Time Series Reconciliation through Flexible Least Squares Estimation

Luis Frank *

Abstract
The research improves the current flexible least squares procedure for time series reconciliation. This kind of estimation was already introduced by the author but assuming a priori knowledge of a key weighting parameter. The improved procedure allows the estimation of the weighting paramater from the sample data by means of the Newton-Raphson iterative method. Encouraging results arise from the reconciliation of the monthly growth rates of Argentina's Monthly Economic Activity Estimator (EMAE) and the quarterly growth rates of the Gross Domestic Product.

Key Words: Time series reconciliation, flexible least squares, EMAE, Gross Domestic Product, GDP, Argentina

1. Introduction

In a recent paper [2] the author proposed a linear estimator (based on the flexible least squares criterion) to reconcile time series of different frequencies, understanding by "reconciliation" a rescaling of the high-frequency series to match the low-frequency series as close as possible. Recall that in the bibliography (see e.g. [1]) reconciliation is understood as a procedure that leads to a perfect fit of a high-frequency series to a low-frequency series, assuming that the latter is observed without error while the former is just a rough proxy of the low-frequency series. Frank discussed the traditional view mainly for four reasons. First, because traditional reconciliation methods do not allow adjustment of high-frequency series in real time but only up to the last available figure of the low-frequency series. Second, because the common practice of government statistical offices in time series reconciliation hides to the public the corrections and updates done in the high-frequency series after the figures of the low-frequency series become available. Third, traditional reconciliation transfers the low-frequency errors to the high-frequency series instead of removing them. Fourth, reconciliation as performed in practice is not informative about the true relationship that links series of different frequencies. The reconciliation procedure proposed in [2] overcomes these drawbacks althoug it still requires that the practitioner knows some parameters in advance. In particular, the practitioner should know the parameter μ that weights the error sum of squares of high-frequency series and the sum of squared deviations of the parameters. The procedure also requires that the time period in which the parameters of the model remain constant be fixed in advance.

2. Objectives

The aim of this paper is to improve the reconciliation procedure proposed in 2017 regarding the knowledge of the aforementioned weighting constant μ. The improved estimator will be used to reconcile Argentina's Monthly Economic Activity Estimator (EMAE, for its acronym in spanish) with the quarterly Gross Domestic Product (GDP) series and the results will be compared with those obtained in 2017.

*Universidad of Buenos Aires. Facultad de Agronomía. Av. San Martín 4453, C1417DSE. Buenos Aires, Argentina.

3. Theoretical Background

Let us recall briefly Kalaba and Tesfatsion's flexible least squares (FLS) estimator [3] and Frank's estimator [2] for time series reconciliation. For ease of reading we shall write Kalaba and Tesfatsion's original formulas in matrix notation. Throughout the paper we shall use the following notation: matrices are writen in bold capitals, vectors in bold lower case, and scalars in italic; unless otherwise mentioned, all vectors are column vectors; greek letters are used for parameters; a tilde on a matrix means that the matrix has been reshaped in a way useful for FLS, as shown next.

3.1 Flexible Least Squares Estimation

Consider a vector of observations **y** and a set of explanatory variables **X** such that each y_i is related linearly to the corresponding row vector $\mathbf{x}_i'$. If the parameters of of such relationship vary along observations, the relationship may be writen as

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

where **y** is the usual $n \times 1$ vector of observations; $\tilde{\mathbf{X}}$ is a $n \times nk$ block diagonal matrix arranged as shown below; vec(**B**) stands for the $nk \times 1$ vectorized matrix of parameters and $\boldsymbol{\epsilon}$ is the usual error term of normal i.i.d random variables. Hereinafter we will use the tilde to refer to block diagonal matrices in which each block is a row of the matrix mentioned below the tilde. So,

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

To estimate **B** under the least squares criterion the function to be minimized, which Kalaba and Tesfatsion [3] called "incompatibility cost function", is

$$\begin{aligned} C(\mathbf{B}, \mu, n) &= \left[\mathbf{y} - \tilde{\mathbf{X}}\,\mathrm{vec}\left(\mathbf{B}\right)\right]' \left[\mathbf{y} - \tilde{\mathbf{X}}\,\mathrm{vec}\left(\mathbf{B}\right)\right] + \mu\,\mathrm{vec}\left(\mathbf{B}\right)' \mathbf{D}'\mathbf{D}\,\mathrm{vec}\left(\mathbf{B}\right) \\ &= \mathbf{y}'\mathbf{y} - 2\,\mathbf{y}'\tilde{\mathbf{X}}\,\mathrm{vec}\left(\mathbf{B}\right) + \mathrm{vec}\left(\mathbf{B}\right)' \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right)\,\mathrm{vec}\left(\mathbf{B}\right) \end{aligned} \tag{1}$$

where μ is a weighting constant and **D** is a $(n-1)k \times nk$ differentiation matrix so that $\mathbf{D}'\mathbf{D}$ has the form

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

and **I** is a $k \times k$ identity matrix. Although the incompatibility function was originally defined by Kalaba and Tesfatsion as a function of **B**, μ and n, in practice it is a function of **B** only and conditional on μ and n since the parameters and the sample size are supposed to be given. Then, by solving the first order conditions for the incompatibility cost function we get the so-called "normal" equations and the FLS 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. That is,

$$\mathrm{vec}(\widehat{\mathbf{B}})_{\mathrm{OLS}} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right)^{-1}\tilde{\mathbf{X}}'\mathbf{y}. \tag{2}$$

3.2 Fixing the FLS estimator to estimate μ

Let us return to the incompatibility cost function (1) and the solution to the first order condition which leads to the FLS estimator (2). This estimator is conditional on the parameter μ which up to now was supposed to be known. However, it is possible to estimate μ adding a second condition $\partial C(\mathbf{B},\mu)/\partial\hat{\mu} = 0$ in the following fashion.

$$C(\mathbf{B},\mu) = \mathbf{y}'\mathbf{y} - 2\,\mathbf{y}'\tilde{\mathbf{X}}\,\mathrm{vec}(\mathbf{B}) + \mathrm{vec}(\mathbf{B})'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right)\mathrm{vec}(\mathbf{B}),$$

Then, the first order condition for μ is

$$\frac{\partial C(\mathbf{B},\mu)}{\partial\hat{\mu}} = \mathrm{vec}(\hat{\mathbf{B}})'\,\frac{\partial \mathbf{A}(\mu)}{\partial\hat{\mu}}\,\mathrm{vec}(\hat{\mathbf{B}}) = 0,$$

where we called $\mathbf{A}(\mu) = \tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}$.[1] Replacing vec($\hat{\mathbf{B}}$) by the solution given in (2) we get the more explicit expression

$$\frac{\partial C(\mathbf{B},\mu)}{\partial\hat{\mu}} = \mathbf{y}'\tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \hat{\mu}\,\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{D}'\mathbf{D}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \hat{\mu}\,\mathbf{D}'\mathbf{D}\right)^{-1}\tilde{\mathbf{X}}'\mathbf{y} = 0. \tag{3}$$

Simple inspection of (3) suggests that (a) μ cannot be easily cleared because it is not related lineraly with $\mathbf{X}$, $\mathbf{D}$ and $\mathbf{y}$; and (b) the first order condition does not have a unique solution for μ. To show the last point let's call

$$\mathbf{u} = \mathbf{D}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \hat{\mu}\,\mathbf{D}'\mathbf{D}\right)^{-1}\tilde{\mathbf{X}}'\mathbf{y},$$

so that the first-order condition can be written as $\partial C/\partial\hat{\mu} = \mathbf{u}'\mathbf{u} = 0$. Without loss of generality let's also replace the matrix $\mathbf{D}$, which pre-multiplicates the right hand side of the equality, by the square matrix $\mathbf{D}^{*'} = [\mathbf{D}', \mathbf{0}']'$. Note that $\mathbf{D}^{*'}\mathbf{D}^{*} = \mathbf{D}'\mathbf{D}$ so that the condition $\mathbf{u}'\mathbf{u} = 0$ is still fullfilled. Then, excluding the trivial solution for $\tilde{\mathbf{X}}'\mathbf{y} = \mathbf{0}$, the condition is satisfied if $\mathbf{u} = \mathbf{0}$ and

$$|\mathbf{D}^{*}|\,\left|\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \hat{\mu}\,\mathbf{D}'\mathbf{D}\right)^{-1}\right| = \frac{|\mathbf{D}^{*}|}{|\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \hat{\mu}\,\mathbf{D}'\mathbf{D}|} = 0.$$

But as $|\mathbf{D}^{*}|$ is null, it is clear that $|\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \hat{\mu}\,\mathbf{D}'\mathbf{D}|$ need not be bounded. In such circumstance, there would not be an optimal $\hat{\mu}$ to find out. To overcome this drawback we propose a slight amendment to the original incompatibility cost function of Kalaba and Tesfatsion. The amendment proposed is a true weighted average of the error sum of squares and the squared differences among parameters. That is,

$$C(\mathbf{B},\mu) = (1-\mu)\,\mathbf{y}'\mathbf{y} - 2\,(1-\mu)\,\mathbf{y}'\tilde{\mathbf{X}}\,\mathrm{vec}(\mathbf{B}) + \mathrm{vec}(\mathbf{B})'\left[(1-\mu)\,\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right]\mathrm{vec}(\mathbf{B}).$$

[1]We called $\mathbf{A}$ to the matrix between parenthesis, although the reader should be aware that this matrix has nothing to do with other matrices called so elsewhere in the paper.

Then, optimizing C with respect to $\text{vec}(\mathbf{B})$ and μ results in the first order conditions

$$\frac{\partial C(\mathbf{B},\mu)}{\partial \text{vec}(\hat{\mathbf{B}})} = -2\,(1-\mu)\,\tilde{\mathbf{X}}'\mathbf{y} + 2\left[(1-\mu)\,\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mu\,\mathbf{D}'\mathbf{D}\right]\text{vec}(\hat{\mathbf{B}}) = \mathbf{0}$$

so that,

$$\text{vec}(\hat{\mathbf{B}}) = \left[(1-\hat{\mu})\,\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \hat{\mu}\,\mathbf{D}'\mathbf{D}\right]^{-1}\tilde{\mathbf{X}}'\mathbf{y}\,(1-\hat{\mu}) = \mathbf{A}^{-1}\tilde{\mathbf{X}}'\mathbf{y}\,(1-\hat{\mu}), \tag{4}$$

and

$$\frac{\partial C(\mathbf{B},\mu)}{\partial \hat{\mu}} = 2\,\mathbf{y}'\tilde{\mathbf{X}}\,\text{vec}(\hat{\mathbf{B}}) + \text{vec}(\hat{\mathbf{B}})'\left(-\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mathbf{D}'\mathbf{D}\right)\text{vec}(\hat{\mathbf{B}}) - \mathbf{y}'\mathbf{y} = 0.$$

However, the second condition can be manipulated as shown below to get a more friendly expression in order to compute the second derivative, whose usefulness will become apparent shortly.

$$\begin{aligned}\frac{\partial C(\mathbf{B},\mu)}{\partial \mu} &= 2\,(1-\mu)\,\mathbf{y}'\tilde{\mathbf{X}}\mathbf{A}^{-1}\tilde{\mathbf{X}}'\mathbf{y} + (1-\mu)^2\,\mathbf{y}'\tilde{\mathbf{X}}\mathbf{A}^{-1}\left(-\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mathbf{D}'\mathbf{D}\right)\mathbf{A}^{-1}\tilde{\mathbf{X}}'\mathbf{y} - \mathbf{y}'\mathbf{y}\\ &= (1-\mu)\,\mathbf{y}'\tilde{\mathbf{X}}\mathbf{A}^{-1}\left[2\mathbf{A} + (1-\mu)\,\left(-\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mathbf{D}'\mathbf{D}\right)\right]\mathbf{A}^{-1}\tilde{\mathbf{X}}'\mathbf{y} - \mathbf{y}'\mathbf{y}.\end{aligned}$$

Note that the matrix between brackets may be rewritten as

$$\begin{aligned}2\mathbf{A} + (1-\mu)\,\left(-\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + \mathbf{D}'\mathbf{D}\right) &= 2\,(1-\mu)\,\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + 2\,\mu\,\mathbf{D}'\mathbf{D} - (1-\mu)\,\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + (1-\mu)\,\mathbf{D}'\mathbf{D}\\ &= (1-\mu)\,\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + (1+\mu)\,\mathbf{D}'\mathbf{D},\end{aligned}$$

so

$$\frac{\partial C(\mathbf{B},\mu)}{\partial \hat{\mu}} = \mathbf{y}'\tilde{\mathbf{X}}\mathbf{A}^{-1}\left[(1-\hat{\mu})^2\,\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + (1-\hat{\mu}^2)\,\mathbf{D}'\mathbf{D}\right]\mathbf{A}^{-1}\tilde{\mathbf{X}}'\mathbf{y} - \mathbf{y}'\mathbf{y} = 0.$$

This expression is, however, non-linear with respect to μ (recall that $\mathbf{A}$ is also a function of μ), so we ought to solve it by a computational method such as Newton-Raphson's iterative procedure. The reader may find a brief explanation of this method in the appendix at the end of the paper. Then, in the context of the incompatibility cost function $C^*(\mathbf{B},\mu)$, the recursion relationship of Newton-Raphson's method may be restated as

$$\hat{\mu}_{(h+1)} = \hat{\mu}_{(h)} - \left[\frac{\partial^2 C^*(\mathbf{B},\mu)}{\partial \hat{\mu}^2_{(h)}}\right]^{-1}\frac{\partial C^*(\mathbf{B},\mu)}{\partial \hat{\mu}_{(h)}}, \tag{5}$$

where the superscript between parenthesis refers to the iteration number. To compute $\hat{\mu}$ in this fashion we must compute the second derivative of $C^*(\mathbf{B},\mu)$ with respect to μ. To that end, we call $\mathbf{C} = (1-\mu)^2\,\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + (1-\mu^2)\,\mathbf{D}'\mathbf{D}$. Then,

$$\frac{\partial^2 C(\mathbf{B},\mu)}{\partial \mu^2} = \mathbf{y}'\mathbf{X}\left[\frac{\partial \mathbf{A}^{-1}}{\partial \mu}\,\mathbf{C}\mathbf{A}^{-1} + \mathbf{A}^{-1}\left(\frac{\partial \mathbf{C}}{\partial \mu}\,\mathbf{A}^{-1} + \mathbf{C}\,\frac{\partial \mathbf{A}^{-1}}{\partial \mu}\right)\right]\mathbf{X}'\mathbf{y}.$$

The reader may check that $\partial\mathbf{C}/\partial\mu = -2\mathbf{A}$ and $\mathbf{C} = (1-\mu)(\mathbf{A}+\mathbf{D}'\mathbf{D})$. In the appendix we develop the derivative of $\mathbf{A}^{-1}$ with respect to μ. Replacing the derivatives involving $\mathbf{A}^{-1}$ and $\mathbf{C}$ and arranging the terms conveniently we get

$$\begin{aligned}\frac{\partial^2 C(\mathbf{B},\mu)}{\partial \mu^2} &= \mathbf{y}'\tilde{\mathbf{X}}\left[\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mu}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1} - \mathbf{A}^{-1}\left(2\,\mathbf{I}_{nk} + \mathbf{C}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mu}\mathbf{A}^{-1}\right)\right]\tilde{\mathbf{X}}'\mathbf{y}\\ &= \mathbf{y}'\tilde{\mathbf{X}}\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mu}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mu}\mathbf{A}^{-1} - 2\,\mathbf{A}^{-1}\right)\tilde{\mathbf{X}}'\mathbf{y}\end{aligned}$$

Note that every term between brackets is a symmetric matrix, so the expression above can be reduced to

$$\frac{\partial^2 C(\mathbf{B},\mu)}{\partial \mu^2} = 2\,\mathbf{y}'\tilde{\mathbf{X}}\mathbf{A}^{-1}\left(\frac{\partial \mathbf{A}}{\partial \mu}\mathbf{A}^{-1}\mathbf{C} - \mathbf{A}\right)\mathbf{A}^{-1}\tilde{\mathbf{X}}'\mathbf{y}. \tag{6}$$

Writing $\partial\mathbf{A}/\partial\mu$ in terms of $\tilde{\mathbf{X}}$ and $\mathbf{D}$, and $\mathbf{C}$ in terms of $\mathbf{A}$ and $\mathbf{D}$, (6) is

$$\frac{\partial^2 C(\mathbf{B},\mu)}{\partial \mu^2} = 2\,\mathbf{y}'\mathbf{X}\mathbf{A}^{-1}\left[(1-\mu)\,\left(\mathbf{D}'\mathbf{D} - \mathbf{X}'\mathbf{X}\right)\left(\mathbf{I}_{nk} + \mathbf{A}^{-1}\mathbf{D}'\mathbf{D}\right) - \mathbf{A}\right]\mathbf{A}^{-1}\mathbf{X}'\mathbf{y}.$$

We are now able to solve the first order condition for $\hat{\mu}$ by Newton-Raphson's recursion. Replacing the first and second derivatives in (5) results in

$$\hat{\mu}_{(h+1)} = \hat{\mu}_{(h)} - \frac{1}{2}\,\frac{\mathbf{y}'\tilde{\mathbf{X}}\mathbf{A}^{-1}\left[(1-\hat{\mu}_{(h)})^2\,\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + (1-\hat{\mu}^2_{(h)})\,\mathbf{D}'\mathbf{D}\right]\mathbf{A}^{-1}\tilde{\mathbf{X}}'\mathbf{y} - \mathbf{y}'\mathbf{y}}{\mathbf{y}'\tilde{\mathbf{X}}\mathbf{A}^{-1}\left[\left(1-\hat{\mu}_{(h)}\right)\,\left(\mathbf{D}'\mathbf{D} - \tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)\left(\mathbf{I}_{nk} + \mathbf{A}^{-1}\mathbf{D}'\mathbf{D}\right) - \mathbf{A}\right]\mathbf{A}^{-1}\tilde{\mathbf{X}}'\mathbf{y}}.$$

Although this expression looks quite cumbersome it may be rewritten in a more friendly fashion just calling $\mathbf{G}_1 = \tilde{\mathbf{X}}'\tilde{\mathbf{X}}$, $\mathbf{G}_2 = \mathbf{D}'\mathbf{D}$, $\mathbf{A} = (1-\mu)\,\mathbf{G}_1 + \mu\,\mathbf{G}_2$ and $\mathbf{d} = \mathbf{A}^{-1}\tilde{\mathbf{X}}'\mathbf{y}$. Then,

$$\hat{\mu}_{(h+1)} = \hat{\mu}_{(h)} - \frac{1}{2}\,\frac{\mathbf{d}'\mathbf{C}\mathbf{d} - \mathbf{y}'\mathbf{y}}{\mathbf{d}'\left[(\mathbf{G}_2 - \mathbf{G}_1)\mathbf{A}^{-1}\mathbf{C} - \mathbf{A}\right]\mathbf{d}} \tag{7}$$

3.3 Time Series Reconciliation through FLS

As already mentioned, Frank [2] proposed a FLS estimator for time series reconciliation. The goal was to fit one or more high-frequency series to a low-frequency series assuming that boths sets are observed with error. To that end, he proposed a data generating process expressible by two sets of equations.

$$\begin{aligned} \mathbf{H}\bar{\mathbf{y}} &= \tilde{\mathbf{Z}}\text{vec}(\mathbf{B}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Omega}), \quad \mathbf{H} = \mathbf{I}_m \otimes \mathbf{1}_q, \quad \bar{\mathbf{y}} = \mathbf{P}\mathbf{y} \\ \text{vec}(\widehat{\mathbf{B}}_0) &= \text{vec}(\mathbf{B}) + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim N\left[\mathbf{0}, \sigma^2_\nu\left(\mathbf{I}_n \otimes \boldsymbol{\Psi}\right)\right]. \end{aligned} \tag{8}$$

In the first set, the high-frequency series in $\mathbf{Z}$ are related linearly to the "expanded" version of the low-frequency series $\bar{\mathbf{y}}$ by fixed, but time-varying, parameters. The expanded low-frequency series $\mathbf{H}\bar{\mathbf{y}}$ is just a time series in which each observation appears repeated q times to match the lenghth of the high-frequency series. $\mathbf{H}$ is an "expansion matrix" while $\mathbf{P} = \mathbf{H}'/q$, is an $m \times mq$ matrix that averages the elements of the unobservable time series $\mathbf{y}$. $\tilde{\mathbf{Z}}$ is an $n \times nk$ matrix of high-frequency series with its rows placed blockwise, as in $\tilde{\mathbf{X}}$. Perhaps, a clearer way of defining the relationship between $\tilde{\mathbf{Z}}$ and $\mathbf{y}$ would be

$$\bar{\mathbf{y}} = \mathbf{P}\tilde{\mathbf{Z}}\,\text{vec}(\mathbf{B}) + \boldsymbol{\xi}, \quad \boldsymbol{\xi} \sim N\left(\mathbf{0}, \sigma^2\boldsymbol{\Phi}\right) \quad \text{and} \quad \boldsymbol{\Phi} = \mathbf{P}\boldsymbol{\Omega}\mathbf{P}'.$$

This relationship, however, is useless for reconciling time series in real time since the number of rows of $\tilde{\mathbf{Z}}$ must match the length of the low-frequency series, that is, $n = mq$. The first set of equations in (8) has a random error term whith unknown covariance $\sigma^2\boldsymbol{\Omega}$. We shall explain below how to estimate $\sigma^2\boldsymbol{\Omega}$. The second set introduces a prior estimate $\widehat{\mathbf{B}}_0$ of the true parameters of the first set. This estimate is supposed to be unbiased with a known block-diagonal covariance structure $\mathbf{I}_n \otimes \boldsymbol{\Psi}$. Under this specification the incompatibility cost function to be optimized is the extended version of Kalaba and Tesfatsion's original incompatibility function given next.

$$C^*(\mathbf{B}|\mu) = \left[\mathbf{H}\bar{\mathbf{y}} - \tilde{\mathbf{Z}}\,\text{vec}\,(\mathbf{B})\right]' \left(\sigma^2\boldsymbol{\Omega}\right)^{-1} \left[\mathbf{H}\bar{\mathbf{y}} - \tilde{\mathbf{Z}}\,\text{vec}\,(\mathbf{B})\right] + \frac{\mu}{\sigma^2}\,\text{vec}\,(\mathbf{B})'\,\mathbf{D}'\mathbf{D}\,\text{vec}\,(\mathbf{B}) + \\ + \frac{1}{\sigma_\nu^2}\,\left[\text{vec}(\widetilde{\mathbf{B}}) - \text{vec}\,(\mathbf{B})\right]' (\mathbf{I}_n \otimes \boldsymbol{\Psi})^{-1} \left[\text{vec}(\widehat{\mathbf{B}}_0) - \text{vec}\,(\mathbf{B})\right]. \tag{9}$$

The first order condition for minimizing C^* conditional to μ, σ^2, σ_ν^2 and $\widehat{\mathbf{B}}_0$ is

$$\frac{\partial C^*(\mathbf{B}|\mu)}{\partial \text{vec}(\widehat{\mathbf{B}})} = -2\,\frac{1}{\sigma^2}\,\tilde{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\mathbf{H}\bar{\mathbf{y}} + 2\,\frac{1}{\sigma^2}\,\tilde{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{Z}}\,\text{vec}(\widehat{\mathbf{B}}) + 2\,\frac{\mu}{\sigma^2}\,\mathbf{D}'\mathbf{D}\,\text{vec}(\widehat{\mathbf{B}}) + \\ + 2\,\frac{1}{\sigma_\nu^2}\,\left(\mathbf{I}_n \otimes \boldsymbol{\Psi}^{-1}\right)\left[\text{vec}(\widehat{\mathbf{B}}) - \text{vec}(\widehat{\mathbf{B}}_0)\right] = \mathbf{0}.$$

Rearranging terms, calling $\alpha = \sigma_\nu^2/\sigma^2$ and manipulating conveniently the expression above the solution for $\text{vec}(\widehat{\mathbf{B}})$ is

$$\text{vec}(\widehat{\mathbf{B}}) = \left[\mathbf{I}_{nk} + \frac{1}{\alpha}\left(\tilde{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right)^{-1}\left(\mathbf{I}_n \otimes \boldsymbol{\Psi}^{-1}\right)\right]^{-1} \text{vec}(\widehat{\mathbf{B}})_{\text{FGLS}} + \\ + \left[\mathbf{I}_{nk} + \alpha\left(\mathbf{I}_n \otimes \boldsymbol{\Psi}^{-1}\right)^{-1}\left(\tilde{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right)\right]^{-1} \text{vec}(\widehat{\mathbf{B}}_0). \tag{10}$$

where $\text{vec}(\widehat{\mathbf{B}})_{\text{FGLS}}$ is the generalized version of the FLS estimator given in (2).

$$\text{vec}(\widehat{\mathbf{B}})_{\text{FGLS}} = \left(\tilde{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right)^{-1}\tilde{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\mathbf{H}\bar{\mathbf{y}}.$$

Caution! If $\text{vec}(\widehat{\mathbf{B}})_{\text{FGLS}}$ were a true GLS estimator, the term containing μ should appear multiplied by α. However, the formula given above is correct because actually we introduced the matrix $\tilde{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}$ to reformulate $\text{vec}(\widehat{\mathbf{B}})$ as a function of $\text{vec}(\widehat{\mathbf{B}})_{\text{GLS}}$ and $\text{vec}(\widehat{\mathbf{B}}_0)$.

The reader may check that the weighting matrices that pre-multiply $\text{vec}(\widehat{\mathbf{B}})_{\text{FGLS}}$ and $\text{vec}(\widehat{\mathbf{B}}_0)$ add up to $\mathbf{I}_{nk}$ so that the current estimate might be interpreted as a weighted average of the estimate of **B** known before sampling and that computed from the sample. Frank [2] provided an alternative expression to compute the reconciled series $\widehat{\mathbf{H}\bar{\mathbf{y}}}$ when all the time periods share the same single prior $\hat{\mathbf{b}}_0$. The expression is as follows

$$\widehat{\mathbf{H}\bar{\mathbf{y}}}\,|\mu,\alpha = \tilde{\mathbf{Z}}^*\left[\mathbf{I}_{nk} + \frac{1}{\alpha}\left(\tilde{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right)^{-1}\left(\mathbf{I}_n \otimes \boldsymbol{\Psi}^{-1}\right)\right]^{-1} \text{vec}(\widehat{\mathbf{B}})_{\text{FGLS}} + \\ + \mathbf{Z}^*\left[\mathbf{I}_{nk} + \alpha\left(\mathbf{I}_n \otimes \boldsymbol{\Psi}^{-1}\right)^{-1}\left(\tilde{\mathbf{Z}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{Z}} + \mu\,\mathbf{D}'\mathbf{D}\right)\right]^{-1}\hat{\mathbf{b}}_0. \tag{11}$$

To make this estimation feasible it is necessary to replace $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$ and $\hat{\mathbf{b}}_0$ by proper *proxies*. In fact, Frank's paper focused on finding a good proxy for $\boldsymbol{\Omega}$. The one found is a Toeplitz matrix **W** whose first column **w** is the linear programing solution satisfying

$$\min_{\mathbf{w}}\left\{\mathbf{1}'\mathbf{w}\right\} \quad \text{subject to} \quad w_1 = 1,\ \mathbf{A}_1\mathbf{w} = q^2\left[\frac{\phi_1}{2},\, \boldsymbol{\phi}'_{i>1}\right]',\ \text{and } \mathbf{A}_2\mathbf{w} \geq \mathbf{0}_{2n-1},$$

where $\mathbf{A}_1$ is a set of m linear constraints (see appendix in [2]) that relates the elements of $\boldsymbol{\Omega}$ to those of a covariance matrix $\boldsymbol{\Phi}$ and $\mathbf{A}_2$ is a set of $2n-1$ constraints introduced to guarantee that each $w_i \geq 0$ and the difference $w_i - w_{i+1} \geq 0$, as are supposed to be the typical covariance structures with positive autocorrelation. The covariance structure that arises from $\mathbf{w}$ is completely non-parametric and can be plugged in directly in (10) if $\mathbf{\Omega}$ were non-singular. Anyway, to avoid numerical unstabilities if $\mathbf{W}$ were ill-conditioned, Frank assumed an AR(1) covariance structure for $\mathbf{\Omega}$ and used $\mathbf{w}$ to compute the single parameter of this structure, ρ.

The previous result is extensible to the FGLS estimator just calling $\tilde{\mathbf{X}} = \mathbf{\Omega}^{-\frac{1}{2}}\tilde{\mathbf{Z}}$ and $\mathbf{y} = \mathbf{\Omega}^{-\frac{1}{2}}\mathbf{H}\bar{\mathbf{y}}$. However, it should be pointed out that in the new incompatibility cost function the parameter μ has a different meaning than that in the original function. In fact, the relationship between μ in Kalaba and Tesfatsion's estimator and μ^* in (4) is exactly $\mu = \mu^*/(1-\mu^*)$, provided $\mu^* \neq 1$. So two courses of action may be followed to compute the reconcilied series $\widehat{\mathbf{H}\bar{\mathbf{y}}}$. One course would be to compute directly

$$\widehat{\mathbf{H}\bar{\mathbf{y}}}\,|\mathbf{\Omega},\alpha = \tilde{\mathbf{Z}}^* \left\{\mathbf{I}_{nk} + \frac{1}{\alpha}\left[(1-\hat{\mu}^*)\,\tilde{\mathbf{Z}}'\mathbf{\Omega}^{-1}\tilde{\mathbf{Z}} + \hat{\mu}^*\,\mathbf{D}'\mathbf{D}\right]^{-1}\left(\mathbf{I}_n \otimes \mathbf{\Psi}^{-1}\right)\right\}^{-1} \mathrm{vec}(\widehat{\mathbf{B}}) +$$
$$+\,\mathbf{Z}^* \left\{\mathbf{I}_{nk} + \alpha\left(\mathbf{I}_n \otimes \mathbf{\Psi}^{-1}\right)^{-1}\left[(1-\hat{\mu}^*)\,\tilde{\mathbf{Z}}'\mathbf{\Omega}^{-1}\tilde{\mathbf{Z}} + \hat{\mu}^*\,\mathbf{D}'\mathbf{D}\right]\right\}^{-1}\hat{\mathbf{b}}_0, \qquad (12)$$

where $\hat{\mu}^*$ is the estimation that arises from Newton-Raphson's iterative procedure, performed when computed $\mathrm{vec}(\widehat{\mathbf{B}})$. The other course of action would be to compute μ in a preliminar round, but assuming $\mathbf{\Omega} = \mathbf{I}$ and $\alpha = 1$, and then proceed in the usual way as if μ were known.

4. Example: Reconciling EMAE with the quarterly GDP

Next, we reconcile EMAE's growth rates with the quarterly GDP growth rates for the period ranging from January 2010 to September 2015. We set aside the period January 2007 to December 2009 to compute the *prior* estimates $\hat{\mathbf{b}}_0$. Note that the growth rates to be reconciled are EMAE's first published growth rates without any further revision, so the EMAE series used in the example that follows is not the version downloadable from INDEC's official homepage, but one compiled by the author from INDEC's press releases. Both the monthly and the quarterly series chosen to examplify the proposed reconciliation procedure are the same that those used in [2] to let the results be fully comparative.

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 the reconciled series computed according to the proposal done in [2] and the improved version given above. 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. So 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.

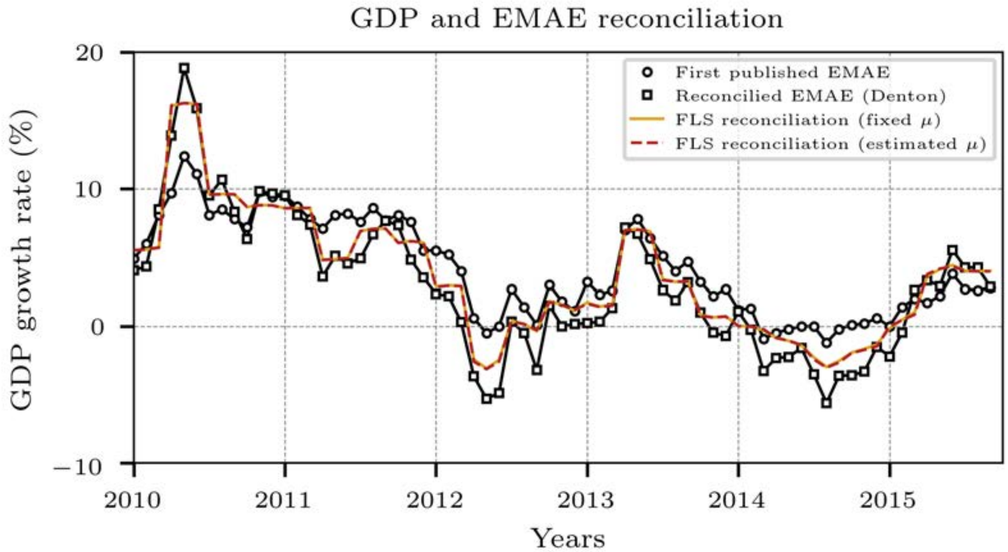

Figure 1: EMAE's first reported growth rates, quarterly GDP growth rates reconciled with EMAE's last revised series by Denton's method and FLS reconciliation.

5. Concluding remarks

In this paper we focused on the estimation of the parameter μ, a topic that seems to have been neglected in the literature referring to the FLS estimator. During the course of the investigation we noted that the so-called "incompatibility cost function" proposed by Kalaba and Tesfatsion cannot be optimized with respect to μ because the parametric space of μ is unbounded. Instead, we proposed a slightly modified version of the incompatibility cost function to bound μ in the interval $\mu \in (0, 1)$. We explicitly excluded the posibility that $\mu = 0$ or $\mu = 1$ because they lead to singular matrices $\mathbf{A}$, or degeneracy in the estimation of the type

$$\lim_{\mu \to 1} \mathrm{vec}(\mathbf{B}) = \lim_{\mu \to 1} \left[(1-\mu)\mathbf{X}'\mathbf{X} + \mu \mathbf{D}'\mathbf{D}\right]^{-1} \mathbf{X}'\mathbf{y}(1-\mu) = \mathbf{0}$$

The new reconciled series of quarterly and monthly growth rates did not show a major improvement with a previous series calculated keeping μ fixed in 1. However, it is yet premature to state that estimating μ from a sample does not improve substantially the reconciliation of time series computed by assuming that μ is fixed because the value of μ computed in our particular example was 0.4626, equivalent to 0.8608 in the traditional scale, which is pretty close to the widespread value $\mu = 1$ often used as a default value in flexible estimation.

Besides, we found that starting the Newton-Raphson algorithm with values close to 0 or 1, convergence cannot always be guaranteed. We attribute these failures, mostly, to the slowness of the algorithm and, in a lesser extent, to numerical instabilities that hindered the inversion of matrices $\mathbf{A}$. These latter cases could have been avoided by replacing problematic inverses of $\mathbf{A}$ with approximate inverses obtained by SVD-inversion. However, we prefered to invert all matrices by pure Gaussian elimination and back-substitution in order to detect this type of numerical drawbacks. When the initial values chosen were not close to 0 or 1, and $\delta < 0.001$, the convergence was reached in a few steps.

References

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

[2] Frank, L.(2017). "Time Series Reconciliation: A Least-Squares Solution for the Case of High-Frequency Series Exceeding the Period covered by the Low-Frequency Series" *2017 Proceedings of the American Statistical Association, Government Statistics Section, Alexandria, VA: American Statistical Association*: 3172-3184

[3] 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.

A. The Newton-Raphson Method

We recall Newthon-Raphson's recursive procedure for finding the roots of a real-valued function. To do so, recall first the Taylor decomposition of a function $f(\mathbf{x})$ in the neighborhood of a certain point $\mathbf{x}_m$.

$$f(\mathbf{x}) = f(\mathbf{x}_m) + (\mathbf{x} - \mathbf{x}_m)' \frac{\partial f(\mathbf{x}_m)}{\partial \mathbf{x}_m} + \frac{1}{2} (\mathbf{x} - \mathbf{x}_m)' \frac{\partial^2 f(\mathbf{x}_m)}{\partial \mathbf{x}_m \mathbf{x}_m'} (\mathbf{x} - \mathbf{x}_m) + \dots$$

In particular, at some point $\mathbf{x}_{m+1}$, the function $f(\mathbf{x})$ can be approximated by the first two terms of the Taylor series

$$f(\mathbf{x}_{m+1}) \approx f(\mathbf{x}_m) + (\mathbf{x}_{m+1} - \mathbf{x}_m)' \frac{\partial f(\mathbf{x}_m)}{\partial \mathbf{x}_m}$$

and, in case $f(\mathbf{x}_{m+1}) = 0$

$$\mathbf{x}_{m+1} \approx \mathbf{x}_m - \left[\frac{\partial f(\mathbf{x}_m)}{\partial \mathbf{x}_m} \right]^{-1} f(\mathbf{x}_m). \tag{13}$$

If $f(\mathbf{x})$ were a function to be optimized, for instance the errors sum of squares of a linear model, expression (13) is useful to find recursively the estimated parameters of the model. In terms of a standard regression (13) may be rewriten as

$$\mathbf{b}_{m+1} = \mathbf{b}_m - \left[\frac{\partial^2 L}{\partial \mathbf{b}_m \mathbf{b}_m'} \right]^{-1} \frac{\partial L}{\partial \mathbf{b}_m}.$$

where $\partial L / \partial \mathbf{b}_m$ is the first derivative of the error sum of squares or the log likelyhood function, and m is the iteration number.

B. Some results in matrix calculus

Recall that for any non-singular square matrix $\mathbf{A}$ there exists a unique matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. Then, if $\mathbf{A}$ is a function of μ, deriving this identity with respecto to μ on both sides of the equality yields,

$$\frac{\partial \mathbf{A}^{-1}}{\partial \mu} \mathbf{A} + \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mu} = \mathbf{0}, \quad \text{so that} \quad \frac{\partial \mathbf{A}^{-1}}{\partial \mu} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mu} \mathbf{A}^{-1}.$$