Benchmarking and Reconciliation of Time Series

An Applied Bayesian Method

José Luis Rojo-García, and José Antonio Sanz-Gómez

Department of Applied Economics, University of Valladolid, Spain

Abstract: The present article features a hierarchical Bayes method applied to solving problems of benchmarking and contemporaneous reconciliation across time series. This method enables the use of high frequency series to be either approximations or one or several related indicators. This method may be applied when facing flow or index disaggregation problems. The authors compare their results to classical procedures (viz. Denton univariate and Rossi multivariate methods) through the use of indicators. This article concludes that the suggested method bestows greater importance on the low frequency series profile, consequently providing smoother solutions than its counterparts.

Keywords: benchmarking, reconciliation, Bayes estimator, normal-gamma family

Introduction

When applied to time series data, benchmarking refers to the adjustment for discrepancies among related statistical information. Frequently, the problem arises when we are faced with two series corresponding to the same variable but measured with different frequencies, the more frequent series usually proving more accurate than the less frequent one.

Sometimes, the less frequent (henceforth, “annual”) and the more frequent (henceforth, “quarterly”) series do not refer to the same concept. However, both the annual series and the various quarterly indicators display related patterns linked to a common data generating process. For example, a country’s quarterly industrial added value may be estimated when the corresponding annual added value is available, using as indicators the Industrial Production Index (IPI) and Industrial employment, both of which have a quarterly frequency.


A second interesting topic is so-called reconciliation for time series. When analyzing economic time series, we are typically faced with systems of series together with their total (or mean). Sometimes, contemporaneous aggregation is not fulfilled (e.g., if the seasonal components have been removed). For example, we may disaggregate quarterly Spanish gross domestic product (GDP) by branch of activity and regions, using Social Security Contributor (SSC) series as indicators.

Even though this problem was tackled 70 years ago (Deming & Stephan, 1940), over the past few decades progress has been made in the use of these methods owing to the increased availability and vastly improved power of computers. Interested readers may refer to Cholette (1988), Chen and Dagum (1997), or Di Fonzo (2002, 2003), among others.

Two groundbreaking papers are cited as key references on benchmarking and reconciliation. Firstly, Di Fonzo and Marini (2005) solve several problems by using the so-called Denton’s movement preservation principle (Denton, 1971). Di Fonzo and Marini’s proposal includes temporal or contemporaneous aggregation constraints and one- or two-way systems of time series or marginal reconciliation problems, among others. More recent contributions from the same authors may be found in Di Fonzo and Marini (2011, 2013).
Secondly, the book by Dagum and Cholette (2006) analyzes and solves benchmarking, calendarization, or reconciliation problems by using regression-based models. In addition, they derive solutions based on autoregressive integrated moving average (ARIMA) and structural time series models. Their approach includes distribution or interpolation problems and uses flow and index time series. As in the Di Fonzo and Marini (2011) paper, the authors deal with one- and two-way problems.

Recently, Davies, Elliot, Aston, and Sayal (2015) compared the Denton and Cholette-Dagum methods with a wavelet benchmarking method proposed by Sayal et al. (2014) that performs well with outliers. Daalmans, Di Fonzo, Bikker, and Mushkudiani (2016) compare Denton Proportionate First Differences (Denton, 1971) with Causey and Trager Growth Rates Preservation methods (Causey & Trager, 1981), and an original time-symmetric variation combining both schemes.

Concurrently, related software has been constructed, although its use might at times be restricted to statistical agencies. Relevant items are BENCH (Statistics Canada, see Cholette, 1994), ECOTRIM (Barcellan, 1994), or the MATLAB library implemented by Quilis (2002).

The present paper puts forward a hierarchical normal-gamma model in solving the two temporal benchmarking and cross-sectional reconciliation problems for time series. Temporal aggregation might be performed by sum or by mean (for stocks or index series, respectively), and quarterly series might also present incomplete observations for the last available year.

Unlike mathematical adjustment methods, our Bayesian econometric model explicitly states the statistical relationship between the indicators (or preliminary series) and the unknown quarterly dependent ones. White noise, AR(1) (Chow & Lin, 1971), or Random Walks (Fernández, 1981) schemes are allowed for the error terms.

By using “difference” matrices, prior distributions are built to favor regular patterns for quarterly series. In this regard, our method focuses particular attention on the regularity of the estimated series, which offers an advantage when the indicators display volatile behavior. However, if the aim is to approximate the values of the indicator (in other words, when the series to be corrected is an approximation of the estimated one) methods such as the Denton approach might prove preferable.

This paper is inspired by a previous study carried out by Rojo and Sanz (2005) in which they solved benchmarking problems with only temporal restrictions. Introducing minor changes will, however, allow for the inclusion of cross-sectional ones. Operations have been performed using MATLAB (R2012b).

A Review of the Previous Work

Rojo and Sanz (2005) proposed a hierarchical Bayes method (the prior probability density function being obtained by chaining conditional distributions) allowing for temporal benchmarking between annual and quarterly series by using high frequency indicators. Despite its limited scope, the method can be applied to more complex situations. We here review the results of such previous works.

Let \( y = (y_1, \ldots, y_n)^T \) be the non-stochastic annual time series, observed for \( N \) consecutive years, and let \( x = (x_{11}, \ldots, x_{41}, x_{12}, \ldots, x_{42}, x_{13}, \ldots, x_{43}, x_{14}, \ldots, x_{44}) \) be the corresponding (unknown) quarterly series. Rojo and Sanz (2005) assumed that the quarterly series has no missing values. Thus, we can establish that \( n = 4N \). We also assumed that both series match, fulfilling the temporal restriction

\[
y_{i1} = x_{4i-1} + x_{4i-2} + x_{4i-3} + x_{4i}, \quad i = 1, \ldots, N. 
\]  

As far as the prior distribution is concerned, we assume a normal distribution \( N(\mu, \tau^{-1}P^{-1}) \) for \( x \), given the precision \( \tau \) and \( y \) data, and obeying the temporal restriction (1). We then modify the prior distribution by adding a second factor,

\[
\pi(x|\tau, y) \propto \tau^{\delta} \cdot \exp \left\{ -\frac{\tau}{2} \left[ (x - \mu)' P(x - \mu) + x'Dx \right] \right\},
\]  

\[
(2)
\]

\( D \) being a \((n-1) \times n\) matrix, with \( d_{ii} = -1, d_{i(i+1)} = 1, \) \( i = 1, \ldots, n-1 \) (first differences) or a \((n-2) \times n\) matrix, with \( d_{ii} = 1, d_{i(i+1)} = -2, d_{i(i+2)} = 1, \) \( i = 1, \ldots, n-2 \) (second differences).

For the \( \tau \) parameter, we take a gamma distribution \( \pi(\tau|y) \propto \tau^{\delta+1} e^{-\beta \tau}, \tau > 0 \), with shape \( \alpha \) and scale \( 1/\beta \) (rate parameter \( \beta \)) and for the likelihood function, we assume an econometric model with \( k \) indicators,

\[
x_i = \sum_{j=1}^{k} \delta_j z_j^i + \epsilon_i, \quad i = 1, \ldots, n, 
\]

written in matricial form as \( x = Z\delta + \epsilon \), being \( \delta = (\delta_1, \ldots, \delta_k), \) \( Z = (z^i) \) and \( \epsilon = (\epsilon_1, \ldots, \epsilon_n)^T \). We also assume a normal prior distribution \( (\delta|\tau,y) \sim N(\bar{\delta}, \tau^{-1}P_\delta^{-1}) \) for the parameters of the model, and normal, zero

---

1 We assumed that the annual series is a flow. The constraint (1) can be written as \( y_i = (x_{i-3} + x_{i-2} + x_{i-1} + x_{i})/4 \) for index (level) series, and as \( y_i = x_i \) for end-of-period stock data, \( i = 1, \ldots, N \). Although Rojo and Sanz (2005) do not solve these other problems, the required changes are obvious.
mean, error terms, \( e \sim N_\nu(0, \tau^{-1}P_e^{-1}) \). Specifically, the likelihood is

\[
L(x, \delta, \tau | y, Z) \propto \tau^{m/2} \exp \left\{ -\frac{\tau}{2} (x - Z\delta)^T P_e (x - Z\delta) \right\}.
\]  

(4)

By joining prior and likelihood we obtain the posterior distribution

\[
p(x, \delta, \tau | y, Z) \propto \tau^{a+1+k/2} \cdot \exp \left( -\tau \times U/2 \right),
\]

with

\[
U = 2\beta + (x - \mu)^T P(x - \mu) + x^T D x
\]

\[
+ (\delta - \delta')^T P_\delta (\delta - \delta') + (x - Z\delta)^T P_e (x - Z\delta),
\]

\[x\] fulfilling the temporal restriction (1).

We include that restriction in the model because the \( n \) linear restrictions concentrate the values of the quarterly series into a low-dimensional subspace. Linear restrictions of (1) imply replacing \( n \) values from \( x \). With no loss of generality, we replace the last (fourth) quarter for each year,\(^2\) writing the restricted distribution as a function of the \( 3n \) values, \( x_r = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, \ldots, x_{n-3}, x_{n-2}, x_{n-1}) \). We can easily prove that the quarterly objective matrix, \( x \), can be expressed as a function of \( x_r \),

\[
x = W x_r + y_r,
\]

where \( W = I_N \otimes A \) and \( y_r = (0, 0, 0, y_1, 0, 0, 0, y_2, \ldots, 0, 0, 0, y_{2n}) \), being \( A = (I_3, -I_3) \), with \( t_i = (1, 1, 1) \) and \( I_3 \) the identity matrix of order 3. Replacing (6) in the unrestricted posterior provides the restricted posterior distribution in terms of \( x_r \). It results\(^3\) in a multivariate\(^4\) Multivariate Student \( t \) (MS-\( t \), from now)

\[
p(x_r | y, Z) \propto \left[ B_r + (x_r - A_r)^T W M W (x_r - A_r) \right]^{-\frac{n+a}{2}}.
\]

(7)

The posterior distribution for \( \delta \) is derived similarly. It also gives an MS-\( t \), specifically \( p(\delta|y, Z) \propto \left[ B_\delta + (\delta - A_\delta)^T M_\delta (\delta - A_\delta) \right]^{-\frac{3n/2}{2}} \). Finally, we obtain the posterior distribution for the precision parameter, \( \tau \), a gamma density \( p(\tau | y, Z) \propto \tau^{m+a-3n/2} \cdot \exp \left\{ -\frac{\tau}{2} B_\tau \right\} \), with shape \( m + a - 3n/2 \) and scale \( 2B_\tau^{-1} \).

Assigning specific values for the hyper-parameters included in the prior distributions is carried out in an interesting manner. Rojo and Sanz (2005) do not employ non-informative priors for the implied parameters, using instead the marginal priors to derive the appropriate rules for the assignment. Partitioning the time interval into batches is an initial step toward obtaining the most adequate values. In our paper, we use the same method.

## Problem-Solving Involving Temporal Benchmarking and Contemporaneous Reconciliation

As stated above, we apply the previous ideas to solving more general benchmarking and reconciliation problems. The series involved can be stock or level ones. In addition, the method allows incomplete years to be processed.

As in Rojo and Sanz (2005), our model contemplates an econometric, indicator-based model for each estimated series, the error terms following either White noise, AR(1), or Random walk schemes. In this paper, we assume that the specifications for error terms are the same for all the series; otherwise, changes would be easy to implement. Moreover, we assume prior independence among the models, which implies block-diagonal covariance matrices. Such a hypothesis is necessary due to the high number of hyper-parameters involved. Nevertheless, the posterior covariance matrix is not, of course, block-diagonal.

As in the previous section, the following symbolic notation is used for the temporal framework: The low frequency periods are called “years,” and we employ the term “quarters” for the high frequency ones. There are \( N \) full years with \( m \) quarters each, and \( r \) extra quarters for the following year (otherwise, we take \( r = 0 \)). As such, the total size for the quarterly series will be \( n = N \cdot m + r \).

As mentioned above, Rojo and Sanz (2005) used a method that established for the vector \( x \), the sub-vector \( x_r \) including only the “free” components, linking both vectors by means of the linear restriction, \( x = W \cdot x_r + y_r \). Thus, the goal now is to translate the temporal and/or contemporaneous constraint into the required format.

Readers can find in the Appendix the specifications of the nine cases solved by the authors, and which we present schematically in Table 1. We detail only some of the specifications that will prove useful when implementing the proposed methods. These are described in the context of Case 0. Specifically, if the series were an index rather

---

\(^2\) It can be seen that selecting any quarterly ordinal does not affect the solution.

\(^3\) The complexity of the expressions which the matrices define means they have not been included in the present work. Readers may refer to them in Rojo and Sanz (2005).

\(^4\) A continuous \( d \)-dimensional random vector, \( Z \), has a Multivariate Student \( t \) distribution if \( Z \) has density

\[
f(z) \propto |v + (z - \mu_v)| A |z - \mu_v|^{-d/2}.
\]

The parameters are the degrees of freedom, \( v \), the position matrix, \( \mu_v \), and the scale matrix, \( A \). Their covariance matrix is \( |v| (v-2)A^{-1} \) for \( v > 2 \). See, for example, Zellner (1971), p. 383, who justifies the choice of this version of Multivariate Student distribution for Bayesian problems.

© 2017 Hogrefe Publishing

Methodology (2017), 13(4), 123–134
than a stock, we would transform \( y_i^* = m \cdot y_i \), \( i = 1, \ldots, R \) and would then use \( y_i^* \) as the annual series.

Moreover, if we wished to estimate several quarters for a final incomplete year, we would introduce several modifications: Firstly, the matrix \( W \) from expression (6), defined in Rojo and Sanz (2005) as \( W = I_n \otimes A \), would be replaced by

\[
W = \begin{pmatrix} I_n \otimes A & 0 \\ 0 & I_r \end{pmatrix},
\]

where the null matrices are of the right dimension.

Secondly, the column matrix, \( y_v \), from (6), defined as \( y_v = (0, 0, 0, y_1, \ldots, 0, 0, 0, y_N) \), would become \( y_v = (0_{1x(m-1)}, y_1, \ldots, 0_{1x(m-1)}, y_N, 0_{1w}) \), then including a final null matrix corresponding to the incomplete year.

Thirdly, the “free” estimated \( x_r \), including only the three first quarters (see Footnote 2) for each year in the previous proposal, will now add the values corresponding to the quarters from the incomplete year, that is to say,

\[
x_r = \left( x_1, \ldots, x_{m-1}, \ldots, x_{(N-1)m+1}, \ldots, x_{(N-1)m+(m-1)}, x_{Nm+1}, \ldots, x_{Nm+r} \right).
\]

Finally, as in Rojo and Sanz (2005), the prior average for \( x \) (to be included in the prior distribution) is now obtained by applying the Boot, Feibes, and Lisman disaggregation method (Boot & Feibes, 1967) to the annual series. The advantage of this method is the absence of indicators in the temporal disaggregation procedure. It should be noted that the information from the indicators already enters the model through the likelihood function.

Processing for incomplete years can also be carried out using the expression (7.4) from Bloem, Dippelsman, and Maelhe (2001). Specifically, we estimate \( y_{N+1} \) as

\[
y_{N+1} = y_N \cdot \left[ \frac{3}{6} \cdot \frac{y_N}{y_{N-1}} + \frac{2}{6} \cdot \frac{y_{N-1}}{y_{N-2}} + \frac{1}{6} \cdot \frac{y_{N-2}}{y_{N-3}} \right],
\]

and then apply the Rojo and Sanz (2005) method to obtain the temporal disaggregation of \( (y_1, \ldots, y_N, y_{N+1})' \).

Finally, we remove the last \( m - r \) terms for the quarterly estimated series.

Which procedure is now advisable? We have detected high volatility for the estimates of the quarters corresponding to incomplete years due to the large number of parameters involved. Stability in the estimation is best achieved by complementing the use of the formulae (10) with the parallel extrapolation for the indicators, thus making use of expression (7.5) from Bloem et al. (2001). Said modification stabilizes the estimation for these parameters at the expense of turning the posterior average into a conditional one.

### Illustration of the Bayesian Benchmarking Methods

In this section, we present several examples showing the performance of the proposed Bayesian method. The first example illustrates the feasibility of the method and compares the results with the so-called multivariate Denton method (Bloem et al., 2001). The second and third examples compare the Bayesian method with the univariate Denton (1971) and the multivariate Rossi method (Rossi, 1982), respectively.

### One-Way Example

We now consider the number of employed persons estimated by the Economically Active Population Survey (EAPS) between 2008 and 2013 (24 quarters) for the four branches of activity: EAG (Agriculture), EIN (Industry, including Energy), ECO (Construction), and ETS (Services). The total number of employed persons, ETT, is also known. The National Statistics Institute (Instituto Nacional de Estadística, 2014) provides the raw data and the rate of variation for the seasonally adjusted (SA) series.

By using the X-12(ARIMA) method as is included in IHS-EViews software (IHS-Eviews, IHS, Englewood, CO, USA),

---

Table 1. Summary of the problems solved in the paper

<table>
<thead>
<tr>
<th>Proc. No.</th>
<th>Dimension</th>
<th>Constraints</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Temporal</td>
<td>No contemporaneous constraint was found.</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Temporal and contemporaneous</td>
<td>Total quarterly series is known.</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Temporal and contemporaneous</td>
<td>Total quarterly series is estimated.</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>Temporal and contemporaneous</td>
<td>Total quarterly series is known.</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>Temporal and contemporaneous</td>
<td>Total quarterly series is known.</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>Temporal and contemporaneous</td>
<td>The row and column marginals are known.</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>Temporal and contemporaneous</td>
<td>The row and column marginals are known.</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>Temporal and contemporaneous</td>
<td>The row and column marginals are known.</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>Temporal and contemporaneous</td>
<td>The row and column marginals are known.</td>
</tr>
</tbody>
</table>

Table 2. Posterior average and variance for the parameters included in the Econometric models when reconciling Branch employed persons with Total employed persons, 2008:1–2013:4

<table>
<thead>
<tr>
<th></th>
<th>OPTEPS = 1</th>
<th>OPTEPS = 2</th>
<th>OPTEPS = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Var.</td>
<td>Average</td>
</tr>
<tr>
<td>AG</td>
<td>1.0001</td>
<td>0.6942</td>
<td>1.0003</td>
</tr>
<tr>
<td>IN</td>
<td>0.9999</td>
<td>0.0609</td>
<td>0.9999</td>
</tr>
<tr>
<td>CO</td>
<td>0.9899</td>
<td>0.9292</td>
<td>0.9778</td>
</tr>
<tr>
<td>TS</td>
<td>0.9999</td>
<td>0.0351</td>
<td>1.0000</td>
</tr>
<tr>
<td>TT</td>
<td>0.9999</td>
<td>0.0081</td>
<td>0.9999</td>
</tr>
</tbody>
</table>

The SA series have been derived (EAGSA, EINSA, ECOSA, ETSSA, and ETTSAs). Unfortunately, these series do not fulfill the contemporaneous constraint, which is to say, the SA series for the branches of activity do not totalize ETTSAs.

We then applied the Case 2 procedure by imposing temporal constraints (the mean of the four quarterly data for each year will be the annual value for the original series. However, the constraint is not fulfilled by the SA series) and contemporaneous constraints (the sum of the four branch series will match the total series for each quarter).

For the quarterly variables, a Bayesian econometric model (with the SA series as the only indicator, and without constant term) is assumed. Table 2 shows the estimated models with the posterior average and variance for each parameter (it should be remembered that the posterior distribution for each parameter is a noncentral rescaled Student distribution). Note that the posterior average does not differ greatly from one because the SA series is an approximation to the estimated series.

We here present several estimations that differ in the order for differences (DIF = 1 or 2) or in the scheme used for the error term for models (OPTEPS = 1, 2, or 3, for White noise, AR(1), or Random walk, respectively).

The indicator called MADR (Mean Absolute Difference Ratio) is now used to evaluate the relative performance of series. Specifically, the more regular the series estimated, the higher the values it takes. Given two series, \( u_t, v_t, t = 1, \ldots, T \), we define

\[
MADR (u, v) = \left( \frac{1}{(T-1)} \right) \frac{\sum_{t=2}^{T} |u_t - u_{t-1}|}{\sum_{t=2}^{T} |v_t - v_{t-1}|}. \tag{11}
\]

Note that MADR > 1 if series \( v_t \) is smoother than \( u_t \). Using the said indicator, the authors analyze the relative regularity between each indicator and the corresponding final estimation, with no relevant patterns being found.

We also consider the MADRR (Mean Absolute Difference of Rates Ratio) indicator, a measure that evaluates the similarity between the growth rates of the estimated series and those of the indicator or preliminary series. If, for example, \( u_t, v_t, p_t, t = 1, \ldots, T \) are, respectively, two estimated series and the preliminary one, MADRR is evaluated as

\[
MADRR (u, v) = \left( \frac{1}{(T-1)} \right) \frac{\sum_{t=2}^{T} |u_t - u_{t-1}|}{\sum_{t=2}^{T} |v_t - v_{t-1}|}. \tag{12}
\]

This measure is denoted by Di Fonzo and Marini (2010) as \( r_1 \), with said authors also evaluating a similar measure of second order, denoted as \( r_2 \). Readers should note that MADRR may only be calculated for models with one indicator, and that it makes no sense unless the (unique) indicator is an approximation to the estimated series. Table 3 presents MADRR values for several specifications of our model. For example, when using DIF = 1, the approximation to the growth rates of ETSSA if OPTEPS = 2 (AR(1) model for errors) is 98.39% of the same approximation as for OPTEPS = 1 (White noise). Table 3 also shows that considering distinct values for DIF or for OPTEPS has very limited effects on the approximations to the growth rates of the SA series.

As pointed out above, the Bayesian method yields smoother series than the multivariate Denton method and, consequently, the growth rates for the former are further from the SA growth rates than those for the latter. Table 4 shows that this fact is especially relevant for ETS.

The authors also replicated the Denton multivariate and Bayesian (Case 2) procedures, this time with series ending for 2013:2. The outcome does not differ from the one given above.

### Comparison Between Bayesian and Denton One-Way Methods

In this section, we compare our method to the one-dimen-
sional Denton additive method (Denton, 1971). Said author adjusts a five-year long quarterly artificial series with the values 50, 100, 150, and 100, in the four calendar quarters, for each year. The annual totals to which the series is to be
benchmark are assumed to be 500, 400, 300, 400, and 500. Denton employs either additive or proportional loss functions. By using this example, we wish to show that our procedure provides more regular solutions than those of Denton (1971), although they are more distanced from the values of the indicators, as pointed out in the Introduction.

Table 5 contains the results. We used $\text{DIF} = 1$ for priors the three allowed schemes for errors. Comparisons with the Bayesian method are carried out with respect to the Denton additive loss function of first order (first and third columns) or with respect to the proportional loss function, which is also first order (second and fourth columns). The comparison is made by using MADRR (left panel) and MADR (right panel).

Readers can see that MADRR values are much greater than 1, especially for the Denton Proportional Loss function. MADR takes values smaller than 1 (values of approx. 0.2). Two conclusions emerge: First, the Denton method adjusts to the indicator values more, and second, our solutions are smoother than those of the Denton method.

Figure 1 presents the balance between the two methods graphically. The piecewise horizontal bold line shows annual values divided into four equal parts (for reference). The solid line is the above-defined indicator (an artificial approximation to the quarterly series). The dashed and dotted lines show the Denton estimate for an additive loss function (of first order). Note that the Denton solution follows the pattern of the indicator, thus showing a highly volatile path. Finally, the dotted line depicts the Bayesian solution $(\text{DIF} = 1$, White noise for errors$)$. As can be seen, even if the path follows the profile of the indicator, the general performance adjusts more to the annual behavior. The conclusions are virtually the same for other specifications.

### One Two-Way Example

Consider now a two-way classification for employed persons in Spain with data drawn from the EAPS (as before). Data are arranged in four rows (the four branches of activity) and two columns (by gender, male and female). From the raw data, we know the quarterly and annual values for the eight series, and the Total Employed Persons (ETT). As in 4.1, we derived the series corrected from seasonality (nine SA series, including the total SA). The seasonal adjustment of the series does not guarantee compliance of the contemporaneous restrictions, although the lack of fit between the total series and the sum of the individual series (all of them seasonally adjusted) is not very large (between $-0.15\%$ and $0.12\%$).

Even though the Bayesian method allows the total to be estimated (Case 8), the Rossi (1982) method (an extension of the Chow-Lin procedure to the multivariate case) that we use for comparison requires the total to be known. Thus, we take ETTSA as binding and estimate the eight individual series (EAGM, EINM, ECOM, ETSM, EAGF, EINF, ECOF, ETSF) restricted to the temporal and contemporaneous constraints. Beyond numerical comparisons, theoretical comparisons of the two methods do not seem possible due to the different conceptual nature, since the method proposed here has a Bayesian structure.

Table 6 provides us with the posterior average and variance of the eight individual econometric models. Their interpretation is the same as above. Table 7 presents the MADRR value for the eight series when compared for the allowed schemes for the error terms. Generally speaking, the values obtained do not differ greatly from 1, which seems to indicate that, for different schemes, the growth rates of the estimated series keep the same ratio with.

---

### Table 3. MADRR indicator for the reconciled Total employed persons, 2008:1–2013:4

<table>
<thead>
<tr>
<th>DIF = 2</th>
<th>OPTEPS</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10.013</td>
<td>10.170</td>
<td></td>
</tr>
<tr>
<td>DIF = 1</td>
<td>OPTEPS</td>
<td>2</td>
<td>0.9839</td>
<td>10.157</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.9841</td>
<td>10.002</td>
</tr>
</tbody>
</table>

Note: MADRR = Mean Absolute Difference of Rates Ratio.

### Table 4. MADRR indicator for the estimated branch series, 2008:1–2013:4

<table>
<thead>
<tr>
<th>EAG</th>
<th>EIN</th>
<th>ECO</th>
<th>ETS</th>
<th>ETT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0741</td>
<td>2.0233</td>
<td>4.3596</td>
<td>10.2520</td>
</tr>
<tr>
<td>OPTEPS</td>
<td>2</td>
<td>2.3825</td>
<td>1.7528</td>
<td>3.4335</td>
</tr>
<tr>
<td>3</td>
<td>2.0886</td>
<td>1.6634</td>
<td>4.0504</td>
<td>10.2360</td>
</tr>
</tbody>
</table>

Notes. The upper side of the ratio is for the Bayesian method, and the lower one is for the multivariate Denton method. MADRR = Mean Absolute Difference of Rates Ratio.

### Table 5. MADRR and MADR values for the reconciled Total Employed Persons, 2008:1–2013:4

<table>
<thead>
<tr>
<th></th>
<th>Additive</th>
<th>Proportional</th>
<th>Additive</th>
<th>Proportional</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPTEPS</td>
<td>2.848</td>
<td>5.816</td>
<td>0.234</td>
<td>0.227</td>
</tr>
<tr>
<td>2</td>
<td>2.731</td>
<td>5.576</td>
<td>0.268</td>
<td>0.280</td>
</tr>
<tr>
<td>3</td>
<td>2.942</td>
<td>6.008</td>
<td>0.201</td>
<td>0.195</td>
</tr>
</tbody>
</table>

Notes. MADR = Mean Absolute Difference Ratio; MADRR = Mean Absolute Difference of Rates Ratio. The upper side of the indicators correspond to the Bayesian method, and to the Denton method for the lower side.
regard to the rates for the SA series. However, relevant differences can be observed for small series, such as ECOF, women employed in Construction, probably due to the presence of unstable shapes.

Table 8 shows the comparison between the revisions for SA growth rates for the Bayesian method (with several schemes for error models) and the Rossi (1982) method. In general, the MADRR values are greater than 1 (which is to say, the Bayesian method causes greater corrections to the SA growth rates than the Rossi method). The only exception is found for ECOF, perhaps due to the same arguments cited above.

![Figure 1. Estimation from Denton artificial data and comparison with Bayesian estimate.](image-url)
Methodology

In the present article, we deal with solving several benchmarking and reconciliation problems, including temporal and contemporaneous constraints. Di Fonzo and Marini (2005) or Dagum and Cholette (2006), among others, have proposed solutions for some of these problems. We suggest an innovative method, specifically a Bayesian normal-gamma model, also including Bayesian econometric models to explain the relations between series and indicators, allowing for several schemes for their error terms. Prior distributions include difference matrices aiming to increase the probability for regular paths in the high frequency series, when correcting the information provided by the indicators.

The Bayesian method is applied to one- and two-way systems of series, in this case solving the estimations for marginals or for the whole system. In any case, series can be of the stock or index kind, and no problems emerge if the last year is not completed. Unlike Denton univariate or multivariate methods, more than one indicator is allowed for the stock or index kind, and no problems emerge if the last year is not completed. Unlike Denton univariate or multivariate methods, more than one indicator is allowed for the Bayesian method, and the econometric models can include a constant term.

We have illustrated the method with data taken from the Spanish Economically Active Population Survey (EAPS) between 2008 and 2013. The example reconciles Employees by Industries with Total Employees, when data have been seasonally adjusted. The Bayesian solution is compared with the Denton multivariate method (Bloem et al., 2001). The second example takes artificially generated data from Denton (1971), the goal being temporal benchmarking between annual series and quarterly approximation. The last example proposes temporal and contemporaneous reconciliation across a two-way system of employed persons series (also taken from EAPS) classified by branch (four rows) and gender (two columns). Our solution is then compared with that of the Rossi (1982) method.

For this example, we show the feasibility of the Bayesian method. Additionally, we compare our method with the classical solutions by using two statistics measuring the smoothness of the estimated series (MADR) or the change in the growth rates of the approximated series (MADRR).

The outcome suggests that the Bayesian method provides series with more regular paths than the indicators, a particularly interesting property when the indicators are volatile, a frequent handicap in benchmarking. It should also be noted that the Bayesian method proves optimal in a certain sense, whereas the Denton and Rossi methods are two-stage optimal, and display only sequential optimality.

The authors will seek future improvements in basically two directions: First, benchmarking volume indicators in the presence of contemporaneous constraints, as they appear in national or regional accounts, and second, deriving the Bayesian model and its practical implementation when using proportional adjustment criterion, which is extremely interesting when the order of magnitude of the series is very different.

Conclusions

Table 8. MADRR values when comparing the Bayesian solution (upper side for ratio) and the Rossi method (lower side) for distinct models for error terms

<table>
<thead>
<tr>
<th>Series</th>
<th>EAGM</th>
<th>EINM</th>
<th>ECOM</th>
<th>ETSM</th>
<th>EAGF</th>
<th>EINF</th>
<th>ECOF</th>
<th>ETSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>W.N./Rossi</td>
<td>3.0513</td>
<td>2.6578</td>
<td>10.1318</td>
<td>6.9882</td>
<td>1.1732</td>
<td>2.4816</td>
<td>0.2118</td>
<td>11.2163</td>
</tr>
<tr>
<td>Ar(1)/Rossi</td>
<td>3.1081</td>
<td>2.7935</td>
<td>9.1261</td>
<td>5.7790</td>
<td>1.2756</td>
<td>1.8708</td>
<td>0.2241</td>
<td>10.4402</td>
</tr>
<tr>
<td>R.W./Rossi</td>
<td>3.1578</td>
<td>2.7870</td>
<td>9.6877</td>
<td>6.4523</td>
<td>1.3194</td>
<td>1.9938</td>
<td>0.2507</td>
<td>11.2987</td>
</tr>
</tbody>
</table>

Notes. MADRR = Mean Absolute Difference of Rates Ratio. DIF = 1 has been used for prior distributions.

References


MR2527.


Received September 26, 2014
Revision received January 31, 2017
Accepted February 17, 2017
Published online November 24, 2017

José Luis Rojo García is a Professor in the Department of Applied Economics at The University of Valladolid, Spain. His research interests are in the applications of Bayesian statistics to regional economics and regional accounts.

José Antonio Sanz Gómez is an Associate Professor in the Department of Applied Economics, and is currently also Dean of the Faculty of Economics at The University of Valladolid, Spain. His main fields of research are applied Bayesian analysis and forecasting in regional economics.

José Luis Rojo García
Department of Applied Economics
University of Valladolid
Avenida Valle Esgueva, 6
47011 Valladolid
Spain
rojo_s@eco.uva.es
Appendix

Here we describe the particularities of each case studied. As a general rule, “\( x \)” denotes the variables in high frequency (quarterly), and “\( y \)” describes the low frequency series (annual). The indicator-based model is described in the same way in all cases, in the form

\[
x_t = \sum_{j=1}^{k} \delta_{ij} y_{ij} + e_{it}, \quad t = 1, \ldots, n, \quad i = 1, \ldots, R,
\]

(A.1)

for each \( x_i \) series to be estimated.

The temporal restrictions always establish the relation between the series to be estimated and those which are low frequency. The procedure is implemented with minor changes for averages \( \left( y_i = (x_{m(i-1)+1} + \cdots + x_{m(i)}/m \right) \) or for sums \( \left( y_i = x_{m(i-1)+1} + \cdots + x_{m(i)} \right) \).

In all cases, the temporal and contemporaneous restrictions are included in the model, such that the probabilistic model is not a degenerate one. This is done by building the relation \( x = W \cdot x_r + y_r \), where \( x \) is the vector column that is the set of high frequency series and \( x_r \) brings together the “free” components of \( x_i \), in other words, not subject to restrictions. In each case, we show how \( x \) and \( y_r \) are specified.

Together with each case studied, we provide an illustrative example taken from the national and regional accounts so that the reader may be able to distinguish them. As the temporal restrictions are always included, we make no reference to the availability of the annual series in these illustrations.

One-Way Classified System of Time Series

Case 0: Only Temporal Constraints

Objective: The goal is to estimate \( x = (x'_1, \ldots, x'_R)' \), where \( x_i = (x_{1i}, \ldots, x_{ni})' \), \( i = 1, \ldots, R \).

Data: Annual values for all the series, \( y_1, \ldots, y_R \), with \( y_i = (y_{i1}, \ldots, y_{iN}) \), \( i = 1, \ldots, R \), and quarterly values for the indicators.

Restrictions: Only temporal ones.

Likelihood: Derived from the indicator-based econometric model as (A.1) for each series.

The matrices involved in the linear relation are \( W = I_n \otimes A = \begin{pmatrix} I_n \otimes A & 0 \\ 0 & I_r \end{pmatrix} \),

\[
y_r = \left( O_{1(m-1)}, y'_1, \ldots, O_{1(m-1)}, y'_N, O_{1(r)} \right)', \text{ and}
\]

\[
x_r = \left( x_1, \ldots, x_{m-1}, \ldots, x_{(N-1)m+1}, \ldots, x_{(N-1)m+(m-1)}, \right. \]

\[
x'_{nm+1}, \ldots, x'_{Nnm+r} \right).
\]

Example: Same as Case 0, with the quarterly national GDP official estimates now being known.

Case 1: Contemporaneous Restriction (Total Sum Is Known)

Objective: \( x = (x'_1, \ldots, x'_R)' \) as for Case 0.

Data: Annual values for each series, \( y'_1, \ldots, y'_R \), quarterly total, \( x = (x_1, \ldots, x_R)' \) and quarterly values for the indicators.

Restrictions: In addition to the temporal ones, we have \( n \) contemporaneous restrictions.

\[
x_t = \sum_{i=1}^{R} x_{it}, \quad t = 1, \ldots, n.
\]

Likelihood: Derived from the indicator-based econometric model as (A.1) for each series.

Now, denoting \( x_t = (x'_1, \ldots, x'_{R-1})' \), with

\[
x'_i = (x_{i1}, \ldots, x_{i(m-1)}), \ldots, x_{i(N-1)m+1}, \ldots, x_{i(N-1)m+(m-1)}, x_{im+1}, \ldots, x_{Nm+r}) \text{ for } i = 1, \ldots, R - 1,
\]

we can write \( x = W x_r + y_r \), now being

\[
W = \begin{pmatrix} I_{R-1} & \otimes & I_n \otimes A \\ \vdots & \vdots & \vdots \end{pmatrix}.
\]

\[
y_r = \left( \begin{pmatrix} I_{R-1} & \otimes & I_n \end{pmatrix} \right) \cdot \left( \begin{pmatrix} y'_{1} \\ \vdots \\ y'_{R-1} \end{pmatrix} \right), \text{ and}
\]

\[
y_{ir} = \left( O_{1(m-1)}, y'_{1i}, \ldots, O_{1(m-1)}, y'_{Ni}, O_{1(r)} \right)', \quad i = 1, \ldots, R - 1.
\]

Example: Same as Case 0, with the quarterly national GDP official estimates now being known.

Case 2: Contemporaneous Restriction (Total Sum Is Estimated)

Objective: \( x = (x'_1, \ldots, x'_R, x')' \).

Data: Annual values for each series, \( y'_1, \ldots, y'_R \), with \( y_i = (y_{i1}, \ldots, y_{iN}) \), \( i = 1, \ldots, R \), and quarterly values for the indicators, now including those corresponding to the total series \( x \).
**Restrictions:** In addition to the temporal ones, we have \( n \) contemporaneous restrictions \( x_t = \sum_{i=1}^{R} x_{it}, \ t = 1, \ldots, n \) as in Case 1.

**Likelihood:** We have an econometric model (A.1) for each individual series, but now have an additional econometric model, corresponding to the total, \( x_t = \sum_{i=1}^{k} \delta_{i} x_{it} + \varepsilon_t, \ t = 1, \ldots, n. \)

Thus, we can write \( x = W x_r + y_v, \) now being \( x = (x'_1, \ldots, x'_R, x'_C)' \) and \( x_r = (x''_1, \ldots, x''_R)' \), with \( x''_i = (X_{t1}, \ldots, X_{t(m-1)}, \ldots, X_{t(N-1)m+1}, \ldots, X_{t(N-1)m+m-1}, X_{tNm+1}, \ldots, X_{tNm+r})', \ i = 1, \ldots, R - 1. \) The matrices involved are

\[
W = \begin{pmatrix} I_R \\ I_R \end{pmatrix} \otimes \begin{pmatrix} I_n \otimes A & 0 \\ 0 & I_r \end{pmatrix} \quad \text{and} \quad y_v = \begin{pmatrix} y_{1v} \\ \vdots \\ y_{Rv} \\ y_{rv} \end{pmatrix},
\]

with \( y_{1v} = (y_{11}, \ldots, y_{1R}, y_{1C})' \) and \( y_{r} = (y_{R1}, \ldots, y_{R1}, y_{RC})'. \)

**Example:** Quarterly regional GDP estimates by Industry, also estimating quarterly regional aggregate GDP for all industries.

### Marginals for a Two-Way Classified System of Time Series

Consider a two-way system of quarterly time series, \( x_{ij}, \ i = 1, \ldots, R, \ j = 1, \ldots, C. \)

Assume that annual series \( y_{ij}, \ i = 1, \ldots, R, \ j = 1, \ldots, C. \)

In this case, we aim to estimate the marginal series, \( x_{it} = \sum_{j=1}^{C} x_{ij}, \ t = 1, \ldots, n, \ i = 1, \ldots, R \) and \( x_{jt} = \sum_{i=1}^{R} x_{ij}, \ t = 1, \ldots, n, \ j = 1, \ldots, C. \)

**Case 3: Estimation of the Marginals, Total Sum Known**

**Objective:** We now wish to estimate the row and column marginals, \( x_1, \ldots, x_R, \) and \( x_{1c}, \ldots, x_{C}. \)

**Data:** In addition to the annual series \( y_{ij}, \ i = 1, \ldots, R, \ j = 1, \ldots, C, \) we know the annual marginals, \( y_{1}, \ldots, y_{R}, \) and \( y_{1c}, \ldots, y_{C}. \)

**Restrictions:** In addition to the temporal ones, we have one contemporaneous restriction, \( x_1 + \cdots + x_R = x. \)

**Likelihood:** The indicator-based econometric models for the marginal quarterly series also come into effect,

\[
x_{it} = \sum_{l=1}^{k} \delta^l_{i} x_{il} + \varepsilon_{it}, \ t = 1, \ldots, n, \ i = 1, \ldots, R \quad \text{and} \quad \text{Case 4: Estimation of the Marginals, Total Sum Unknown}.

**Objective:** We now wish to estimate the row and column marginals, \( x_{1}, \ldots, x_{R}, \) and \( x_{1c}, \ldots, x_{C}. \)

**Data:** In addition to the annual series \( y_{ij}, \ i = 1, \ldots, R, \ j = 1, \ldots, C, \) we know the annual marginals, \( y_{1}, \ldots, y_{R}, \) and \( y_{1c}, \ldots, y_{C}. \)

**Restrictions:** In addition to the temporal ones, we have one contemporaneous constraint, \( x_1 + \cdots + x_R = x. \)

**Likelihood:** The indicator-based econometric models provide the likelihood function,

\[
x_r = \sum_{l=1}^{k} \delta^l_{r} x_{rl} + \varepsilon_{r}, \ t = 1, \ldots, n, \ j = 1, \ldots, C \quad \text{and} \quad y_v = \begin{pmatrix} y_{1v} \\ \vdots \\ y_{rv} \end{pmatrix}, \text{and}
\]

\[
W = \begin{pmatrix} I_R & 0 \\ 0 & I_C \\ I_R & -I_C \end{pmatrix} \otimes \begin{pmatrix} I_N \otimes A & 0 \\ 0 & I_r \end{pmatrix}.
\]

**Example:** Now, we estimate the quarterly GDP for the Provinces (NUTS3) of a Region and the quarterly regional added value by Industries, also estimating the quarterly GDP for the region in question.

© 2017 Hogrefe Publishing

Methodology (2017), 13(4), 123–134
Two-Way Classified System of Series

Now we have a two-way system, the target being \( R \times C \) series, \( x_{ij} \), \( i = 1, \ldots, R, j = 1, \ldots, C \).

We solve four cases in line with the available information. Some can be reduced to previous cases, except Case 5.

Case 5: Estimation of the Two-Way System, With Known Marginals

Objective: Estimation of the \( R \times C \) quarterly series, \( x_{ij} \), \( i = 1, \ldots, R, j = 1, \ldots, C \).

Data: We know the annual corresponding series, \( y_{ij} \), \( i = 1, \ldots, R, j = 1, \ldots, C \), and the quarterly marginal series, \( x_{1i}, \ldots, x_{Ri} \), and \( x_{i1}, \ldots, x_{iC} \).

Restrictions: In addition to the temporal constraints, we have \( R + C \) constraints,

\[
x_{ij} = \sum_{j=1}^{C} x_{ijt}, \quad t = 1, \ldots, n, \quad i = 1, \ldots, R \text{ and } j = 1, \ldots, C.
\]

Likelihood: It derives from the \( R \times C \) individual indicator-based models,

\[
x_{ijt} = \sum_{i=1}^{R} \delta_{ij} x_{ijt} + \epsilon_{ijt}, \quad i = 1, \ldots, R, \quad j = 1, \ldots, C.
\]

Now, letting \( x = (x'_{11}, \ldots, x'_{R-11}, \ldots, x'_{1(C-1)}, \ldots, x'_{R-1(C-1)}, x'_R C, \ldots, x'_{R1}, \ldots, x'_{R(C-1)} \) be the total column vector and \( x = (x'_1, \ldots, x'_{R-11}, \ldots, x'_{1(C-1)}, \ldots, x'_{R(C-1)} \) the “free” values vector, we can write \( x = Wx_r + y_r \), now being

\[
W = \left( \begin{array}{cc}
I_{R(C-1)}
& 0
& \otimes
& I_{R}
& 0
& \otimes
& I_r
\end{array} \right)
\]

and

\[
y_r = \left( y^r_{11}, \ldots, y^r_{R-11}, \ldots, y^r_{1(C-1)}, \ldots, y^r_{R(C-1)} \right)'
\]

where

\[
y^r_{v} = \left( y^r_{11}, \ldots, y^r_{(R-1)1}, \ldots, y^r_{1(C-1)}, \ldots, y^r_{R(C-1)} \right)'
\]

Case 6: Estimation of the Two-Way System, With Known Row Marginals

It is like Case 5, although this time round the only marginals known are the row marginals, \( x_{1i}, \ldots, x_{Ri} \). This case can be reduced to \( R \) consecutive executions of Case 1, once per row.

The same procedure applies when only the column marginals are known.

Example: Same as Case 5, although now we only know the national quarterly employment by economic sector.

Case 7: Estimation of the Two-Way System, Only the Total Sum Is Known

Row and column marginals are unknown, but the total, \( x_{ij} \), is known.

Case 7 is thus reduced to a single execution for Case 1, with \( R \times C \) variables.

Example: We estimate the two-way breakdown for the official quarterly national GDP by Industries and by regions.

Case 8: Estimation of the Two-Way System, Total Sum Is Estimated

Like Case 7, although now the total sum is unknown.

We have an additional econometric indicator-based model for the total,

\[
x_{ijt} = \sum_{i=1}^{R} \delta_{ij} z_{ijt} + \epsilon_{ijt}, \quad t = 1, \ldots, n.
\]

Observe that this case can be solved by applying Case 2 for the \( R \times C \) variables of the system.

Example: We estimate the two-way breakdown by Provinces (NUTS3) and Industries of the quarterly regional GDP, which is also estimated.