On the Estimation of Nonlinearly Aggregated Mixed Models

Tommaso Proietti

Statistics Department,

University of Udine, Italy

Abstract

The article proposes an iterative algorithm for the estimation of fixed and random effects of a nonlinearly aggregated mixed model. The latter arises when an additive Gaussian model is formulated at the disaggregate level on a nonlinear transformation of the responses, but information is available in aggregate form. The nonlinear transformation breaks the linearity of the aggregate model, yielding a nonlinear tight observational constraint.

The algorithm rests upon the sequential linearization of the nonlinear aggregation constraint around proposals that are iteratively updated until convergence. Likelihood inferences on the hyperparameters are also discussed. As a by product we provide a solution to the problem of disaggregating over the units of analysis the aggregate responses, enforcing the nonlinear observational constraints.

Illustrations are provided with reference to the temporal disaggregation problem, concerning the distribution of annual time series flows to the quarters making up the year.

Keywords: Temporal and spatial disaggregation; Best linear unbiased prediction; Box-Cox transformation; Constrained nonlinear optimization.

1 Introduction

The available statistical information often refers to space or time units that are wider than the units of analysis. Using aggregate data, we are typically interested in estimating a model that is specified at the disaggregate level; as a related problem, we are also concerned with distributing the available information over the units of analysis that make up the aggregate (disaggregation).

Linear aggregation, that arises when the aggregate is linear in the unknown disaggregate responses, has received a lot of attention in the literature and the corresponding disaggregation problem has a closed form solution. This article is concerned instead with a situation when the disaggregated model is a linear mixed model formulated in terms of a transformation of the response, e.g. the Box-Cox transformation (Box and Cox, 1964), and the aggregated value is a nonlinear function of the transformed disaggregated responses.

A leading example is provided by the distribution of of annual time series totals of a flow variable to the quarters, using a linear mixed model formulated for the logarithms of the original variables, rather than the levels. We specify a linear time series model for the logarithms of a variable, as we deem that the assumptions of additivity, normality and homoscedasticity are more likely to hold on the transformed scale, rather than the levels. As the annual aggregate results from the sum of the levels of the quarters making up the year, a nonlinear observational constraint arises.

This article proposes an iterative algorithm for estimating the fixed and random effects of the disaggregated mixed model, and that solves the nonlinear disaggregation problem. The algorithm is based on a Taylor first order approximation of the nonlinear observational constraint around a trial value that is sequentially improved; one of its virtues is that it can be implemented using standard linear estimating equations.

Although in our applications we refer to temporal disaggregation, the solution is applicable to spatial disaggregation using intrinsic random functions (see

Cressie, 1993, sec. 5.4) and to the estimation of contingency tables with known margins, which can also be though of as a particular instance of disaggregation.

Section 2 introduces the problem of nonlinear aggregation of mixed models and briefly reviews the most popular temporal disaggregation procedures. The estimation of fixed and random effects is performed by our proposed iterative algorithm, which is presented in section 3. The algorithm is a particular instance of a sequential linear constrained method for solving an optimization problem with nonlinear constraints (see Gill *et al.*, 1989) and its properties are illustrated using geometric arguments. We also discuss how likelihood inference on the hyperparameters is carried out.

Section 5 is devoted to two empirical illustration concerning the temporal disaggregation of the total production series from the annual frequency to the quarterly frequency, using related indicators. Finally, in section 6 we draw our conclusions.

2 Nonlinear aggregation

Suppose that n disaggregated responses, $\mathbf{y}$, follow a mixed linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are known matrices, $\boldsymbol{\beta}$ is a vector of k fixed unknown parameters, $\boldsymbol{\alpha}$ is a vector of random effects, $\boldsymbol{\alpha} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of residuals, $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2_\epsilon \mathbf{I}_n)$, that are distributed independently of $\boldsymbol{\alpha}$.

The vector $\mathbf{y}$ is not observed, but a nonlinear non-injective (many-to-one) transformation is available, $\mathbf{Y} = \mathbf{f}(\mathbf{y})$, where $\mathbf{f}(\cdot)$ is a $N \times 1$, $N < n$, vector function of $\mathbf{y}$, and $\mathbf{Y}$ denotes the $N \times 1$ vector stacking the aggregated transformed observations, $\mathbf{Y} = \{Y_i, i = 1, \ldots, N\}$.

This situation arises when a linear Gaussian mixed model is assumed to hold for a scale that is different from the original scale of measurement, e.g. on the logarithms of the variable, but the aggregation is linear in the original measurements, which induces a nonlinear aggregation constraints in terms of the elements of the vector $\mathbf{y}$.

For instance, in the temporal disaggregation of flow variables or time averaged stocks measured on a ratio scale (such as production, income and prices), a typical situation is when a linear time series model is formulated for the logarithms of the quarterly values, $\mathbf{y}$, and the available data are only annual and arise from the sum of the levels of the flow variable over the four quarters that make up the year.

In such cases, if s denotes the aggregation interval, $s = 4$ in our example, the observations can be expressed in terms of the y's as follows:

$$Y_i = \sum_{j=0}^{s-1} f(y_{is-j}), i = 1, \ldots, N; \tag{2}$$

we shall mostly concentrate on $f(y) = \exp(y)$, but the theory applies to the general class of Box-Cox inverse transformation with parameter λ, $f(y) = (1 + \lambda y)^{1/\lambda}$. Throughout the paper we assume that the transformation is smooth in that the function $f(\cdot)$ is twice continuously differentiable.

The problem of temporal disaggregation of flow variables using related indicators has received a lot of attention in the literature and has practical relevance: as a matter of fact, in many countries disaggregation techniques are an essential ingredient for the construction of quarterly national accounts estimates from annual data ($\mathbf{Y}$) and quarterly related indicators, $\mathbf{X}$. These techniques rest upon the linearity assumption, by which $\mathbf{Y} = \mathbf{A}\mathbf{y}$, where the matrix $\mathbf{A}$ is a constant aggregation matrix. Usually, the observations $\mathbf{Y}$ pertain to the sum of s consecutive disaggregated values, so that $\mathbf{A} = \mathbf{I}_N \otimes \mathbf{i}'_s$, $\mathbf{i}'_s = [1, \ldots, 1]$.

The most popular disaggregation techniques postulate a simple time series process for the random component. For instance, in the Chow-Lin (1981) linear disaggregation procedure the disaggregated model is a linear regression with first order autoregressive errors, $y_t = \mathbf{x}'_t\boldsymbol{\beta} + \alpha_t$, $\alpha_t = \phi\alpha_{t-1} + \eta_t$, $\eta \sim \mathrm{N}(0, \sigma^2)$. In the representation (1), $\mathbf{Z} = \mathbf{I}$ and $\boldsymbol{\Omega}$ has elements $\omega_{ij} = \phi^{|i-j|}\sigma^2/(1 - \phi^2)$, and $\boldsymbol{\epsilon} = \mathbf{0}$.

Litterman (1983) proposed a linear disaggregation procedure based upon the disaggregate model $y_t = \mathbf{x}_t'\boldsymbol{\beta} + \alpha_t$, where α_t is an ARIMA(1,1,0) process: $\alpha_t = \alpha_{t-1} + \phi(\alpha_{t-1} - \alpha_{t-2}) + \eta_t$. Denoting by $\boldsymbol{\Delta}_\rho$ the $n \times n$ (quasi) differencing matrix with ones on the diagonal and $-\rho$ on the first subdiagonal, i.e.

$$\boldsymbol{\Delta}_\rho = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

the Litterman model admits the representation (1) with $\mathbf{Z} = \mathbf{I}$, $\boldsymbol{\Omega} = \sigma^2 \left(\boldsymbol{\Delta}_1'\boldsymbol{\Delta}_\phi'\boldsymbol{\Delta}_\phi\boldsymbol{\Delta}_1\right)^{-1}$ and $\boldsymbol{\epsilon} = \mathbf{0}$.

The Fernandez (1981) model is such that α_t is a random walk, and thus can be seen as a restricted version of the Litterman model featuring $\phi = 0$. The case when α_t is an ARIMA process has been considered by Wei and Stram (1990), and in general, (1) can be viewed as the stacked version of the the general linear state space model:

$$\begin{aligned} y_t &= \mathbf{z}'\boldsymbol{\alpha}_t + \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t, \qquad \epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2) \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}\boldsymbol{\alpha}_t + \mathbf{c}_t + \mathbf{R}\boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \text{NID}(\mathbf{0}, \mathbf{Q}), \quad \text{E}(\boldsymbol{\eta}_t\epsilon_j) = \mathbf{0}, \forall j. \end{aligned}$$

3 The iterative algorithm

Let $\mathbf{A}(\mathbf{y}) = \{a_{it}(\mathbf{y})\}$ denote the $N \times n$ Jacobian matrix, containing the partial derivatives $a_{it}(\mathbf{y}) = \partial f_i / \partial y_t$; for instance if in (2) $f(\cdot) = \exp(\cdot)$, $a_{it}(\mathbf{y}) = \exp(y_t), t = is - j, j = 0, \ldots, s-1$, and $a_{it}(\mathbf{y}) = 0$ otherwise.

Let us denote by $\mathbf{y}^*$ a trial value and set $\mathbf{A}^* = \mathbf{A}(\mathbf{y}^*)$. Writing $\mathbf{Y}^* = \mathbf{f}(\mathbf{y}^*)$, the first order Taylor approximation of $\mathbf{Y} = \mathbf{f}(\mathbf{y})$ around the trial value $\mathbf{y}^*$ is:

$$\mathbf{Y} \approx \mathbf{Y}^* + \mathbf{A}^*(\mathbf{y} - \mathbf{y}^*). \tag{3}$$

Replacing the mixed model representation for $\mathbf{y}$ into (3), we obtain the pseudo linear aggregated model:

$$\tilde{\mathbf{Y}}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{Z}^*\boldsymbol{\alpha} + \boldsymbol{\epsilon}^*, \tag{4}$$

where $\tilde{\mathbf{Y}}^* = \mathbf{A}^*\mathbf{y}^* + \mathbf{Y} - \mathbf{Y}^*,\ \mathbf{X}^* = \mathbf{A}^*\mathbf{X},\ \mathbf{Z}^* = \mathbf{A}^*\mathbf{Z}$, and $\boldsymbol{\epsilon}^* = \mathbf{A}^*\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma_\epsilon^2\mathbf{A}^*\mathbf{A}^{*\prime})$.

Letting $\boldsymbol{\Sigma} = \sigma_\epsilon^2\mathbf{I}_T + \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}'$, and applying standard optimal prediction principles (see Robinson, 1991), the estimators of the fixed and random effects are respectively:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left[\mathbf{X}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}\mathbf{X}^*\right]^{-1}\mathbf{X}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}\tilde{\mathbf{Y}}^*, & (5)\\
\hat{\boldsymbol{\alpha}} &= \boldsymbol{\Omega}\mathbf{Z}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}\left(\tilde{\mathbf{Y}}^* - \mathbf{X}^*\hat{\boldsymbol{\beta}}\right), & (6)\\
\hat{\boldsymbol{\epsilon}} &= \sigma_\epsilon^2\mathbf{A}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}\left(\tilde{\mathbf{Y}}^* - \mathbf{X}^*\hat{\boldsymbol{\beta}}\right), & (7)
\end{aligned}$$

These inference can be combined so as to construct a new trial value

$$\begin{aligned}
\hat{\mathbf{y}}^* &= \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\epsilon}}\\
&= \mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}\mathbf{A}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}\left(\tilde{\mathbf{Y}}^* - \mathbf{X}^*\hat{\boldsymbol{\beta}}\right).
\end{aligned} \tag{8}$$

The latter can be used to form a new linear approximating model via a Taylor first order approximation.

The previous arguments suggest the following iterative scheme:

1. Start from a trial value $\mathbf{y}^*$. A possibility is to solve the linear disaggregation problem assuming $\mathbf{Y} = \mathbf{A}\mathbf{y}$, for a fixed aggregation matrix $\mathbf{A}$; in this case the trial value is said to be feasible, as it satisfies the constraint $\mathbf{f}(\mathbf{y}^*) = \mathbf{Y}$. In general, $\mathbf{y}^*$ does not have to be feasible.

2. Form the linear approximating model using the first order Taylor expansion around $\mathbf{y}^*$.

3. Estimate the fixed and random effects and $\boldsymbol{\epsilon}$ from the linearized model using (5)-(7), and combine them to form $\hat{\mathbf{y}}^*$ as in (8).

4. If $||\mathbf{y}^* - \hat{\mathbf{y}}^*||$, or equivalently $||\mathbf{Y} - \mathbf{f}(\hat{\mathbf{y}}^*)||$, is greater than a specified tolerance value, set $\mathbf{y}^* = \hat{\mathbf{y}}^*$ and return to step 2.

4 The nature of the solution

The iterative algorithm outlined in the previous section is a *sequential linear constrained* (SLC) method for solving a constrained nonlinear optimization problem. Denoting by $g(\cdot)$ a Gaussian density, the problem consists of choosing $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\epsilon}}$ so as to maximize the joint density $g(\mathbf{y}, \boldsymbol{\alpha})$ subject to the observational constraints: $\mathbf{Y} = \mathbf{f}(\mathbf{y})$, that is:

$$\max_{\beta,\alpha,\epsilon} \{\ln g(\mathbf{y}|\boldsymbol{\alpha}) + \ln g(\boldsymbol{\alpha})\} \quad \text{subject to:} \quad \mathbf{Y} = \mathbf{f}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}).$$

SLC methods, reviewed in Gill *et al.* (1989), section 7, rests upon the linearization of the constraint around a trial value $\mathbf{y}^*$, which does not have to be a feasible value. This yields the optimization problem with linear constraints:

$$\min_{\beta,\alpha,\epsilon} \left\{\sigma_\epsilon^{-2}\boldsymbol{\epsilon}'\boldsymbol{\epsilon} + \boldsymbol{\alpha}\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}\right\} \quad \text{subject to:} \quad \tilde{\mathbf{Y}}^* = \mathbf{A}^*\mathbf{X}\boldsymbol{\beta} + \mathbf{A}^*\mathbf{Z}\boldsymbol{\alpha} + \mathbf{A}^*\boldsymbol{\epsilon},$$

for which an exact solution is available. The latter is obtained in two stages: for a given $\boldsymbol{\beta}$, the solution for $\boldsymbol{\alpha}$ and $\boldsymbol{\epsilon}$ is given as in (6)-(7). Replacing into the objective function yields the solution for $\hat{\boldsymbol{\beta}}$ as given in (5). The new value is then obtained by a linear combination of these estimates and the process is iterated until convergence.

At convergence, $\mathbf{Y} = \mathbf{f}(\hat{\mathbf{y}}^*)$, so that $\tilde{\mathbf{Y}}^* = \mathbf{A}^*\hat{\mathbf{y}}^*$, and the log-likelihood of the linearized model, concentrated with respect to $\boldsymbol{\beta}$, is

$$\mathcal{L}(\mathbf{Y}; \boldsymbol{\Sigma}) = -\frac{1}{2}\left\{\ln|\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*'}| + (\tilde{\mathbf{Y}}^* - \mathbf{X}^*\hat{\boldsymbol{\beta}})'(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*'})^{-1}(\tilde{\mathbf{Y}}^* - \mathbf{X}^*\hat{\boldsymbol{\beta}})\right\}. \quad (9)$$

The restricted log-likelihood (Patterson and Thompson, 1971, Harville, 1977) is defined as follows:

$$\mathcal{L}^{[R]}(\mathbf{Y}; \boldsymbol{\Sigma}) = \mathcal{L}(\mathbf{Y}; \boldsymbol{\Sigma}) - \frac{1}{2}\ln|\mathbf{X}^{*'}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*'})^{-1}\mathbf{X}^*|.$$

The solution $\hat{\mathbf{y}}^*$ provides the mode of the distribution of $\mathbf{y}$ conditional on $\mathbf{Y}$ and the aggregation constraint. Thus, if we apply the inverse transformation, e.g. if $f(\cdot) = \exp(\cdot)$, we exponentiate the elements of $\hat{\mathbf{y}}^*$, a set of disaggregated estimates of the unknown observations, that are consistent with the aggregated totals, are obtained.

It should be noticed that in the linear case $\mathbf{Y} = \mathbf{A}\mathbf{y}$, starting from any trial value $\mathbf{y}^*$, the algorithm converges at the first iteration. This is a reflection of the fact that the linear disaggregation problem admits a closed form solution.

A similar iterative algorithm arises in the estimation of the nonlinear mixed model (NLMM):

$$\begin{aligned} \mathbf{Y} &= \mathbf{f}(\mathbf{y}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}), \\ \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}, \quad \boldsymbol{\alpha} \sim \mathrm{N}(\mathbf{0}, \Omega) \end{aligned}$$

This model is considered in Lindstrom and Bates (1990), see also Pinheiro and Bates (2000, ch. 7) and the references therein, and addresses a different situation, in which the observable $\mathbf{Y}$ is nonlinearly related to a signal, composed of fixed and random effects, and is affected by measurement error. As a matter of fact, the mapping $\mathbf{y} \mapsto \mathbf{f}(\mathbf{y})$ is one to one and the Jacobian is a square diagonal matrix. In our perspective, it is a different model since the observational constraint is not binding and the measurement error is absent from the disaggregated mixed model.

Inference for the NLMM is carried out iteratively via a linearization of $\mathbf{f}(\mathbf{y})$ around proposals $\mathbf{y}^*$ that are sequentially updated. The values of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ that maximize $\ln g(\mathbf{Y}, \boldsymbol{\alpha}) = \ln g(\mathbf{Y}|\boldsymbol{\alpha}) + \ln g(\boldsymbol{\alpha})$ are formally given as in (5)-(6), but the updated $\hat{\mathbf{y}}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\boldsymbol{\alpha}}$ (and thus $\tilde{\mathbf{Y}}^*$) obviously differs.

Moreover, the evaluation of the likelihood poses different issues: for the NLMM it is needed to compute the integral $\int g(\mathbf{Y}|\boldsymbol{\alpha}) g(\boldsymbol{\alpha}) d\boldsymbol{\alpha}$, which can be done by Monte Carlo simulation methods using importance sampling techniques and other approximating methods reviewed in Pinheiro and Bates (1995).

In the framework considered by this paper the only option that is available is to use the Gaussian likelihood (9) when the iterative estimation scheme has converged. Denoting $R = \{\mathbf{y} : \mathbf{Y} = \mathbf{f}(\mathbf{y})\}$, $R^* = \{\mathbf{y} : \mathbf{A}^*\mathbf{y} = \tilde{\mathbf{Y}}^*\}$, the likelihood, defined by a multiple integral over the surface R, $\int_R g(\mathbf{y}) d\mathbf{y}$, is approximated by $\int_{R*} g(\mathbf{y}) d\mathbf{y} = g(\tilde{\mathbf{Y}}^*)$. In practice, the surface R is replaced by the hyperplane tangent to the surface at the optimized $\hat{\mathbf{y}}^*$, defined by $\mathbf{A}^*(\mathbf{y} - \hat{\mathbf{y}}^*) = 0, \tilde{\mathbf{Y}}^* = \mathbf{A}^*\hat{\mathbf{y}}^*$.

Maximum likelihood estimation of the variance parameters $\boldsymbol{\Omega}$ and σ_ϵ^2 can thus be based on a quasi-Newton algorithm, which at each iteration approximates the likelihood for a given parameter configuration by the Gaussian likelihood for $\tilde{\mathbf{Y}}^*$ at convergence, given by (9).

4.1 Illustration of the algorithm

For simplicity, consider the case when regression effects are absent. Then, denoting by $\mathbf{D}^* = \mathbf{Y} - \mathbf{f}(\mathbf{y}^*)$ the discrepancy between the observed aggregate values and the transformed initial values, the new estimate of the disaggregate observations arises as follows:

$$\begin{aligned} \hat{\mathbf{y}}^* &= \mathbf{Z}\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\epsilon}} \\ &= \boldsymbol{\Sigma}\mathbf{A}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}\tilde{\mathbf{Y}}^* \\ &= \mathbf{y}^* + \left[\mathbf{I} - \boldsymbol{\Sigma}\mathbf{A}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}\mathbf{A}^*\right]\mathbf{y}^* + \boldsymbol{\Sigma}\mathbf{A}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}\mathbf{D}^* \end{aligned}$$

Setting $\mathbf{M}_1 = \left[\mathbf{I} - \boldsymbol{\Sigma}\mathbf{A}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}\mathbf{A}^*\right]$ and $\mathbf{M}_2 = \boldsymbol{\Sigma}\mathbf{A}^{*\prime}(\mathbf{A}^*\boldsymbol{\Sigma}\mathbf{A}^{*\prime})^{-1}$, we have that

$$\mathbf{A}^*\mathbf{M}_1 = \mathbf{0}, \quad \mathbf{A}^*\mathbf{M}_2 = \mathbf{I}, \;\; \mathbf{M}_1\mathbf{M}_2 = \mathbf{0}.$$

$\mathbf{M}_1$ is a projection matrix that spans the null space of $\mathbf{A}^*$; as a result $\mathbf{M}_1\mathbf{y}^*$ is a movement along the hyperplane normal to $\mathbf{A}^*$, defined by the equation $\mathbf{A}^*(\mathbf{y} - \mathbf{y}^*) = \mathbf{0}$. On the contrary, $\mathbf{M}_2$ lies in the range space of $\mathbf{A}^*$, and thus it projects a point onto the subspace generated by the rows of $\mathbf{A}^*$.

The previous decomposition shows that the new proposal results from two distinct movements: the first determines the optimal solution (BLUP) along the hyperplane that is orthogonal to $\mathbf{A}^*$ (this hyperplane is parallel to that tangent to the curve $\mathbf{Y} = \mathbf{f}(\mathbf{y})$); the second aims at reducing the distance from the curve $\mathbf{Y} = \mathbf{f}(\mathbf{y})$; $\mathbf{M}_2\mathbf{D}^*$ is thus a movement towards the nonlinear attractor $\mathbf{Y} = \mathbf{f}(\mathbf{y})$.

In the presence of known fixed effects, the previous decomposition becomes $\hat{\mathbf{y}}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{M}_1(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}) + \mathbf{M}_2\mathbf{D}^*$. In the general case, when fixed effects are estimated, a further additional component comes out in the revision of a trial estimate, which depends on the change in the estimates of $\boldsymbol{\beta}$.

Figure 1 illustrates the algorithm with respect to the simple case when $\mathbf{y}$ is two-dimensional and is drawn from a bivariate Gaussian distribution, $\mathbf{y} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = [1, 2]'$, $\boldsymbol{\Sigma} = \{\sigma_{ij}, i, j = 1, 2\}, \sigma_{11} = \sigma_{22} = 1, \sigma_{12} = 0.8$. The plotted ellipsoids are density contours corresponding to the probability levels 0.25, 0.50 and 0.75; the true value, drawn at random from this distribution, is $\mathbf{y} = [0.35, 1.75]'$, giving an aggregated value $Y = \exp(y_1) + \exp(y_2) = 7.20$.

The set of points in the plane satisfying the above nonlinear observational constraints is the solid curve labelled $Y = f(y)$. Suppose we start from a trial value $\mathbf{y}^* = [0, 4]'$, which yields a discrepancy equal to -48.40. The first five iterations of the sequential algorithm are reproduced in the following table:

Iteration	$\tilde{\mathbf{y}}^*$	Discrepancy $\mathbf{D}^*$
1	[1.87, 3.08]′	-21.03
2	[1.26, 2.29]′	-6.24
3	[0.83, 1.82]′	-1.25
4	[0.70, 1.66]′	-0.09
5	[0.68, 1.65]′	-5.e-04

The value obtained at the second iteration, $\hat{\mathbf{y}}_2 = [1.26, 2.29]'$ is obtained from the previous, $\hat{\mathbf{y}}_1 = [1.87, 3.08]'$, by performing two movements: the first is along the subspace $\mathbf{A}^*(\mathbf{y} - \hat{\mathbf{y}}_1^*) = 0$, which is a line in our two dimensional illustration, and aims at minimizing the estimation error variance along that subspace; the second is a movement towards the curve $\exp(y_1) + \exp(y_2) = 7.20$ that reduces the bias due to the violation of the observational constraints. After five iterations the $\hat{\mathbf{y}}_5 = [1.87, 3.08]'$ is already very close to the solution, which is a point along the attractor.

5 Empirical illustrations: disaggregation of economic time series

Our empirical illustrations deal with the temporal disaggregation of two economic flows referring to total annual production. The annual observations are to be distributed across the quarters using the quarterly information on related series.

Both the annual series and the indicators are made available by Istat, the Italian National Statistical Institute, which carries out routinely the disaggregation using a variant of the Chow-Lin procedure, that was briefly recalled in section 2.

The series under scrutiny are annual total production at current prices for the *Communication* sector (which accounts for 2.4% of total GDP in the year 2000), and for *Food, Beverages and Tobacco* (2.6% of GDP).

The quarterly indicator for Communication is a survey based measure of turnover, whereas for the Food, Bev. & Tob. sector it consists of the quarterly index of industrial production, inflated by the producer price index of the same sector. The series are plotted in figure 2, which illustrates a high degree of concordance between the annual series on the left, and the corresponding indicator on the right.

The Chow-Lin procedure adopted by Istat assumes that the aggregated annual observations are the sum of the unknown disaggregated observations, that are assumed to follow an AR(1) model with regression effects in their levels. Hence, the observational constraint is $Y_i = y_{4i} + y_{4i-1} + y_{4i-2} + y_{4i-3}$.

Our aim is to assess the sensitivity of the disaggregated total production series to the linearity assumption, by comparing the linear standard Chow-Lin method with the nonlinear alternative, which arises when the linear Gaussian mixed model (1) is assumed to hold with $\mathbf{y}$ representing the logarithms of the disaggregated unknown values. In such case the annual observations arise as $Y_i = \exp y_{4i} + \exp y_{4i-1} + \exp y_{4i-2} + \exp y_{4i-3}$.

This is a more consistent and realistic framework, as total production is measured on a ratio measurement scale (it cannot assume negative values) and 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 total production. 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.

In both cases the disaggregate model is formulated as follows:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 x_t + \alpha_t, \quad \alpha_t - \phi\alpha_{t-1} = \eta_t \sim \mathrm{NID}(0, \sigma^2),$$

where x_t denotes the indicator and t is time. The model entails that y_t and x_t are cointegrated (Engle and Granger, 1987), so that α_t has a stationary distribution, possibly around a quadratic deterministic trend.

The linear and nonlinear models were estimated by restricted maximum likelihood (REML). The disaggregated series (in the nonlinear case $\exp \hat{\mathbf{y}}^*$, where $\hat{\mathbf{y}}^*$ is obtained from the iterative algorithm of section 3 using the REML estimates of the hyperparameters) are displayed in the left upper panel of figures 3 and 4. The right upper panel is a plot of the profile log-likelihood for the ϕ parameter, in the range $[0, 1)$, adjusted for a vertical shift; for the *Communication* series the REML estimates of the autoregressive parameter resulted 0.58 for the linear specification and 0.77 for the nonlinear one. For *Food, Bev. & Tob.* the estimates were 0.87 and 0.76, respectively.

For the nonlinear specifications the estimated regression coefficient on the quarterly indicator (β_3) were 0.77 and 0.67, respectively for the *Communication* and the *Food, Bev. & Tob.* series. Their approximate standard error, computed for the linear Gaussian approximating model based on the Taylor expansion around the optimized $\hat{\mathbf{y}}^*$, were 0.11 and 0.10.

There lower panels of figures 3 and 4 compare the quarterly and annual growth rates of the disaggregated series. They convey the message that the linear and nonlinear specifications may entail important differences in the estimation of growth rates, the identification of their turning points, and in characterization of the sharpness of the turning points, As a matter of fact, in both of the cases considered in this section the linear estimates will tend display lower amplitude. In sum, essential business cycle features, such as the depth of the fluctuations, the location and sharpness of turning points seem to be affected by the choice of the specification.

6 Conclusions

This article has proposed an algorithm for the estimation of fixed and random effects of a disaggregate linear mixed model with nonlinear aggregation. The algorithm rests upon the sequential linearization of the nonlinear observational constraint around proposals that are iteratively updated until convergence. Likelihood inferences on the hyperparameters have also been discussed.

The proposed algorithm is easily implemented as it involves linear estimating equations, and provides a solution to the nonlinear disaggregation problem of distributing the observed aggregate values over the more refined unit of analysis.

Linear disaggregation methods have the attractive property of having a closed form solution; however, they come at odds with the need of formulating a disaggregate mixed model that features additivity of effects, normality and homogeneity of variance. Statistical models are in fact often formulated in terms of a transformation of the scale of the response variable, e.g. belonging to the class considered by Box and Cox (1964).

The examples concerning the estimation of quarterly time series from annual ones have illustrated that the linearity assumption may bear relevant implications for the measurement of business cycle features, such as the positioning of turning points and the amplitude of economic fluctuations.

References

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.

Cressie, N.A.C. (1993). *Statistics for spatial data*, Revised edition, Wiley, New York.

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

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

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.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, 320-340.

Lindstrom, M. J., and Bates, D. M. (1990). Nonlinear mixed-effects models for repeated measures data. *Biometrics*, 46, 673687.

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.

Patterson, H. D., and Thompson, R. (1971). Recovery of inter-block information when blocks sizes are unequal. *Biometrika*, 58, 545554.

Pinheiro, J.C., and Bates, D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effect model, *Journal of Computational and Graphical Statistics*, 4, pp. 12-35.

Pinheiro, J.C., and Bates, D.M. (2001). *Mixed-Effects Models in S and S-PLUS*, Springer, Statistics and Computing.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science*, 6, pp. 15-32.

Wei, W. W. S., and Stram D. O. (1990). Disaggregation of time series models, *Journal of The Royal Statistical Society*, B, 52, 3, pp. 453-467.

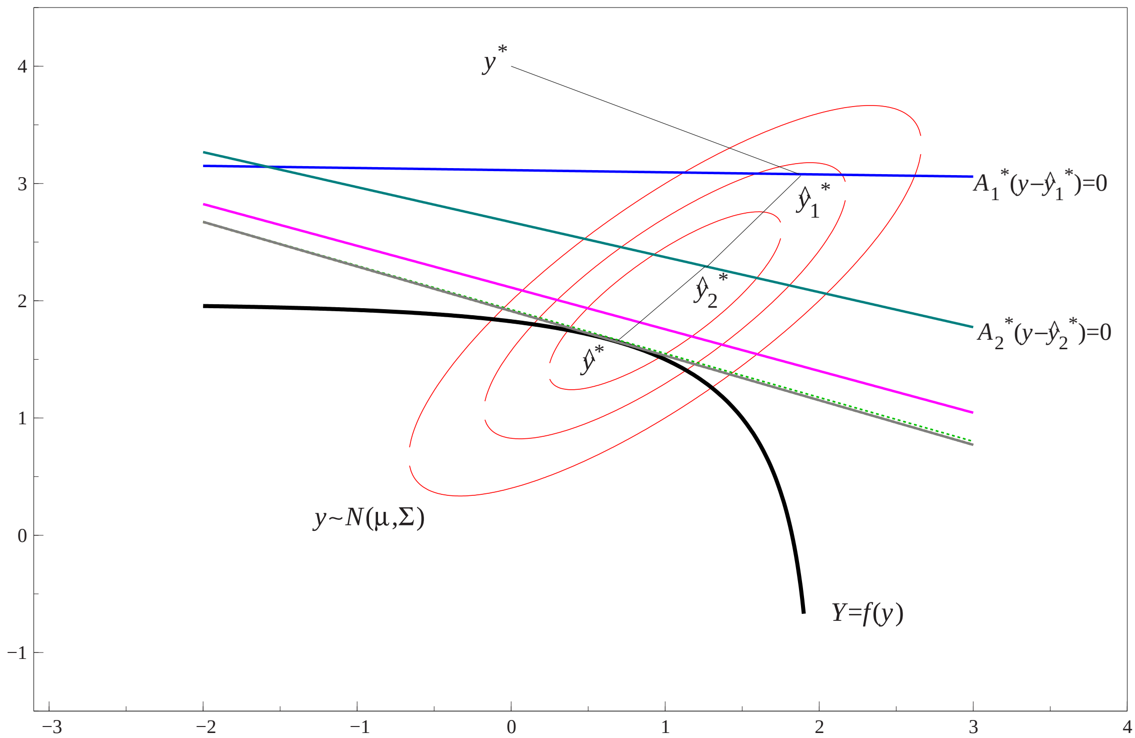

Figure 1: Illustration of the iterative algorithm.

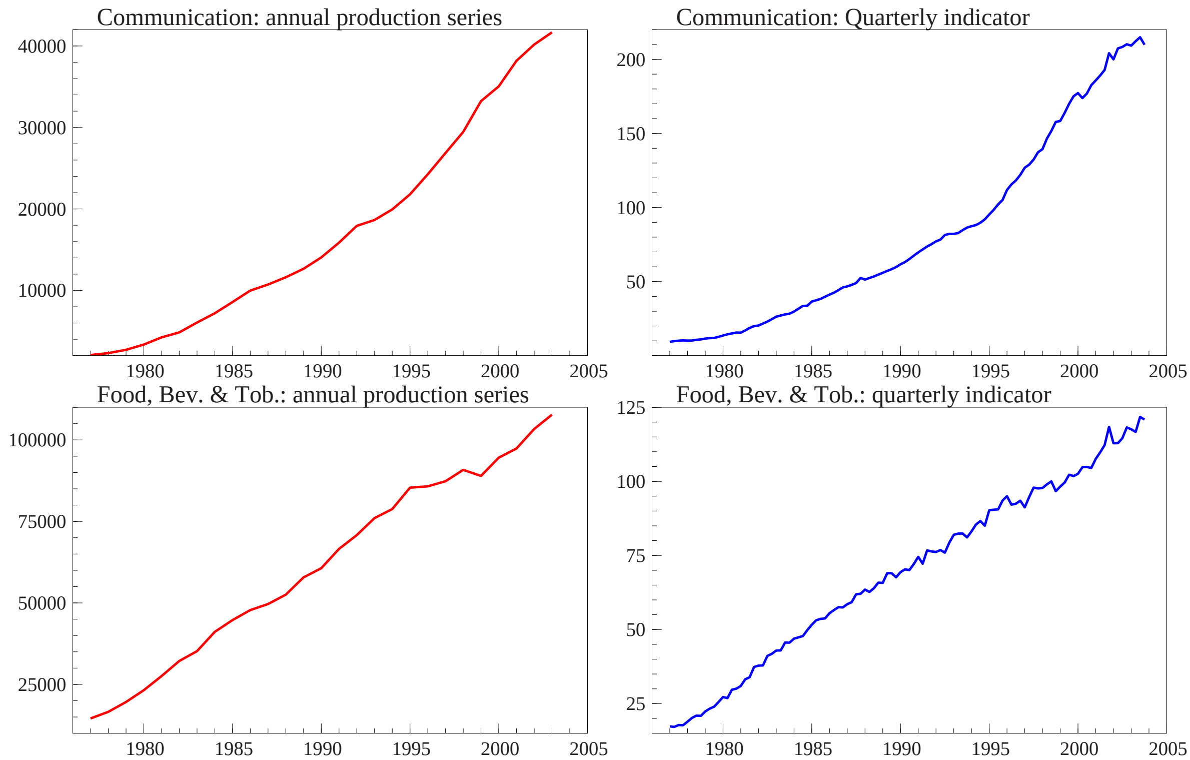

Figure 2: Plot of the annual production series for the Communication and Food, Beverages and Tobacco sectors, and the corresponding quarterly indicators.

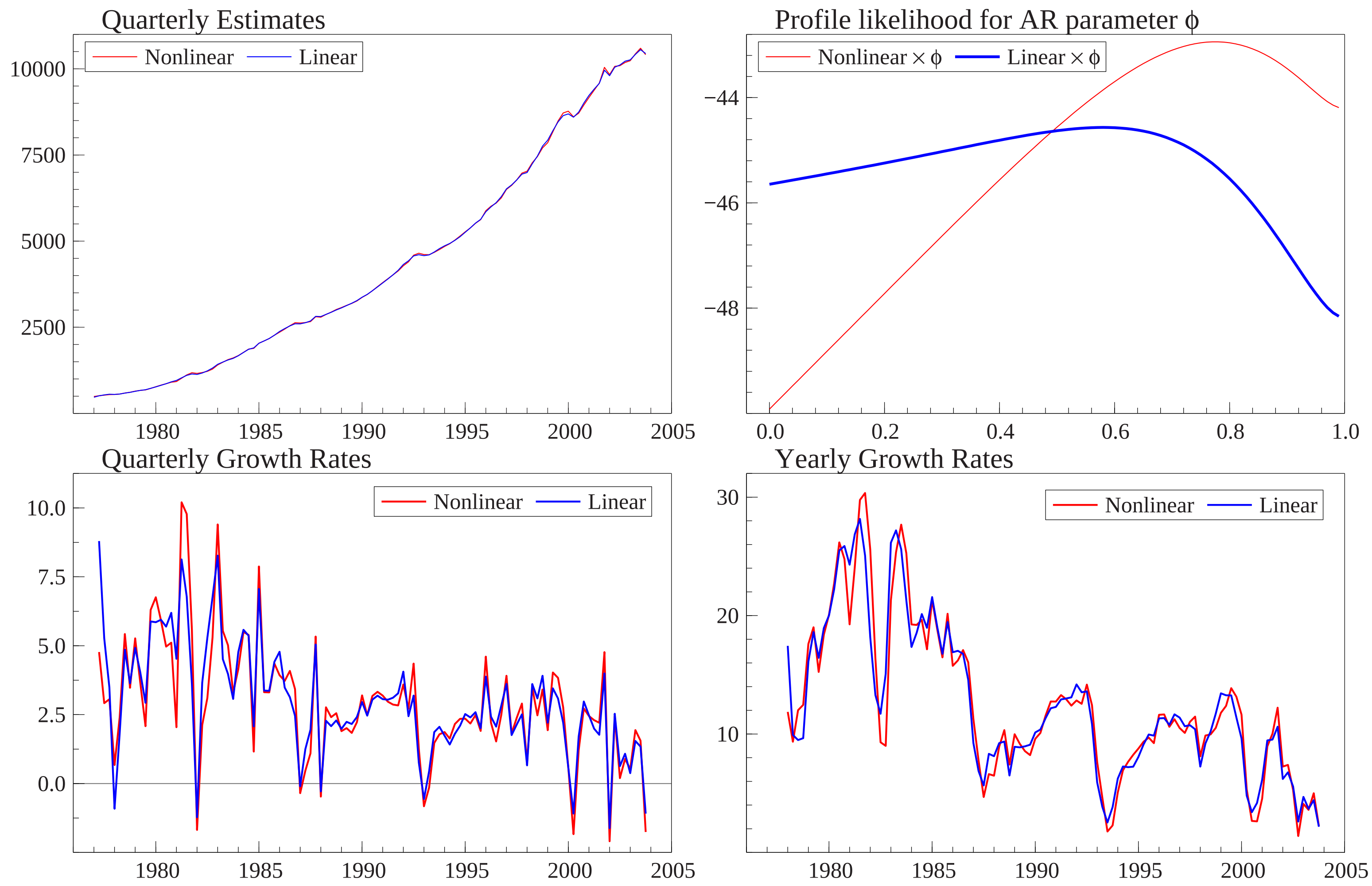

Figure 3: Temporal disaggregation of the annual production series for the Communication sector using the linear and nonlinear Chow-Lin disaggregated model.

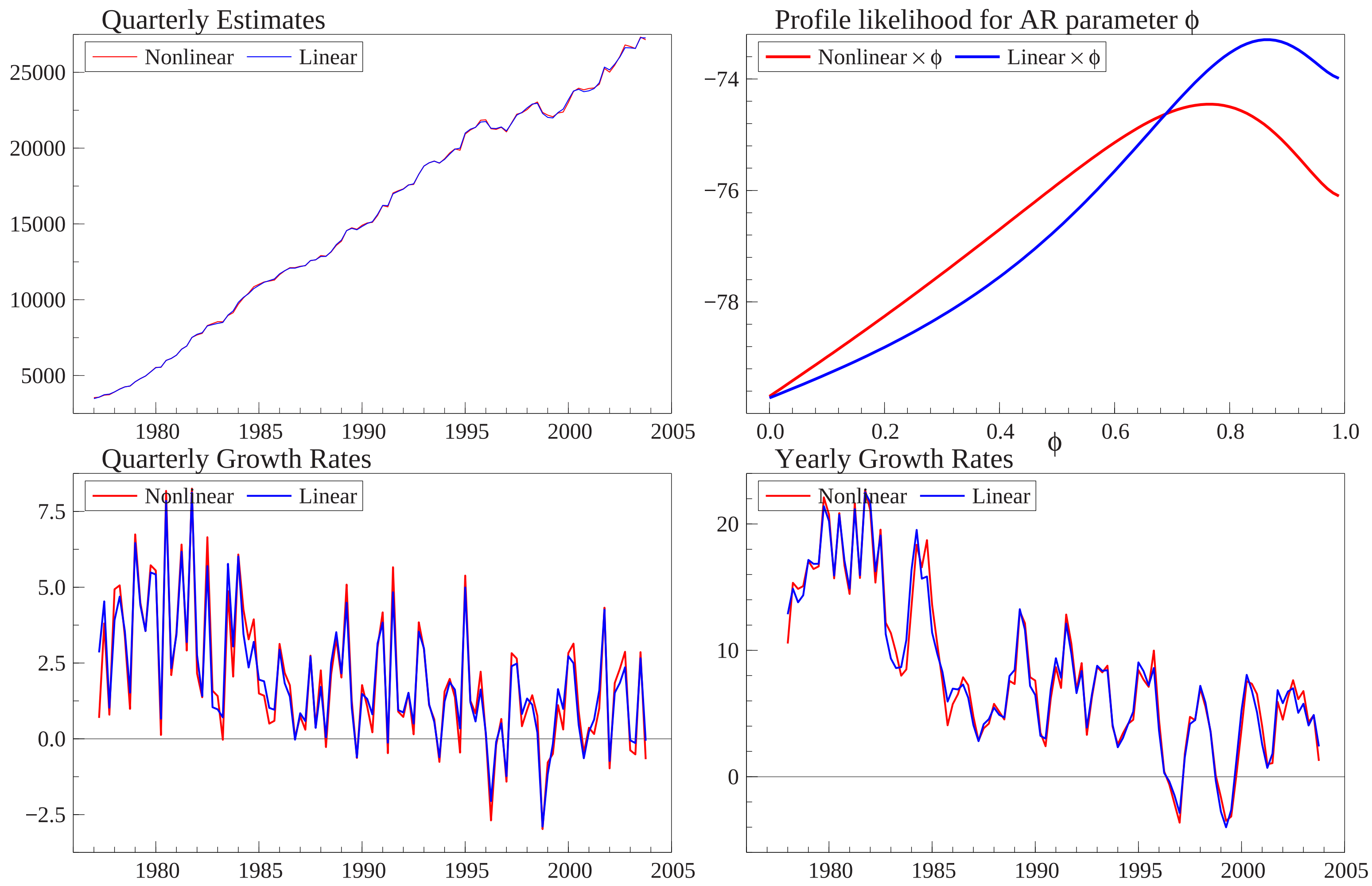

Figure 4: Temporal disaggregation of the annual production series for the Food, Bev. & Tob. sector using the linear and nonlinear Chow-Lin disaggregated model.