Time Series Reconciliation: A Least-Squares Solution for the Case of High-Frequency Series Exceeding the Period covered by the Low-Frequency Series

Luis Frank*

Abstract

The paper presents an alternative method - based on the least squares criterion - to reconciliate time series observed at different frequencies. The main advantage of this method - besides its simplicity to reconciliate one or more high frequency (e.g. monthly) series with a low frequency (e.g. quarterly) series - is that it enables the extrapolation of the reconciled monthly time series beyond the last available quarter, avoiding the usual practice of forecasting an additional quarter to achieve a soft reconciliation with the last months of the monthly series. It is worth noting however that the "reconciled" monthly values obtained with this procedure will not add up (or average) exactly to the corresponding values of the quarterly time series. The paper also compares the new procedure with other methods widely spread in official statistical bureaus using real data from the National Accounts of Argentina.

Key Words: Time series reconciliation, economic time series, official statistics, least squares criterion.

1. Introduction

Worldwide, the national statistics offices publish economic indicators of different frequency (usually monthly, quarterly and annual) to describe the evolution of certain macroeconomic aggregates.[1] In Argentina, for example, the National Accounts Office (INDEC) publishes the Monthly Economic Activity Estimator (EMAE), quarterly estimates of the Gross Domestic Product (GDP), an annual estimate of the GDP and a later review [5] of this last estimate. However, it is known that monthly, quarterly and annual estimates of the same macroeconomic aggregate hardly ever agree because each series comes from a different set of information sources. To force the match among series of different frequencies, various econometric methods - known as reconciliation methods - have been proposed (see [3] for a thorough review of these methods) which essentially distribute the discrepancies between the high-frequency series and the low-frequency series along a new synthetic high-frequency series that fits perfectly the given low-frequency series, under the assumption that the low-frequency series is the one that best represents the "true" evolution of the macroeconomic aggregate being studied. Such criterion is based on the reasonable assumption that low-frequency indicators are more accurate because they incorporate more information than their high-frequency counterparts. The declared goal of time series reconciliation is to avoid confusing the public with different figures for the same macroeconomic aggregate, even at the expense of transfering any measurement defect of the low-frequency series to the reconciled high-frequency series.

A review of the most popular reconciliation methods used in national accounting [12] reveals that most of them consider that there are one or more (observed) high-frequency series that may be explained by a single true but unknown low-frequency series, plus random noise. Such relationship may be written as a linear model subject to a set of linear

*Universidad de Buenos Aires. Facultad de Agronomía. Departamento de Métodos Cuantitativos y Sistemas de Información. Av. San Martin 4453 - C1417DSE. Buenos Aires, Argentina

[1] In this paper "frequency" refers only to the reporting period of the referred series.

constraints. For example, Denton's method [4], considers the model

$$\mathbf{D}\mathbf{z} = \mathbf{D}\boldsymbol{\upsilon} + \boldsymbol{\epsilon} \text{ subject to } (q\,\mathbf{P})\boldsymbol{\upsilon} = \mathbf{y},\ \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$$

where $\mathbf{D}$ is a differenciating matrix of -1 and 1, $\mathbf{z}$ is a single (observed) high-frequency series, $\boldsymbol{\upsilon}$ is the true (but unknown) high-frequency series - to be estimated - and $q\,\mathbf{P}$ is a zero-ones matrix that adds up the values of $\boldsymbol{\upsilon}$ in order to match the low-frequency series $\mathbf{y}$. Denton's model presents several undesirable features also shared by Chow-Lin's [1] and Fernández's [6, 7] models. First, these models assume that the low-frequency series $\mathbf{y}$ is observed without error, while $\mathbf{z}$ is observed with random noise. This assumption is crucial because it implies that any revision of the low-frequency series operates as a new model-specification, which in turn requires the computation of a new reconciled high-frequency series even if the observed high-frequency series did not change. Second, these methods are only useful for fitting high-frequency series to low-frequency series when the quantity of periods of the high-frequency series matches exactly that of the low-frequency series. In other words, these methods prevent the reporting of reconciled values in real-time. Third, these models reverse the true causal relationship between high and low-frequency series according to which the low-frequency series should result from the aggregation (adding or averaging) of high-frequency series instead of the disaggregation of low-frequency series. As presented, these models assume that the knowledge of the low-frequency series is *prior* to the observation of the high-frequency series. Fourth, by establishing an exact (linear) relationship between the true high-frequency and the low-frequency series, it assumes that the latter is the simple aggregation of the (unobserved) high-frequency series, although in practice it is sometimes verified that the observed high-frequency series has a completely different origin than the low-frequency series. Fifth, these methods are useful if and only if there is at least one high-frequency series is observable. However, if not a single high-frequency series is observable but only some high-frequency "stylized facts" are known, the model specification becomes doubtful and other methods (known as benchmarking methods) ought to be used instead.

2. Objectives

The aim of this paper is fourfold. First, we wish to develop a method for time series reconciliation computable in real time, that is, each time a new monthly datum becomes available, particularly after the end of the last available quarter. Second, we want the reconciled series not to show appreciable changes backward each time a new quarter is added. Third, we require the method to be simple, computable with standard statistical software. Four, we want to preserve the reconciled high-frequency series as a separate series, not necessarily perfectly fitting the quarterly series, in order to make clear to the public that the low-frequency series is not a straightforward aggregate of the high-frequency series. Regarding this last point recall that most statistical offices only publish the reconciled high-frequency series, so that the user can not keep track of the original (although revised) figures of the original high-frequency series.

3. The Proposed Method

Consider a low-frequency time series $\mathbf{y}$ (e.g. an annual series) of m periods and a set of $k-1$ related high-frequency series (e.g. quarterly series) arranged in a $n \times k$ matrix $\mathbf{Z}$ whose first column is $\mathbf{1}_n$.[2] Both $\mathbf{y}$ and $\mathbf{Z}$ represent the same underlying data generating

[2] $\mathbf{Z}$ may also include series of "stylized facts" or stationarity patterns known from the economic theory.

process, so two models may be addressed to relate them,

$$\mathbf{Hy} = \mathbf{Z}\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1,\ \boldsymbol{\epsilon}_1 \sim N\left(\mathbf{0}, \sigma^2\boldsymbol{\Omega}_1\right) \tag{1}$$

and

$$\mathbf{y} = \mathbf{PZ}\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2,\ \boldsymbol{\epsilon}_2 \sim N\left(\mathbf{0}, \sigma^2\boldsymbol{\Omega}_2\right), \tag{2}$$

where $\mathbf{H} = \mathbf{I}_m \otimes \mathbf{1}_q$, $\mathbf{P} = \mathbf{I}_m \otimes \mathbf{1}_q'/q$, and $q = n/m$.[3] As usual, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $k \times 1$ unknown vectors and $\boldsymbol{\epsilon}_2 = \mathbf{P}\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_1$ are unobserved random errors, presumably identically distributed although not independent. Under this specification, $\sigma^2\boldsymbol{\Omega}_2 = \sigma^2\mathbf{P}\boldsymbol{\Omega}_1\mathbf{P}'$, while $\boldsymbol{\Omega}_1$ is an unknown matrix. We shall assume, however, that $\boldsymbol{\Omega}_1$ is a symmetric and positive definite matrix, and is therefore invertible.

In the first model, $\mathbf{H}$ is a matrix that "expands" the low-frequency series $\mathbf{y}$ up to the length of the high frequency series simply by repeating q times each element of $\mathbf{y}$. In the second model, $\mathbf{P}$ is a matrix that "shrinks" the high-frequency series in $\mathbf{Z}$ by averaging the elements that correspond to the same low-frequency period. Matrix $\mathbf{H}$ is related to $\mathbf{P}$ by the identities $\mathbf{H}' = q\,\mathbf{P}$ and, of course, by $\mathbf{PH} = \mathbf{I}_m$. Although the first model may appear a bit rough compared to the second it cannot be ignored as a possible representation of the (linear) relationship between $\mathbf{Z}$ and $\mathbf{y}$. To these models we may add two equations to account for *prior* estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in order to to facilitate the estimation of these parameters in case the time series to be reconciled have very few observations.

$$\tilde{\mathbf{b}}_1 = \boldsymbol{\beta}_1 + \boldsymbol{\nu}_1 \quad \text{and} \quad \tilde{\mathbf{b}}_2 = \boldsymbol{\beta}_2 + \boldsymbol{\nu}_2, \tag{3}$$

where $\boldsymbol{\nu}_1 \sim N(\mathbf{0}, \sigma_\nu^2\boldsymbol{\Psi}_1)$, $\boldsymbol{\nu}_2 \sim N(\mathbf{0}, \sigma_\nu^2\boldsymbol{\Psi}_2)$ and $cov(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \mathbf{0}$. Besides, we impose a linear stochastic constraint on the estimatores $\mathbf{b}_1$ and $\mathbf{b}_2$ to guarantee that the estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ will be as close as possible.

$$\mathbf{b}_2 - \mathbf{b}_1 = (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1) + (\boldsymbol{\nu}_2 - \boldsymbol{\nu}_1). \tag{4}$$

The complete set of equations may be understood in a seemingly unrelated regressions (SUR) framework with *prior* information. The reader should not understand that $\tilde{\mathbf{b}}_1$ and $\tilde{\mathbf{b}}_2$ are *prior* estimates in a temporal sense, although past estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ can be used as *priors*.

3.1 The Least-Squares Solution

To proceed with the parameter estimation we write the error sums of squares function $L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$

$$\begin{aligned} L = {} & \frac{1}{\sigma^2}\,\boldsymbol{\epsilon}_1'\boldsymbol{\Omega}_1^{-1}\boldsymbol{\epsilon}_1 + \frac{1}{\sigma^2}\,\boldsymbol{\epsilon}_2'\boldsymbol{\Omega}_2^{-1}\boldsymbol{\epsilon}_2 + \frac{1}{\sigma_\nu^2}\,\boldsymbol{\nu}_1'\boldsymbol{\Psi}_1^{-1}\boldsymbol{\nu}_1 + \frac{1}{\sigma_\nu^2}\,\boldsymbol{\nu}_2'\boldsymbol{\Psi}_2^{-1}\boldsymbol{\nu}_2 + \\ & + \frac{1}{\sigma_\nu^2}(\boldsymbol{\nu}_2 - \boldsymbol{\nu}_1)'\mathbf{W}^{-1}(\boldsymbol{\nu}_2 - \boldsymbol{\nu}_1), \end{aligned}$$

where $\mathbf{W}$ is the covariance matrix of the random vector $\mathbf{b}_2 - \mathbf{b}_1$. It is not necessary to give further details about $\mathbf{W}$ (besides being invertible) since the reader will soon realize that the

[3] We'll assume that $\mathbf{Z}$ has exactly the same number of observations in both models, although this is not strictly necesary

term containing $\mathbf{W}^{-1}$ vanishes after deriving the first-order conditions. Then, the first order conditions for this minimization problem $\min\{L\}$ are, respectively,

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{b}_1} &= -2\,\frac{1}{\sigma^2}\,\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{H}\mathbf{y} + 2\,\frac{1}{\sigma^2}\,\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\mathbf{b}_1 + 2\,\frac{1}{\sigma_\nu^2}\boldsymbol{\Psi}_1^{-1}\left(\mathbf{b}_1 - \tilde{\mathbf{b}}_1\right) = \mathbf{0}\\ \frac{\partial L}{\partial \mathbf{b}_2} &= -2\,\frac{1}{\sigma^2}\,\mathbf{Z}'\mathbf{P}'\boldsymbol{\Omega}_2^{-1}\mathbf{y} + 2\,\frac{1}{\sigma^2}\,\mathbf{Z}'\mathbf{P}'\boldsymbol{\Omega}_2^{-1}\mathbf{P}\mathbf{Z}\mathbf{b}_2 + 2\,\frac{1}{\sigma_\nu^2}\boldsymbol{\Psi}_2^{-1}\left(\mathbf{b}_2 - \tilde{\mathbf{b}}_2\right) = \mathbf{0},\end{aligned}$$

which yield the linear system

$$\begin{bmatrix} \boldsymbol{\Psi}_1^{-1} + \alpha\,\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z} & \mathbf{0}_k \\ \mathbf{0}_k & \boldsymbol{\Psi}_2^{-1} + \alpha\,\mathbf{Z}'\mathbf{P}'\boldsymbol{\Omega}_2^{-1}\mathbf{P}\mathbf{Z} \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} \alpha\,\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{H}\mathbf{y} + \boldsymbol{\Psi}_1^{-1}\tilde{\mathbf{b}}_1 \\ \alpha\,\mathbf{Z}'\mathbf{P}'\boldsymbol{\Omega}_2^{-1}\mathbf{y} + \boldsymbol{\Psi}_2^{-1}\tilde{\mathbf{b}}_2 \end{bmatrix}.$$

where $\alpha = \sigma_\nu^2/\sigma^2$. From the expression above it's easy to see that $\mathbf{b}_1$ is a linear combination of the well known generalized least squares estimator (GLS) of $\boldsymbol{\beta}_1$ and the *prior* estimator $\tilde{\mathbf{b}}_1$. Operating conveniently, the solution for $\mathbf{b}_1$ is

$$\begin{aligned}\mathbf{b}_1|\mathbf{Z},\tilde{\mathbf{b}}_1 &= \left(\boldsymbol{\Psi}_1^{-1} + \alpha\,\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\left(\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\right)^{-1}\left(\alpha\,\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{H}\mathbf{y} + \boldsymbol{\Psi}_1^{-1}\tilde{\mathbf{b}}_1\right)\\ &= \left[\mathbf{I}_k + \frac{1}{\alpha}\left(\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\right)^{-1}\boldsymbol{\Psi}_1^{-1}\right]^{-1}\mathbf{b}_1{}^{\mathrm{GLS}} + \left(\mathbf{I}_k + \alpha\boldsymbol{\Psi}_1\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\right)^{-1}\tilde{\mathbf{b}}_1. \qquad (5)\end{aligned}$$

In the same fashion, but recalling that $\sigma^2\boldsymbol{\Omega}_2 = \sigma^2\mathbf{P}\boldsymbol{\Omega}_1\mathbf{P}'$,

$$\begin{aligned}\mathbf{b}_2|\mathbf{Z},\tilde{\mathbf{b}}_2 = &\left\{\mathbf{I}_k + \frac{1}{\alpha}\left[\mathbf{Z}'\mathbf{P}'\left(\mathbf{P}\boldsymbol{\Omega}_1\mathbf{P}'\right)^{-1}\mathbf{P}\mathbf{Z}\right]^{-1}\boldsymbol{\Psi}_2^{-1}\right\}^{-1}\mathbf{b}_2{}^{\mathrm{GLS}} + \\ &+ \left[\mathbf{I}_k + \alpha\,\boldsymbol{\Psi}_2\mathbf{Z}'\mathbf{P}'\left(\mathbf{P}\boldsymbol{\Omega}_1\mathbf{P}'\right)^{-1}\mathbf{P}\mathbf{Z}\right]^{-1}\tilde{\mathbf{b}}_2. \qquad (6)\end{aligned}$$

The reader may check that $\partial^2 L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$ is a positive definite matrix provided $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Psi}_1$ are also positive definite, so that the optimality conditions are fulfilled. Expressions (5) and (6) show that $\mathbf{b}$ may be interpreted as a weighted average of $\mathbf{b}^{\mathrm{GLS}}$ and $\tilde{\mathbf{b}}$. Moreover, adding up the two weighting matrices yields the identity matrix $\mathbf{I}_k$, in the same way scalar weights add up to unity in an index, as can be seen below

$$\left(\mathbf{I}_k + \frac{1}{\alpha}\,\mathbf{A}^{-1}\right)^{-1} + \left(\mathbf{I}_k + \alpha\,\mathbf{A}\right)^{-1} = \left(\mathbf{I}_k + \alpha\,\mathbf{A}\right)\left(\mathbf{I}_k + \alpha\,\mathbf{A}\right)^{-1} = \mathbf{I}_k,$$

where $\mathbf{A}$ is either $\boldsymbol{\Psi}_1\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}$ or $\boldsymbol{\Psi}_2\mathbf{Z}'\mathbf{P}'\boldsymbol{\Omega}_2^{-1}\mathbf{P}\mathbf{Z}$. For completeness, we next give the expression of the variance of $\mathbf{b}_1$. The reader may deduce $var(\mathbf{b}_2)$ by analogy.

$$\begin{aligned}var\left(\mathbf{b}_1|\mathbf{Z},\tilde{\mathbf{b}}_1\right) &= \sigma^2\left[\mathbf{I}_k + \frac{1}{\alpha}\,\left(\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\right)^{-1}\boldsymbol{\Psi}_1^{-1}\right]^{-1}\left(\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\right)^{-1}\left[\mathbf{I}_k + \frac{1}{\alpha}\,\boldsymbol{\Psi}_1^{-1}\left(\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\right)^{-1}\right]^{-1}\\ &= \sigma^2\left\{\left[\boldsymbol{\Psi}_1 + \frac{1}{\alpha}\,\left(\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\right)^{-1}\right]\left(\boldsymbol{\Psi}_1^{-1}\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\boldsymbol{\Psi}_1^{-1}\right)^{-1}\left[\boldsymbol{\Psi}_1 + \frac{1}{\alpha}\left(\mathbf{Z}'\boldsymbol{\Omega}_1^{-1}\mathbf{Z}\right)^{-1}\right]\right\}^{-1}. \qquad (7)\end{aligned}$$

Before moving to the next section it is relevant to make a brief digression to show how $\boldsymbol{\Psi}_1$ is related to $\boldsymbol{\Psi}_2$. First, note that if $\tilde{\mathbf{b}}_1$ is the GLS estimator of $\boldsymbol{\beta}_1$ computed from an earlier time series, then

$$\frac{1}{\sigma_\nu^2}\boldsymbol{\Psi}_1^{-1} = \frac{1}{\tilde{\sigma}^2}\left(\mathbf{Z}_0'\mathbf{U}'\right)\left(\mathbf{U}\mathbf{Z}_0\right) = \frac{1}{\tilde{\sigma}^2}\mathbf{T}'\mathbf{T}.$$

This is so because $\mathbf{\Omega}_1$ is a symmetrix positive-definite matrix as well as $\mathbf{\Omega}_1^{-1}$. Then, $\mathbf{\Omega}_1^{-1}$ may be decomposed into the product of an upper triangular marix by its traspose, that is $\mathbf{\Omega}_1^{-1} = \mathbf{U}'\mathbf{U}$, which is the well-known Cholesky decomposition.[4] Second, note that expanding conveniently the inverse of $\tilde{\sigma}^2\mathbf{\Psi}_2$ we can write

$$\begin{aligned}\frac{1}{\sigma_\nu^2}\mathbf{\Psi}_2^{-1} &= \frac{1}{\tilde{\sigma}^2}\left(\mathbf{Z}_0'\mathbf{U}'\right)\left[\mathbf{U}\mathbf{\Omega}_1\mathbf{P}'\left(\mathbf{P}\mathbf{\Omega}_1\mathbf{P}'\right)^{-1}\mathbf{P}\mathbf{\Omega}_1\mathbf{U}'\right]\left(\mathbf{U}\mathbf{Z}_0\right)\\ &= \frac{1}{\tilde{\sigma}^2}\mathbf{T}'\left\{\left(\mathbf{U}^{-1}\right)'\mathbf{P}'\left[\mathbf{P}\mathbf{U}^{-1}\left(\mathbf{U}^{-1}\right)'\mathbf{P}'\right]^{-1}\mathbf{P}\mathbf{U}^{-1}\right\}\mathbf{T}.\end{aligned}$$

Or, in a more compact way,

$$\frac{1}{\sigma_\nu^2}\mathbf{\Psi}_2^{-1} = \frac{1}{\tilde{\sigma}^2}\mathbf{T}'\left[\mathbf{V}\left(\mathbf{V}'\mathbf{V}\right)^{-}\mathbf{V}'\right]\mathbf{T},$$

where $\mathbf{V}' = \mathbf{P}\mathbf{U}^{-1}$ and the ordinary inverse of $\mathbf{V}'\mathbf{V}$ was replaced by a generalized inverse exploiting the fact that the linear projector between brackets is invariant to the chosen generalized inverse.

3.2 Methodological Background for estimating α, $\mathbf{\Omega}_1$ and $\tilde{\mathbf{b}}_1$

At this point, the reader is completely aware that our primary interest is to compute $\mathbf{b}_1$ for which we need first to estimate α, $\mathbf{\Omega}_1$ and get a *prior* estimate of $\boldsymbol{\beta}_1$. For this reason, the estimation of $\mathbf{b}_1$ must necessarily be performed in several stages, one to find $\mathbf{\Omega}_1$, another to get $\tilde{\mathbf{b}}_1$, a third one to estimate α, and a final step to estimate $\mathbf{b}_1$. However, to justify the procedure that we will follow at each stage we must first introduce some methodological background useful to work out the protocol.

Regarding the first step, recall that $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are symmetric positive definite matrices with Toeplitz-type structure, so the first column of each matrix contains all its distinct elements, n elements in case of $\mathbf{\Omega}_1$ and m in case of $\mathbf{\Omega}_2$.[5] Nevertheless, as the first element of each matrix is equal to 1, we got $m-1$ unknown elements of $\mathbf{\Omega}_2$ to be estimated and $n-1$ elements in case of $\mathbf{\Omega}_1$. So, even if $\mathbf{\Omega}_2$ were perfectly known it would be impossible to estimate the unknown $n-1$ elements of $\mathbf{\Omega}_1$. To circumvent this limitation, we propose to replace the unknown elements of $\mathbf{\Omega}_1$ by the vector that minimizes the overall sum of elements of the first column of $\mathbf{\Omega}_1$ subject to a set of linear constraints to guarantee that the relationship $\mathbf{\Omega}_2 = \mathbf{P}\mathbf{\Omega}_1\mathbf{P}'$ will hold. For example, if $\boldsymbol{\epsilon}_2$ followed an autorregressive process of order 1, with $0 < \rho < 1$, the first column of $\mathbf{\Omega}_1$ could be replaced by the solution to the linear programming problem

$$\min_{\mathbf{x}}\left\{\mathbf{1}_n'\mathbf{x}\right\} \quad \text{subject to} \quad x_1 = 1,\ \mathbf{A}_2\mathbf{x} = \mathbf{c}^*,\ \mathbf{A}_3\mathbf{x} \geq \mathbf{0}_n,\ \text{and } \mathbf{A}_4\mathbf{x} \geq \mathbf{0}_{n-1},$$

where the solution $\mathbf{x}$ is the solution that replaces the first column of $\mathbf{\Omega}_1$; $\mathbf{A}_2$ is a set of linear constraints that relates the first column of $\mathbf{\Omega}_1$ to the first column of $\mathbf{\Omega}_2$; $\mathbf{c}$ is the first column of $\mathbf{\Omega}_2$ multiplied by q^2 and $\mathbf{c}^*$ is the same as $\mathbf{c}$ except that the element $c_1^* = c_1/2$; $\mathbf{A}_3 = \mathbf{I}_n$; and $\mathbf{A}_4$ is a differencing matrix of 1 and -1, introduced to guarantee that $x_i - x_{i+1} > 0$. Under the given specification all the elements of $\mathbf{c}^*$ are positive. At this point, the covariance structure $\mathbf{\Omega}_1$ that arises from $\mathbf{x}$ is completely non-parametric. In the appendix

[4] $\mathbf{U}$ is a non-singular square matrix because all its diagonal elements are real positive numbers if and only if $\mathbf{\Omega}_1$ is a positive-definite matrix.

[5] In fact, both matrices may be decomposed as the product of a triangular matrix by its transpose (the well known Cholesky decomposition), i.e. $\mathbf{\Omega} = \mathbf{L}\mathbf{L}'$, where $\mathbf{L}$ is a lower triangular matrix, and the first column of $\mathbf{L}$ is exactly equal to the first column of $\mathbf{\Omega}$.

we show the system of linear constraints written more explicitly. If $\boldsymbol{\epsilon}_2$ followed and AR(1) process with $-1 < \rho < 0$ the elements of $\mathbf{c}^*$ would be alternately positive and negative, the even elements positive and the odd negative. In such case, if $\boldsymbol{\epsilon}_1$ also follows and AR(1) process with $\rho < 0$, $\mathbf{A}_3^-$ would be a $n \times n$ diagonal matrix with entries $(-1)^{|i-j|}$ and $\mathbf{A}_4^- = (\mathbf{R}\mathbf{A}_3\mathbf{R}')\mathbf{A}_4$ where $\mathbf{R} = [\mathbf{I}_{n-1}, \mathbf{0}_{(n-1)\times 1}]$. The problem to be solved may be written

$$\min_{\mathbf{x}} \left\{ \left(\mathbf{A}_3^- \mathbf{1}_n\right)' \mathbf{x} \right\} \quad \text{subject to} \quad x_1 = 1,\ \mathbf{A}_2\mathbf{x} = \mathbf{c}^*,\ \mathbf{A}_3^-\mathbf{x} \geq \mathbf{0}_n,\ \text{and}\ \mathbf{A}_4^-\mathbf{x} \geq \mathbf{0}_{n-1}.$$

The solutions to the linear programing problems addressed above are just solutions, not estimators of the elements of $\mathbf{\Omega}_1$, regardless the rationale of the assumptions that support them. Nevertheless, the solutions are unique although the reader should keep in mind that the covariance matrix arising from $\mathbf{x}$ might not be invertible. We will return to this point later.

Another issue related to the estimation of $\mathbf{b}_1$ is the availability of a *prior* estimate $\tilde{\mathbf{b}}_1$. To understand how to make available such an estimate, note that the limit when n tends to infinity for the given solutions $\mathbf{b}_1$ and $\mathbf{b}_2$ is

$$\lim_{n\to\infty} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \lim_{n\to\infty} \begin{bmatrix} \left(\mathbf{\Psi}_1^{-1}/n + \alpha\, \mathbf{Z}'\mathbf{\Omega}_1^{-1}\mathbf{Z}/n\right)^{-1} & \mathbf{0}_k/n \\ \mathbf{0}_k/n & \left(\mathbf{\Psi}_2^{-1}/n + \alpha\, \mathbf{Z}'\mathbf{P}'\mathbf{\Omega}_2^{-1}\mathbf{P}\mathbf{Z}/n\right)^{-1} \end{bmatrix} \times$$

$$\times \lim_{n\to\infty} \begin{bmatrix} \alpha\, \mathbf{Z}'\mathbf{\Omega}_1^{-1}\mathbf{H}\mathbf{y}/n + \mathbf{\Psi}_1^{-1}\tilde{\mathbf{b}}_1/n \\ \alpha\, \mathbf{Z}'\mathbf{P}'\mathbf{\Omega}_2^{-1}\mathbf{y}/n + \mathbf{\Psi}_2^{-1}\tilde{\mathbf{b}}_2/n \end{bmatrix} = \lim_{n\to\infty} \begin{bmatrix} \mathbf{b}_1{}^{\text{GLS}} \\ \mathbf{b}_2{}^{\text{GLS}} \end{bmatrix},$$

where we see that $\mathbf{b}_1$ and $\mathbf{b}_2$ approach the standard GLS estimator as the terms containing $\mathbf{\Psi}^{-1}$ and $\tilde{\mathbf{b}}$ vanish. Then, if long enough series of a previous "base year" were available, *prior* estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ could be obtained simply by computing the standard GLS estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ for those series, as long as $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ were also available. In practice, when working with index numbers, this may be achieved with series spaced enough to guarantee that the errors of the *prior* and current series are truly uncorrelated.

Another issue related to the computation of $\mathbf{b}_1$ is the estimation of α. To achieve this goal, two estimation strategies have been proposed (see [8, 7.4, p. 227]). One would be to start with an initial estimate of α, and iteratively compute $\mathbf{b}_1$, the vector of residuals $\mathbf{e}_1 = \mathbf{H}\mathbf{y} - \mathbf{Z}\mathbf{b}_1$, a new estimate of σ^2 and α, a new $\mathbf{b}_1$ and so on until stability is reached. The second strategy would be just a two step estimation of α departing from the GLS estimate of σ^2 and σ_ν^2 which is typically assumed to be known. A more recent alternative is that suggested by Liu et al. [10] to optimally estimate the product $w\,\alpha$, where w is a tuning factor introduced by Schaffrin and Toutenbourg [16] to account for the degree of belief on the *prior* estimate of the parameters. We shall not deepen into this issue as it is out of the scope of the paper.

3.3 The Time Series Reconciliation Protocol

The background given above allows us to propose a multistep procedure for estimating $\mathbf{b}_1$. The first step is to find $\mathbf{\Omega}_2$. We believe that a reasonable structure for $\mathbf{\Omega}_2$ would be an autocorrelated one, due to the fact that this kind of structure is a common feature in economic time series. To that end, we average the values of $\mathbf{Z}$ for each time period of $\mathbf{y}$ and estimate the parameters $\boldsymbol{\rho}$ of $\mathbf{\Omega}_2(\boldsymbol{\rho})$ following Cochrane-Orcutt's [2] iterative procedure. We could also estimate $\boldsymbol{\rho}$ by other methods (e.g. Durbin-Watson's method described in most econometric texts) but Cochrane-Orcutt's method has the advantage of returning $\mathbf{b}_2{}^{\text{GLS}}$ as a by-product. Following the same procedure, we can also compute $\tilde{\mathbf{b}}_2$ as long as a previous sample $\{\mathbf{Z}_0, \mathbf{y}_0\}$, were available.

Once we got $\tilde{\mathbf{b}}_2$ and $\mathbf{\Omega}_2(\hat{\boldsymbol{\rho}})$ we can either estimate $\mathbf{\Psi}_2$ and then α following the recursive procedure proposed by Theil and Goldberger [14] in the context of the so called "mixed" estimation, or move directly to the computation of $\widehat{\mathbf{\Omega}}_1$ which in turn allows the estimation of σ^2 and $\sigma_\nu^2\mathbf{\Psi}_1$. $\widehat{\mathbf{\Omega}}_1$ is the solution to the linear programing problem proposed above, and once it is available we are able to compute $\tilde{\mathbf{b}}_1$ in the same fashion as $\tilde{\mathbf{b}}_2$ and $\sigma_\nu^2\mathbf{\Psi}_1$. Finally, we compute $\mathbf{b}_1$ and the interpolated time series $\mathbf{Z}^{+}\mathbf{b}_1$ where the asterisk indicates that the high-frequency series extend up to the last available value. Below is the protocol just described (in a more straightforward version) and in the appendix the reader may find a computer code written in Euler Math Toolbox's matrix language to carry it out.

(1) Fit a linear regression of $\mathbf{PZ}$ on $\mathbf{y}$ following the Cochrane-Orcutt procedure and assemble the correlation matrix $\widehat{\mathbf{\Omega}}_2(\boldsymbol{\rho})$ using the estimated correlation coefficients in $\hat{\boldsymbol{\rho}}$.

(2) Solve the linear programming problem $\min\{\mathbf{1}_n'\mathbf{x}\}$ subject to $\mathbf{Ax} \geq \mathbf{d}$, where $\mathbf{A}$ and $\mathbf{d}$ refer to the set of constraints defined in the previous section. Normalize $\mathbf{x}$ so that $x_1 = 1$. Then compute the correlation coefficients ρ_h implicit in $\widetilde{\mathbf{\Omega}}_1(\mathbf{x})$. For instance, if $\mathbf{\Omega}_1$ is presumed to be an AR(1) structure and n is big enough, $\hat{\hat{\rho}}$ may be computed by equating $\mathbf{1}_n'\mathbf{x}$ to the upper bound $\mathbf{1}'\mathbf{x} \leq 1/(1-\hat{\hat{\rho}})$.

(3) Compute $\hat{\mathbf{b}}_1{}^{\text{GLS}}$ replacing $\mathbf{\Omega}_1$ by its proxy $\widehat{\mathbf{\Omega}}_1$ assembled from $\hat{\hat{\rho}}$ computed in the previous step. Also compute the "mean square error" (MSE) $s_1^2 = \mathbf{e}_1'\mathbf{e}_1/(n-k)$. To invert $\widehat{\mathbf{\Omega}}_1$ either resort to the close form of the inverse (only possible for simple covariance structures) or decompose $\widehat{\mathbf{\Omega}}_1$ in the SVD form and invert each singular value skipping those close to zero.

(4) Compute $\tilde{\mathbf{b}}_1$ and the MSE in the same way as in the previous step but from an earlier sample. Compute also $\hat{\alpha} \approx s_0^2/s_1^2$ and $\widehat{\mathbf{\Psi}}_1 \approx \hat{\alpha}^{-1}(\mathbf{Z}_0'\widehat{\mathbf{\Omega}}_1^{-1}\mathbf{Z}_0)^{-1}$ where the subscript 0 means that $\mathbf{Z}$ is the matrix of *prior* high-frequency series.

(5) Finally, compute $\hat{\mathbf{b}}_1$ using expression (5), and the sought reconciled series $\widehat{\mathbf{Hy}} = \mathbf{Z}^{+}\hat{\mathbf{b}}_1$, where $\mathbf{Z}^{+}$ stands for the low-frequency series up to the last available figure.

3.4 Extending the Procedure to Time-Varying Parameters

The reconciliation method described above satisfies our initial objectives, in particular regarding simplicity, real-time reconciliation and robustness to backward revisions. However, its simplicity also implies some assumptions that might appear too rigid, or even unrealistic, when dealing with real time series. One of them is the assumption that the parameters of the model that relates high-frequency to low-frequency series remain constant over time. Particularly for long time series the practitioner might find this assumption difficult to justify. To overcome this weakness we extend our five-step-procedure to the field of flexible estimation as developed by Kalaba and Tesfatsion [9] in the context of models with time-varying parameters. We shall not delve into the theory of flexible estimation (called Flexible Least-Squares) but rather bring it up to improve the aforementioned procedure. The reader can find a thorough explanation of the theory behind FLS in Kalaba and Tesfatsion's 1989 paper and in the bibliography cited therein, and a more concise explanation but extended to constrained estimation in appendix B. For our purposes, we shall simply extend Kalaba and Tesfatsion's original estimator by adding a term to the incompatibility cost function in order to incorporate a matrix $\widetilde{\mathbf{B}}_1$ of *prior* estimates, defined as

$$\text{vec}(\widetilde{\mathbf{B}}_1) = \text{vec}(\mathbf{B}_1) + \text{vec}(\boldsymbol{\nu}_1), \quad \text{vec}(\boldsymbol{\nu}_1) \sim N\left(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Psi}_1^{-1}\right)$$

where $\mathbf{I}_n \otimes \mathbf{\Psi}_1^{-1}$ is an $n \times k$ covariance structure. Although this is a general form which admits a *prior* estimate for each period, a more realistic situation would involve only one *prior* estimate from the past series, and only for $\boldsymbol{\beta}_1$. Anyway, rewriting the first order conditions for the extended incompatibility cost function (see appendix) we get

$$\frac{\partial C(\mathbf{B}_1, \mu, n)}{\partial \text{vec}(\widehat{\mathbf{B}}_1)} = 2\frac{1}{\sigma^2}\tilde{\mathbf{Z}}'\mathbf{\Omega}_1^{-1}\tilde{\mathbf{Z}}\,\text{vec}(\widehat{\mathbf{B}}_1) + 2\frac{\mu}{\sigma^2}\,\mathbf{D}'\mathbf{D}\,\text{vec}(\widehat{\mathbf{B}}_1) - 2\frac{1}{\sigma^2}\tilde{\mathbf{Z}}'\mathbf{\Omega}_1^{-1}\tilde{\mathbf{H}}\mathbf{y}$$
$$+ 2\frac{1}{\sigma_\nu^2}\left(\mathbf{I}_n \otimes \mathbf{\Psi}_1^{-1}\right)\left[\text{vec}(\widehat{\mathbf{B}}_1) - \text{vec}(\widetilde{\mathbf{B}}_1)\right] = \mathbf{0}.$$

Then, proceeding in the same fashion as in (5) we get

$$\text{vec}(\widehat{\mathbf{B}}_1) = \left[\mathbf{I}_{n\times k} + \frac{1}{\alpha}\left(\tilde{\mathbf{Z}}'\mathbf{\Omega}_1^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right)^{-1}\left(\mathbf{I}_n \otimes \mathbf{\Psi}_1^{-1}\right)\right]^{-1}\text{vec}(\widehat{\mathbf{B}}_1)^{\text{GLS}} +$$
$$+ \left[\mathbf{I}_{n\times k} + \alpha\left(\mathbf{I}_n \otimes \mathbf{\Psi}_1^{-1}\right)^{-1}\left(\tilde{\mathbf{Z}}'\mathbf{\Omega}_1^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right)\right]^{-1}\text{vec}(\widetilde{\mathbf{B}}_1). \tag{8}$$

where the weighting matrices add up to $\mathbf{I}_{n\times k}$. An alternative expression for $\text{vec}(\widehat{\mathbf{B}}_1)$, which under certain circumstances may be more desirable from a computational standpoint, is

$$\text{vec}(\widehat{\mathbf{B}}_1) = \left[\alpha\left(\tilde{\mathbf{Z}}'\mathbf{\Omega}_1^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right) + \left(\mathbf{I}_n \otimes \mathbf{\Psi}_1^{-1}\right)\right]^{-1} \times$$
$$\times \left[\alpha\,\tilde{\mathbf{Z}}'\mathbf{\Omega}_1^{-1}\mathbf{H}\mathbf{y} + \left(\mathbf{I}_n \otimes \mathbf{\Psi}_1^{-1}\right)\text{vec}(\widetilde{\mathbf{B}}_1)\right]. \tag{9}$$

So far, nothing was said about the tuning parameter μ. Kalaba and Tesfatsion did not attempt to estimate it in their original paper but rather keep it as a tuning parameter chosen by the practitioner. A naive criterion would be to set $\mu = 1$, that is, to give the same weight to the squared sum of errors and to the squared distances of consecutive vectors of estimated parameters. We will follow this criterion hereafter, but we will return to the point at the end of the paper.

4. Example: Reconciling Argentine's EMAE with the quarterly GDP

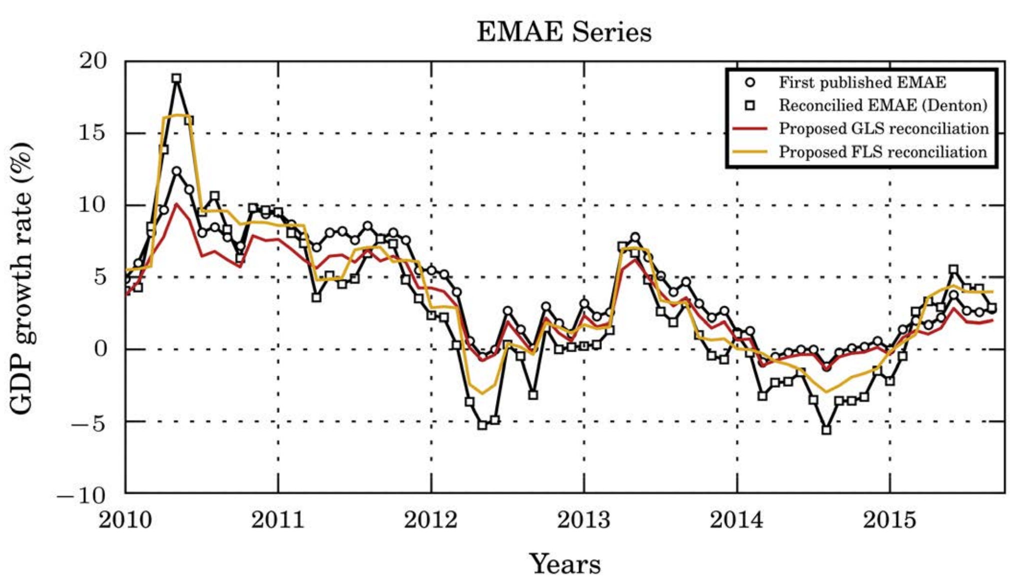
As an example, we reconcile next INDEC's Monthly Economic Activity Estimator's (EMAE, for its acronym in spanish) growth rates with the quarterly Gross Domestic Product (GDP) growth rate.[6] The EMAE (see Methodological Report) is a provisional index of the GDP published 50 to 60 days after the end of the reference month. It is a Laspeyres index that "tries to replicate the methods for calculating the quarterly and/or annual GDP, to the extent allowed by the availability of sources of information for a shorter time period." The index is constructed by aggregation of sectorial indeces that correspond to tabulation categories of the ISIC-3 classification. Because these indexes are provisional, the EMAE has to be revised to incorporate information missing in the first edition or corrections informed by the primary sources. EMAE's revision policy allows only two corrections of the first published value. However, at the end of the quarter, the series is reconciled with the quarterly GDP, which in turn is revised several times over the two years after the first publication. Therefore, the first published monthly growth rates remain available on INDEC's official website only for one or two months, after which they are replaced by corrected or reconciled figures.

[6]We use growth rates instead of the index value because INDEC only published growth rates between January 2017 and September 2015.

The figure below shows EMAE's first reported interannual growth rates, the quarterly GDP interannual growth rates reconciled with EMAE's last revised series (following Denton's method) and overlapped with them two series reconciled using the estimators (5) and (8). The period covered by both series (EMAE and GDP) is nearly five years, starting in January 2010 and ending in September 2015. We omit figures after September 2015 because they are still under review. The same time series but between January 2007 and December 2009 were used to compute the *prior* estimates $\tilde{\mathbf{b}}_1$ and $\widetilde{\mathbf{B}}_1$. The formula (8) assumes implicitly that each column vector of $\mathbf{B}_1$ is estimated from a different *prior*. Nevertheless, a more realistic possibility would be to use the last estimate of $\boldsymbol{\beta}_1$ from the previous period as the *prior* estimate of every column vector of $\mathbf{B}_1$, since it contains all the information available from the old series. Therefore, we decided to compute the reconciled series with the following slightly modified version of (8).

$$\widehat{\mathbf{Hy}} = \tilde{\mathbf{Z}}\left[\mathbf{I}_{n\times k} + \frac{1}{\alpha}\left(\tilde{\mathbf{Z}}'\boldsymbol{\Omega}_1^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right)^{-1}\left(\mathbf{I}_n \otimes \boldsymbol{\Psi}_1^{-1}\right)\right]^{-1} \text{vec}(\widehat{\mathbf{B}}_1)^{\text{GLS}} +$$
$$+\,\mathbf{Z}\left[\mathbf{I}_{n\times k} + \alpha\left(\mathbf{I}_n \otimes \boldsymbol{\Psi}_1^{-1}\right)^{-1}\left(\tilde{\mathbf{Z}}'\boldsymbol{\Omega}_1^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right)\right]^{-1}\tilde{\mathbf{b}}_1.$$

Simple inspection of the graph shows that the FLS series fits better the quarterly series than the GLS series, although both alternatives perform pretty well. The graph also shows that our procedure returns a softer monthly series and avoids spurious values at the end of the series that are a typical outcome of traditional reconciliation methods. Recall that the common practice to overcome this problem is to forecast the low-frequency series one period ahead and then reconcile the whole series as if all the figures were obtained by the same data generating process. This practice, however, also requires forecasts of monthly future values to match the period covered by the quarterly series. Then, the accuracy of the reconciliation procedure, at least at the end of the series, relies heavily on the method chosen to forcast future periods. This issue might obscure the whole reconciliation method.



5. Concluding Remarks

We present an alternative method to reconciliate time series observed at different frequencies. The main advantage of our method is that it enables the extrapolation of the reconciled monthly time series beyond the last available figure of the low frequency series, avoiding the usual practice of forecasting one period ahead to achieve a soft reconciliation with the last months of the monthly series. It is worth noting however that the reconciled high frequency values obtained with this method will not add up (or average) exactly to the corresponding values of the quarterly time series.

Some future research guidelines arise both from the theoretical development of the proposed method and its practical implementation in a typical National Accounts situation. First, it is crucial to develop an optimal weighting factor μ between error sums of squares and the sums of squared distances of the estimated parameters in the flexible reconciliation procedure. Second, it would be desible to combine models (1) and (2) in order to obtain a superior estimator than that developed in the paper. Third, it would be useful to extend the flexible estimator to the minimization of the squared distances between estimated parameters kept fixed over arbitrary time periods, in accordance with the revision policy of each statistical bureau. Another issue that has not yet been explored is the introduction of stylized facts in the reconciliation process, which is perfectly valid in the context of the method described above.

References

[1] Chow, G. and Lin, A. (1971). "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series" *Review of Economics and Statistics*, 53, 372-375.

[2] Cochrane, D. and Orcutt, G. H. (1949). "Application of least squares regression to relationships containing auto-correlated error terms," *Journal of the American Statistical Association*, 44(245), 32-61.

[3] Dagum, E. and Cholette, P. (2006). " Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series' 'Lecture Notes in Statistics. Spinger.

[4] Denton, F. (1971). "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization" *Journal of the American Statistical Association*, 66, 99-102.

[5] Dirección Nacional de Cuentas Nacionales (2016). Metodología del Estimador Mensual de Actividad Económica (EMAE). Metodología INDEC nro. 20. ISBN 978-950-896-480-9.

[6] Fernández, R.(1981a). "Estimación de indicadores económicos de corto plazo. Mensualización del Producto Bruto Industrial Argentino" CEMA nro. 28.

[7] Fernández, R.(1981b). "A Methodological Note on the Estimation of Time Series" *The Review of Economics and Statistics*, vol. 63, issue 3, pages 471-76.

[8] Johnston, J. (1972). "Econometric Methods" Second Edition. McGraw-Hill.

[9] Kalaba, R. and Tesfatsion, L. (1989). "Time-Varying Linear Regression via Flexible Least Squares" *Computers and Mathematics with Applications*, vol. 17 (8/9), pp. 1215-1245.

[10] Liu, C., Yang, H. and Wu J. (2013). "On the Weighted Mixed Almost Unbiased Ridge Estimator in Stochastic Restricted Linear Regression" *Journal of Applied Mathematics* Volume 2013, Article ID 902715.

[11] Mehta, J. S. and Swamy, P. A. V. B. (1970). "The Finite Sample Distribution of Theil's Mixed Regression Estimator and a Related Problem" *Review of the International Statistical Institute*, vol. 38.

[12] Ponce, J. (2004). "Una nota sobre empalme y conciliacin de series de cuentas nacionales" *Revista de Economía, Banco Central del Uruguay*, 11(2), 2-34.

[13] Swamy, P. A. V. B. and Mehta, J. S. (1969). "On Theil's Mixed Regression Estimator." *Journal of the American Statistical Association*, vol. 64,

[14] Theil, H. and Goldberger, A. S. (1961). "On Pure and Mixed Statistical Estimation in Economics" *International Economic Review*, vol. 2.

[15] Theil, H. (1963) "On the Use of Incomplete Prior Information in Regression Analysis" *Journal of the American Statistical Association*, vol. 58.

[16] Schaffrin, B. and Toutenburg H. (1990). "Weighted mixed regression" *Zeitschrift fur Angewandte Mathematik und Mechanik*, 70, 735-738.

A. System of Linear Constraints of the Linear Programing Problem used to compute $\mathbf{\Omega}_1$

We show next the set of linear constraints proposed to compute $\mathbf{\Omega}_1$. Note that the first row of the left hand side matrix sets $x_1 = 1$, the next $m + n$ rows correspond to the equality constraint $\mathbf{A}_2\mathbf{x} = \mathbf{c}^*$ and the inequality constraint $\mathbf{A}_3\mathbf{x} \geq 0$, and the last $n - 1$ rows correspond to the system $\mathbf{A}_4\mathbf{x} \geq \mathbf{0}$. The first row of the subset $\mathbf{A}_3\mathbf{x} \geq 0$, however, is superfluous because the constraint is implicit in the equality $x_1 = 1$, although we retain the inequality for didactical reasons. [7]

$$\begin{bmatrix} q/2 & q-1 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & q-1 & q & q-1 & \dots & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & q-1 & q & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 2 & 1 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 \\ 1 & -1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_{m+1} \\ x_{m+2} \\ \vdots \\ x_{i-1} \\ x_i \\ x_{i+1} \\ \vdots \\ x_n \end{bmatrix} \begin{matrix} = \\ = \\ = \\ = \\ \vdots \\ = \\ \geq \\ \vdots \\ \geq \\ \geq \\ \geq \\ \vdots \\ \geq \end{matrix} \begin{bmatrix} c_1/2 \\ c_2 \\ c_3 \\ \vdots \\ c_m \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

B. Flexible Least Squares Estimation

Under the FLS criterion the parameters of the model vary along observations. This means that, given a set of regressors and observations $\{\mathbf{X}, \mathbf{y}\}$, the underlying model may be writen as

$$\mathbf{y} = \tilde{\mathbf{X}}\,\mathrm{vec}(\mathbf{B}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}_n\right)$$

[7] The second element of the right hand side vector is equal to $q^2/2$ as the first element of $\mathbf{\Omega}_2$ (and all its diagonal elements) is expected to be equal to 1.

where $\mathbf{y}$ is the usual $n \times 1$ vector of observations, $\tilde{\mathbf{X}}$ is a $n \times (n \times k)$ block diagonal matrix arranged as shown below, vec($\mathbf{B}$) stands for the $(n \times k) \times 1$ vectorized matrix of parameters (see below) and $\boldsymbol{\epsilon}$ is the usual error term of normal i.i.d random variables. Then,

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \mathbf{x}_i & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{x}_n \end{bmatrix} \quad \text{and} \quad \text{vec}\,(\mathbf{B}) = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_i \\ \vdots \\ \boldsymbol{\beta}_n \end{bmatrix}.$$

The function to be minimized, which Kalab and Tesfatsion (1989) called "incompatibility cost function", is

$$C(\boldsymbol{\beta}, \mu, n) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i \boldsymbol{\beta}_i)^2 + \mu \sum_{i=1}^{n-1} \left(\boldsymbol{\beta}_{i+1} - \boldsymbol{\beta}_i\right)' \left(\boldsymbol{\beta}_{i+1} - \boldsymbol{\beta}_i\right).$$

However, a much easier and straight forward way of writing $C(\boldsymbol{\beta}, \mu, n)$ in matrix notation would be

$$\begin{aligned} C(\mathbf{B}, \mu, n) &= \left[\mathbf{y} - \tilde{\mathbf{X}}\,\text{vec}\,(\mathbf{B})\right]' \left[\mathbf{y} - \tilde{\mathbf{X}}\,\text{vec}\,(\mathbf{B})\right] + \mu\,\text{vec}\,(\mathbf{B})'\,\mathbf{D}'\mathbf{D}\,\text{vec}\,(\mathbf{B}) \\ &= \mathbf{y}'\mathbf{y} - 2\,\mathbf{y}'\tilde{\mathbf{X}}\,\text{vec}\,(\mathbf{B}) + \text{vec}\,(\mathbf{B})' \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right) \text{vec}\,(\mathbf{B}). \end{aligned} \tag{10}$$

$\mathbf{D}$ is a $(n-1)k \times (nk)$ differentiation matrix, so that $\mathbf{D}'\mathbf{D}$ is equal

$$\mathbf{D}'\mathbf{D} = \begin{bmatrix} -\mathbf{I} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{I} & -\mathbf{I} & \ddots & & \vdots \\ \mathbf{0} & \mathbf{I} & -\mathbf{I} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & & \ddots & \mathbf{I} & -\mathbf{I} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{I} \end{bmatrix} \times \begin{bmatrix} -\mathbf{I} & \mathbf{I} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} & \mathbf{I} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -\mathbf{I} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & -\mathbf{I} & \mathbf{I} \end{bmatrix}$$

where $\mathbf{I}$ is a $k \times k$ identity matrix.

$$\mathbf{D}'\mathbf{D} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ -\mathbf{I} & 2\,\mathbf{I} & -\mathbf{I} & \ddots & & \vdots \\ \mathbf{0} & -\mathbf{I} & 2\,\mathbf{I} & -\mathbf{I} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & & \ddots & -\mathbf{I} & 2\,\mathbf{I} & -\mathbf{I} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & -\mathbf{I} & \mathbf{I} \end{bmatrix}$$

Deriving the incompatibility cost function and equating to $\mathbf{0}$ we get the so-called "normal" equations and the least squares solution for vec($\mathbf{B}$) which will be unique if and only if $\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right)$ is a full rank matrix.

$$\text{vec}(\widehat{\mathbf{B}})^{\text{OLS}} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right)^{-1} \tilde{\mathbf{X}}'\mathbf{y}. \tag{11}$$

B.1 Constrained Flexible Least Squares

Under certain circumstances it may be desirable to keep the parameters' estimates constant over an arbitrary period of time. Such circumstances may arise when the revision policy prevents any change in the model specification within a certain time span (tipicaly an year) or when some robustness on the parameters' estimation is desired. To achieve this goal the incompatibility cost function may be constrained by a set of linear equations of the type $\mathbf{R}\text{vec}(\mathbf{B}) = \mathbf{r}$. For instance, if the estimated parameters are required to remain constant over two consecutive periods, the matrices of constraints would be

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & -\mathbf{I} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & & & & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \dots & \dots & \dots & \mathbf{I} & -\mathbf{I} \end{bmatrix} \text{ and } \mathbf{r} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}.$$

where $\mathbf{R}$ is a $(n-1)k \times nk$ and $\mathbf{r}$ is an $(n-1)k \times 1$ zero column vector. Then, the constrained incompatibility cost function is

$$C^* = \left[\mathbf{y} - \tilde{\mathbf{X}}\,\text{vec}\,(\mathbf{B})\right]' \left[\mathbf{y} - \tilde{\mathbf{X}}\,\text{vec}\,(\mathbf{B})\right] + \mu\,\text{vec}\,(\mathbf{B})'\,\mathbf{D}'\mathbf{D}\,\text{vec}\,(\mathbf{B}) + 2\boldsymbol{\lambda}'\left[\mathbf{R}\,\text{vec}\,(\mathbf{B}) - \mathbf{r}\right]$$

which in turn leads to the system of normal equations

$$\begin{bmatrix} \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right) & \mathbf{R}' \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \text{vec}(\widehat{\mathbf{B}}) \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{X}}'\mathbf{y} \\ \boldsymbol{\lambda} \end{bmatrix},$$

or to the more explicit parameter-estimator

$$\begin{aligned} \text{vec}(\widehat{\mathbf{B}})^{\text{RLS}}|\mu = \text{vec}(\widehat{\mathbf{B}})^{\text{OLS}} - \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right)^{-1} \left[\mathbf{R}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{R}'\right]^{-1} \times \\ \times \left[\mathbf{R}\,\text{vec}(\widehat{\mathbf{B}})^{\text{OLS}} - \mathbf{r}\right]. \end{aligned} \tag{12}$$