Evaluates the profile likelihood $\mathcal{L}_c$ or the diffuse profile likelihood, $\mathcal{L}_\infty$ at ϕ, given respectively by (14) and (15), by means of the augmented Kalman filter (equations (11)).

SsfProfileLikelihood(const vyf, const mx, const cs, const sSel) Uses the previous function for evaluating the profile likelihood over the interval (-0.99,0.99) and plots it versus ϕ. The horizontal line is drawn at the maximum minus one half of the 95-th percentile of the chi square distribution with one degree of freedom. Values of ϕ in the region where the likelihood is above the line do not differ significantly from the maximiser at 5% level.

GridSearch(const vyf, const mx, const cs, const sSel, ...) Performs estimation of the ϕ parameter via a grid search over the interval (-1, 1). The user may modify the range of the search specifying a different lower bound or a different range.

SsfInnovations(const vyf, const dphi, const mx, const cs, const sSel) Runs the augmented Kalman filter (equations (11)) and computes the innovations $\tilde{v}_t$, along with their variance $\tilde{f}_t$, as described in section 4.2.

SsfFilteredEst(const vyf, const dphi, const mx, const cs, const sSel) Computes the real time or filtered estimates for the state space model augmented by appending y_t to the state vector. The last element of the filtered state is the real time estimate of the disaggregated series. Only the estimation error variance for this element is returned.

SsfSmoothedEst(const vyf, const dphi, const mx, const cs, const sSel) Computes the smoothed estimates for the state space model augmented by appending y_t to the state vector. The last element of the smoothed state vector provides the estimate of the disaggregated series. Only the estimation error variance for this element is returned.

The function implements the smoothing algorithm proposed by de Jong (1988), appropriately adapted to hand missing values, discussed in section 4.2.

LinearDisaggregation(const vyf, const mx, const cs, const sSel) This function performs maximum likelihood estimation, plotting the profile likelihood, computes the real time and smoothed estimates of the disaggregated series, and plots them along with 95% confidence bounds, computes the standardised innovations and plots them along with their correlogram and nonparametric density estimate.

A.2 Nonlinear disaggregation

The functions implementing the disaggregation with nonlinear temporal aggregation constraint arising when the model is specified in the logrithms, are described below. Here, `vYf` denotes the $1 \times n$ vector

$$\{\texttt{"."},\texttt{"."},\texttt{"."},Y_4^c,\texttt{"."},\texttt{"."},\texttt{"."},Y_8^c,\ \texttt{"."},\ \texttt{"."},\ \texttt{"."},Y_{12}^c,\ldots\}.$$

`LGAMLogLikc(const vYf, const dphi, const mx, const cs, const sSel, const vyhat)` Evaluates the profile likelihood (under fixed and diffuse effects) for the linear and Gaussian approximating model (19) using $\{\tilde{y}_t\}$ for the Taylor expansion.

`LGAMSmoothedEst(const vYf, const dphi, const mx, const cs, const sSel, const vyhat)` Computes the smoothed estimates of the disaggregate series for the linear and Gaussian approximating model (19) based on $\{\tilde{y}_t\}$.

`SequentialPostMode(const vYf, const dphi, const mx, const cs, const sSel, const vyhat)` Starting from a trial disaggregated series $\{\tilde{y}_t\}$, computes the final feasible estimate of the disaggregated series iterating until convergence the constrained linear sequential algorithm described in section 5.

`LGAMGridSearch(const vYf, const mx, const cs, const sSel, ...)` Performs estimation of the ϕ parameter via a grid search over the interval (-1, 1). The user may modify the range of the search specifying a different lower bound or a different range.

`SsfNLProfileLikelihood(const vYf, const mx, const cs, const sSel)` Evaluates the profile likelihood over the interval (-0.99,0.99) and plots it versus ϕ.

Table 1: Revision history for Istat series B2 (years 1992-2003).

Model	*Mean percentage revision error*			
	1 step	2 steps	3 steps	4 steps
Chow-Lin (constant)	1.44	1.97	2.22	2.23
Chow-Lin (trend)	1.36	1.68	1.85	1.73
Fernandez	0.67	0.83	0.94	0.88
	Mean revision error			
	1 step	2 steps	3 steps	4 steps
Chow-Lin (constant)	259.67	372.11	435.34	443.47
Chow-Lin (trend)	250.70	324.80	372.80	353.21
Fernandez	111.75	147.85	177.93	173.44
	Mean absolute revision error			
	1 step	2 steps	3 steps	4 steps
Chow-Lin (constant)	404.13	574.00	673.56	693.95
Chow-Lin (trend)	414.32	560.54	649.26	647.21
Fernandez	366.79	525.71	632.25	673.81
	Mean square revision error			
	1 step	2 steps	3 steps	4 steps
Chow-Lin (constant)	247363	503008	695426	743491
Chow-Lin (trend)	336728	589829	793115	742075
Fernandez	215468	422454	603165	672839

Acronym	Model type	Notes
"CL"	Chow-Lin	
"ADL10"	ADL(1,0) in levels	Initialisation based on (7)
"ADL10x"	ADL(1,0) in levels	Initialisation: $y_1 \sim \mathrm{N}(c + x_1'\beta, \sigma^2)$
"ADL11"	ADL(1,0) in levels	Initialisation based on (7)
"ADL11x"	ADL(1,1) in levels	Initialisation: $y_1 \sim \mathrm{N}(c + x_1'\beta, \sigma^2)$
"L0"	Litterman	$u_0 = \Delta u_0 = 0$ (zero initialisation)
"L"	Litterman	u_1 in β, $\Delta u_0 \sim \mathrm{N}(0, \sigma^2/(1-\phi^2))$
"ADL10D"	ADL(1,0) in 1st differences	
"ADL11D"	ADL(1,1) in 1st differences	