2/RT/04 April 2004

Research Technical Paper

A RATS subroutine to implement the Chow-Lin distribution/interpolation procedure

John C. Frain [1]

Economic Analysis and Research Department,
Central Bank and Financial Services Authority of Ireland,
P.O. Box 559, Dame Street,
Dublin 2,
Ireland
http://www.centralbank.ie

Abstract

This paper describes a RATS routine to implement the Chow Lin (1971) [2] procedure for the best linear unbiased distribution and interpolation of time series by related series. Various versions of this procedure have been used to distribute/interpolate annual to quarterly time series. One particular use of the routine described here has been to derive quarterly national accounts which have been used to estimate a quarterly macro model of the Irish economy and in various other research studies in the Bank. The paper is accompanied by sample program and data.

Contents

1   Introduction

This note describes a RATS subroutine which implements a version of the Chow & Lin (1971) [2] procedure for the "Best Linear Unbiased Distribution and Interpolation of Time Series by Related Series". Following this introduction there are two sections which deal with

2: A users guide to the use of the subroutine

3: A description of the distribution/interpolation method, and

A listing of the subroutine is given in an annexe.

Official quarterly national accounts are available in Ireland from 1997 and were first published in 1999. Events of interest to central banks are often of a short duration and their effects are not very obvious in annual data. For example in 1992/3 the size of the turbulence in the interest rate and exchange rate markets during the 1992/3 ERM adjustments appears as a slight blip in most annual data. Annual data do not capture a wide variety of effects of interest to a central bank. The Central Statistics Office have always published a wide range of monthly or quarterly indicators covering areas of interest in the Irish economy. A direct study of these indicators provides valuable intelligence on the state of the Irish economy and will always be used for that purpose.

The underlying behavioural relationships in the economy are expressed in terms of national accounts type aggregates. To estimate these relationships on a quarterly basis requires one to have historic estimates of the quarterly data. O'Reilly (1981) [5] estimated the first set of distributed Quarterly National Accounts for Ireland. These estimates were based on OLS regressions with the 'annual residuals' distributed according to the methodology of Boot et al (1967) [1]. This work had two features

- It took account of all relevant indicators

- The quarterly series was constrained to add to the annual

Updates to this work were based on first a partial implementation of the methodology of Chow & Lin (1971) [2]. The latest such estimates were computed to derive quarterly national accounts for use in updating the Irish section of the ESCB multi-country model (See McQuire and Ryan (2000) [4]).

2

The choice of indicators to use is pragmatic. In general one should be able to think of the relationship between a variable as a part/whole relationship and avoid behavioural relationships. In particular one must avoid the use of indicators which would lead to spurious results. For example, if one were to use retail sales to apportion consumption and then to use the same indicator to apportion disposable income, it would not be possible to determine whether any resulting relationship had come from the distribution/interpolation method or from an actual relationship between consumption and income. While this example is very obvious there may be many more subtle pitfalls awaiting the unwary

The aim of the exercise was to produce seasonally adjusted quarterly national accounts. To this end all indicator series were first seasonally adjusted using the TRAMO/SEATS [3] seasonal adjustment program. As expected the distributed series had no measurable seasonal variation.

The annual series distributed in the latest exercise were based on ESA 79 definitions. Official Irish quarterly national accounts are based on ESA 95. They are not compatible with the series derived here. Quarterly data produced using this methodology have been used in a variety of studies as well as in macro model building. The use of the data usually produces reasonable parameter values. One would not regard the derived short-term estimated dynamics as being as reliable as those derived from the analysis of primary quarterly data.

The version of the subroutine described here has been tested with a beta test Rats version 6, versions 5.11 and 5.04. It is not compatible with version 5.10. The sample series distributed with this paper are for illustration purposes only.

This routine is provided without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The entire risk as to the quality and performance of the program is with the user. Should the program prove defective the user assumes the cost of all necessary servicing, repair or correction.

2 A user's guide to the subroutine

The syntax of the subroutine call is

```
@y2q(options) yseries qseries start end
# <list of indicators in regression format>
```

Where

yseries is the series to interpolate. This should be set up as a quarterly series. y2q takes the last quarter of each year as the annual value to be interpolated. Reading an annual series from a `RATS` format file when the `CALENDAR` statement sets up a quarterly periodicity produces this set–up

qseries is the interpolated quarterly series

start end is the range to use in interpolation. These should refer to quarters 1 and 4 respectively.

The list of indicators may include constant and trend terms. Constant and trend terms may also be included as options in the procedure call. If they are included in the list of indicators, they will be used in the calculation of the Principal Component which is used to make the interpolating regression 'homoscedastic'. If they are specified in the option list they are added to the regression after the 'heteroscedasticity' adjustment has been completed. If there is a single indicator do not use the `PC` option in conjunction the `CONST` option. In this case specify a constant in your indicators and use the `NOCONST` option (if this is what you require).

The annual series should contain no missing data (#NA) over the interpolation range. All indicators must contain no #N/A from the first period of the start year to the final period of the end year of the interpolation. If an #N/A is found in either series the procedure halts and returns an error message. If `start` is not the first quarter of a year or `end` is not the fourth quarter the procedure also halts and returns an error message.

The options available are as follows:

PC [NOPC] Heteroscedasticity adjustment required using sole indicator or principal component if more than one indicator. The default is not to adjust.

CONST/[NOCONST] Add or do not add a constant to the regression after the 'homoscedastic' transformation. Default is not to add constant

TREND/[NOTREND] Add or do not add a trend to the regression after the 'homoscedastic' transformation. Default is not to add the trend

STARTRHO Initial value for ρ in maximum likelihood estimation. Default is 0.5. If the estimate does not converge for this value of ρ you may need to try a larger value. Some initial values may cause the program to exit with a message that some matrix is not invertible

The procedure includes an amended version of the principal components procedure distributed with RATS

Example

```
cal 1980 1 4
all 1999:4
open data C:\work\interpolation\tex\example\pcra.rat
data(format=rats) / pcra ;* annual series
data(format=rats) / ind_pcr;* quarterly indicator
set intercept = 1
source(noecho) C:\work\interpolation\tex\example\y2q.src
*
@y2q(startrho=0.95) pcra pcrq_nopc 1980:1 1999:4
#ind_pcr intercept
*
@y2q(pc,startrho=0.95,CONST) pcra pcrq_pc1 1980:1 1999:4
#ind_pcr
*
@y2q(pc,startrho=0.95) pcra pcrq_pc2 1980:1 1999:4
#ind_pcr intercept
```

The first six statements load the data and indicators. The source statement will load the subroutine into memory for further use. The statement

```
@y2q(startrho=0.95) pcra pcrq_nopc 1980:1 1999:4
#ind_pcr intercept
```

runs the procedure using as indicators a constant and `ind_pcr`. The program is run in default mode (no adjustments for heteroskedasticity). The run will not converge with the default quarterly autocorrelation (.5) and a starting value of 0.95 is used. This set of instructions gives rise to an interpolation procedure which is asymptotically equivalent to that used in earlier exercises. The statement

```
@y2q(pc,startrho=0.95,const) pcra pcrq_pc1 1980:1 1999:4
#ind_pcr
```

is similar but this time the heteroskedasticity adjustment `PC` is completed. The statement

```
@y2q(pc,startrho=0.95) pcra pcrq_pc2 1980:1 1999:4
#ind_pcr
```

carries out the heteroskedasticity adjustment using principal components of `ind_pcr` and `intercept`.

```
@y2q(pc,startrho=0.95) pcra pcrq_pc2 1980:1 1999:4
#ind_pcr intercept
```

will not work as the resulting regression will be multicollinear. If a second variable is added to the supplementary card

```
@y2q_2(pc,startrho=0.95,constant) pcra pcrq_const 1980:1 1996:4
#ind_pcr ind2_pcr
```

The `PC` adjustment is based on the two variables `ind_pcr` and `ind2_pcr`. A constant is then added to the regression. There are then three variables is in the regression - rescaled `ind_pcr` (rescaled), `ind2_pcr` (rescaled) and a constant.

The output produced by the program consists of

1. Progress of the maximization process used in estimating the autocorrelation coefficient. If the system fails to converge this may indicate a possible new starting value for ρ (The auto-correlation coefficient in the quarterly model). It may also imply that the indicator model is very bad and that the indicators are not indicating.

2. Estimates of the parameters of the model

3. Summary statistics of the residuals in the final GLS regression

4. Summary statistics of the distributed quarterly series

5. A printout of the interpolated quarterly series and it rate of change quarter on quarter and quarter on same quarter of the previous year

6. A graph giving rates of change of the original annual series and for the distributed series the rates of change quarter on quarter and quarter on same quarter of the previous year

3 A description of the distribution/interpolation method

This section describes the methodology of Chow and Lin (1971) and sets out how it has been adapted and implemented.

Let $\boldsymbol{y}$ be a $(4n \times 1)$ vector of the unknown quarterly series. Let $\boldsymbol{X}$ be a $(4n \times p)$ matrix of related series. Each column of this matrix represents a quarterly data series which will serve as an indicator for the distribution/interpolation of the annual series. We assume that there is a multivariate regression type of relationship between $\boldsymbol{y}$ and $\boldsymbol{X}$ of the form.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where

$$E[\boldsymbol{\varepsilon}] = 0 \text{ and } E[\boldsymbol{\varepsilon\varepsilon'}] = \boldsymbol{\Omega}$$

Let $\boldsymbol{C}$ be the $(n \times 4n)$ matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \ldots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 1 & 1 & 1 \end{pmatrix} \tag{2}$$

and let

$$\boldsymbol{y}_a = \boldsymbol{Cy} \text{ and } \boldsymbol{X}_a = \boldsymbol{CX}$$

Then $\boldsymbol{y}_a$ an $(n \times 1)$ vector of the known annual observations on the national accounts variable and $\boldsymbol{X}_a$ is an annualised version of the indicator variable matrix. Strictly speaking we are distributing an annual flow variable between quarters rather than providing an interpolated end quarter value for a stock variable. If interpolation of a stock variable or a price index is required the matrix $\boldsymbol{C}$ must be amended and replaced by

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \ldots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 1 \end{pmatrix}$$

Let $\boldsymbol{A}$ be matrix such that $\hat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{y}_a$ is a linear unbiased estimator of the quarterly series. Premultiplying equation (1) by $\boldsymbol{C}$ we get

$$
\begin{aligned}
\boldsymbol{y}_a &= \boldsymbol{C}\boldsymbol{y} \\
&= \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}\boldsymbol{\varepsilon} \\
&= \boldsymbol{X}_a\boldsymbol{\beta} + \varepsilon_a
\end{aligned}
\tag{3}
$$

where

$$
E[\varepsilon_a] = 0 \text{ and } E[\varepsilon_a\varepsilon_a{}'] = \boldsymbol{C}\boldsymbol{\Omega}\boldsymbol{C}'
$$

Thus

$$
\begin{aligned}
\hat{\boldsymbol{y}} - \boldsymbol{y} &= \boldsymbol{A}\boldsymbol{y}_a - \boldsymbol{y} \\
&= \boldsymbol{A}(\boldsymbol{X}_a\boldsymbol{\beta} + \varepsilon_a) - (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= (\boldsymbol{A}\boldsymbol{X}_a - \boldsymbol{X})\boldsymbol{\beta} + (\boldsymbol{A}\varepsilon_a - \boldsymbol{\varepsilon})
\end{aligned}
\tag{4}
$$

If $\hat{\boldsymbol{y}}$ is an unbiased estimator of $\boldsymbol{y}$ then $E[\hat{\boldsymbol{y}} - \boldsymbol{y}] = 0$. Taking expectations of equation (4) gives

$$
E[\hat{\boldsymbol{y}} - \boldsymbol{y}] = (\boldsymbol{A}\boldsymbol{X}_a - \boldsymbol{X})\boldsymbol{\beta} = 0
\tag{5}
$$

This represents $4n$ equations in the p unknowns. Thus we must have

$$
\boldsymbol{A}\boldsymbol{X}_a - \boldsymbol{X} = 0
\tag{6}
$$

The covariance matrix of $(\hat{\boldsymbol{y}} - \boldsymbol{y})$ is given by

$$
\begin{aligned}
cov(\hat{\boldsymbol{y}} - \boldsymbol{y}) &= E[(\boldsymbol{A}\varepsilon_a - \boldsymbol{\varepsilon})(\boldsymbol{A}\varepsilon_a - \boldsymbol{\varepsilon})'] \\
&= E[\boldsymbol{A}\varepsilon_a\varepsilon_a'\boldsymbol{A}' - \boldsymbol{\varepsilon}\varepsilon_a'\boldsymbol{A}' - \boldsymbol{A}\varepsilon_a\boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] \\
&= \boldsymbol{A}\boldsymbol{C}\boldsymbol{\Omega}\boldsymbol{C}'\boldsymbol{A}' - \boldsymbol{\Omega}\boldsymbol{C}'\boldsymbol{A}' - \boldsymbol{A}\boldsymbol{C}\boldsymbol{\Omega} + \boldsymbol{\Omega}
\end{aligned}
\tag{7}
$$

The diagonal elements of this covariance matrix represent the variance of the corresponding element of $(\hat{\boldsymbol{y}} - \boldsymbol{y})$. The sum of these elements or the trace of this covariance matrix is a multiple of the variance of the estimator. Thus the minimum variance unbiased estimator is that choice of $\boldsymbol{A}$ which minimises the trace of this covariance matrix subject to the constraints given by equation (6). We may write the Lagrangian for this constrained optimisation problem as—

$$
\mathcal{L} = \frac{1}{2}trace(\boldsymbol{A}\boldsymbol{C}\boldsymbol{\Omega}\boldsymbol{C}'\boldsymbol{A}' - \boldsymbol{\Omega}\boldsymbol{C}'\boldsymbol{A}' - \boldsymbol{A}\boldsymbol{C}\boldsymbol{\Omega} + \boldsymbol{\Omega}) - trace(\boldsymbol{M}'(\boldsymbol{A}\boldsymbol{X}_a - \boldsymbol{X}))
\tag{8}
$$

9

where $\boldsymbol{M}$ is a $(4n \times p)$ matrix of Lagrange multipliers.

Differentiating (8) with respect to the elements of A and equating the derivatives to zero for a minimum we get—

$$\boldsymbol{AC\Omega C'} - \boldsymbol{\Omega C} - \boldsymbol{MX}'_a = 0 \tag{9}$$

Thus

$$\boldsymbol{A} = \boldsymbol{\Omega C'(C\Omega C')}^{-1} + \boldsymbol{MX}'_a\boldsymbol{(C\Omega C')}^{-1} \tag{10}$$

Substitute (10) into (6) to get

$$\boldsymbol{\Omega C'(C\Omega C')}^{-1}\boldsymbol{X}_a + \boldsymbol{M}[\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{X}_a] - \boldsymbol{X} = 0 \tag{11}$$

Thus

$$\boldsymbol{M} = \boldsymbol{X}[\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{X}_a]^{-1} - \boldsymbol{\Omega C'(C\Omega C')}^{-1}\boldsymbol{X}_a[\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{X}_a]^{-1} \tag{12}$$

Substitute (12) into (10) to get —

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{\Omega C'(C\Omega C')}^{-1} + \boldsymbol{X}[\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{X}_a]^{-1}\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1} \\ &\quad - \boldsymbol{\Omega C'(C\Omega C')}^{-1}\boldsymbol{X}_a[\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{X}_a]^{-1}\boldsymbol{X}'_a\boldsymbol{(C\Omega C)}^{-1} \\ &= \boldsymbol{X}[\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{X}_a]^{-1}\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1} \\ &\quad + \boldsymbol{\Omega C'(C\Omega C')}^{-1}[I - [\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{X}_a]^{-1}\boldsymbol{X}'_a\boldsymbol{(C\Omega C)}^{-1}] \end{aligned} \tag{13}$$

The resulting estimator is now

$$\begin{aligned} \hat{\boldsymbol{y}} &= \boldsymbol{Ay}_a \\ &= \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\Omega C'(C\Omega C')}^{-1}\hat{\boldsymbol{u}}_a \end{aligned} \tag{14}$$

where

$$\hat{\boldsymbol{\beta}} = [\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{X}_a]^{-1}\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{y}_a \tag{15}$$

is the GLS estimator in the regression of $\boldsymbol{y}_a$ on $\boldsymbol{X}_a$ and

$$\hat{\boldsymbol{u}}_a = \boldsymbol{y}_a - \boldsymbol{X}[\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{X}_a]^{-1}\boldsymbol{X}'_a\boldsymbol{(C\Omega C')}^{-1}\boldsymbol{y}_a \tag{16}$$

are the residuals in this regression.

We may summarise the result as follows

1. Estimate the GLS estimates of the coefficients in the annual regression of $\boldsymbol{y}_a$ on $\boldsymbol{X}_a$. The distributed series is the sum of two parts

2. The first part is found by applying the estimated regression coefficients to the quarterly indicators

3. The second part is found by distributing the residual in the annual regression over the quarters. This ensures that the distributed series sum to the known annual series.

The implementation of this procedure involves two practical problems. The first is the choice of indicators. This must be dealt with on a pragmatic basis for each series (see Introduction above). The second problem is that the covariance matrix $\mathbf{\Omega}$ is, in general, unknown. Obviously we can not estimate all $(4n \times 4n)$ elements in $\mathbf{\Omega}$. We must assume some structure for $\mathbf{\Omega}$. Four cases are provided in the program. Cases (1) to (3) below are set out in Chow and Lin (1971). In previous distribution/interpolation exercises either a test was carried out for the significance of autocorrelation in regression of the annual variables. If this was not significant case (1) was used. If significant autocorrelation was found then case (3) was used.

In our practical applications of the methodology the estimated equations produced very poor out of sample forecasts of the distributed/interpolated variables. The autocorrelation found in the estimations is likely to be due to misspecification of the variable–indicator model. We are also concerned that the the assumption of homoscedasticity may cause a loading of errors at the start of the sample in the type of regressions estimated for the model.

Case 4 implements an extension to cater for both autocorrelation and heteroscedasticity. The program estimates the unknown parameters in the model using pseudo maximum likelihood rather than the iterative procedure suggested by Chow and Lin (1971). The procedure can be extended to account for various other autocorrelation structures. A maximum likelihood estimate of the other cases can also be completed by an appropriate choice of options in running the procedure.

Case 1 : ε are uncorrelated variables with constant variance

In this case

$$\Omega C'(C\Omega C')^{-1} = \begin{pmatrix} \frac{1}{4} & 0 & \dots & 0 \\ \frac{1}{4} & 0 & \dots & 0 \\ \frac{1}{4} & 0 & \dots & 0 \\ \frac{1}{4} & 0 & \dots & 0 \\ 0 & \frac{1}{4} & \dots & 0 \\ 0 & \frac{1}{4} & \dots & 0 \\ 0 & \frac{1}{4} & \dots & 0 \\ 0 & \frac{1}{4} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{4} \\ 0 & 0 & \dots & \frac{1}{4} \\ 0 & 0 & \dots & \frac{1}{4} \\ 0 & 0 & \dots & \frac{1}{4} \end{pmatrix}$$

$\boldsymbol{\beta}$ is estimated by OLS and the residual is divided equally between the four quarters.

Case 2 : ε are uncorrelated variables but are heteroscedastic

In this case

$$\boldsymbol{\Omega} = \begin{pmatrix} \omega_1 & 0 & \ldots & 0 \\ 0 & \omega_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \omega_{4n} \end{pmatrix}$$

and

$$\boldsymbol{\Omega}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{\Omega}\boldsymbol{C}')^{-1} = \begin{pmatrix} \frac{\omega_1}{\omega_1+\omega_2+\omega_3\omega_4} & 0 & \ldots & 0 \\ \frac{\omega_2}{\omega_1+\omega_2+\omega_3\omega_4} & 0 & \ldots & 0 \\ \frac{\omega_3}{\omega_1+\omega_2+\omega_3\omega_4} & 0 & \ldots & 0 \\ \frac{\omega_4}{\omega_1+\omega_2+\omega_3\omega_4} & 0 & \ldots & 0 \\ 0 & \frac{\omega_5}{\omega_5+\omega_6+\omega_7\omega_8} & \ldots & 0 \\ 0 & \frac{\omega_6}{\omega_5+\omega_6+\omega_7\omega_8} & \ldots & 0 \\ 0 & \frac{\omega_7}{\omega_5+\omega_6+\omega_7\omega_8} & \ldots & 0 \\ 0 & \frac{\omega_8}{\omega_5+\omega_6+\omega_7\omega_8} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{\omega_{4n-3}}{\omega_{4n-3}+\omega_{4n-2}+\omega_{4n-1}\omega_{n4}} \\ 0 & 0 & \ldots & \frac{\omega_{4n-2}}{\omega_{4n-3}+\omega_{4n-2}+\omega_{4n-1}\omega_{n4}} \\ 0 & 0 & \ldots & \frac{\omega_{4n-1}}{\omega_{4n-3}+\omega_{4n-2}+\omega_{4n-1}\omega_{n4}} \\ 0 & 0 & \ldots & \frac{\omega_{4n}}{\omega_{4n-3}+\omega_{4n-2}+\omega_{4n-1}\omega_{n4}} \end{pmatrix}$$

In this case $\boldsymbol{\beta}$ is estimated by GLS and the residuals are distributed in proportion to the variance in the quarter. The variance may be taken as proportional to a function of the indicator variable(s) or perhaps of the principal component of the indicator variables.

Case 3 : Innovations in ε are homoscedastic but ε follows an $AR(1)$ process

In this case we assume that the quarterly residuals follow an $AR(1)$ process. In this case

$$\Omega = \frac{\sigma^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{4n-1} \\ \rho & 1 & \rho & \dots & \rho^{4n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{4n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{4n-1} & \rho^{4n-2} & \rho^{4n-3} & \dots & 1 \end{pmatrix}$$

The covariance of the annual residuals is then given by $\boldsymbol{C\Omega C'}$. It can be shown that the ratio of element (1,2) to element (1,1) in this covariance matrix is given by —

$$\alpha = \frac{\rho + 2\rho^2 + 3\rho^3 + 4\rho^4 + 3\rho^5 + 2\rho^6 + \rho^7}{4 + 6\rho + 4\rho^2 + 2\rho^3}$$

Taking α as a function of ρ

$$\alpha(0) = 0$$
$$\alpha(1) = 1 \text{ and}$$
$$\alpha(\rho) > 0 \text{ for } 0 < \rho \le 1.$$

Thus for a particular value of α in the range $(0,1)$ there corresponds a unique value of ρ.

When $-0.130544 < \alpha] < 0$ there are two values of ρ in the range $(-0.680297, 0)$ for each value of α. For values of α in the range $(-1, -0.130554)$ there is no real solution to the equation. Thus this methodology is not valid for the distribution of quarterly data when α is negative. In previous implementations of this procedure we found no negative values of α

Case 4 : Innovations in ε are heteroscedastic and "pseudo AR(1)" process

This Special Case is a combination of 3.2 and 3.3. We may write

$$\boldsymbol{\Omega} = \boldsymbol{WVW}$$

$$= \frac{1}{1-\rho^2} \begin{pmatrix} \omega_1 & 0 & \ldots & 0 \\ 0 & \omega_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \omega_{4n} \end{pmatrix} \times$$

$$\begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{4n-1} \\ \rho & 1 & \rho & \ldots & \rho^{4n-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{4n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{4n-1} & \rho^{4n-2} & \rho^{4n-3} & \ldots & 1 \end{pmatrix} \begin{pmatrix} \omega_1 & 0 & \ldots & 0 \\ 0 & \omega_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \omega_{4n} \end{pmatrix}$$

where ω_1, ω_2, ... ,ω_{4n} are the square roots of the variances of the quarterly regressions and ρ is the first order auto-correlation of the quarterly residuals. This methodology may be implemented using the following steps:

1. Estimate the ω's using an appropriate function of the indicator variable(s) (e.g. proportional to a principal component of the indicator variable – this is the assumption made in the subroutine). This determines $\boldsymbol{W}$ up to a factor σ
   .

2. If the annual variable $\boldsymbol{y}_a$ is multivariate normal with mean $\boldsymbol{CX\beta}$ and variance $\boldsymbol{C'W'VWC}$ the log of the likelihood is given by

$$-\frac{N}{2}Log(2\pi) - \frac{1}{2}Log(|\boldsymbol{C'W'VWC}|)$$
$$-\frac{1}{2}(\boldsymbol{y}_a - \boldsymbol{CX\beta})'(X_a'(\boldsymbol{C'W'VWC})^{-1}X_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{CX\beta})$$

3. For a given value of ρ the maximum of this expression occurs when

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}_a'(\boldsymbol{C'W'VWC})^{-1}\boldsymbol{X}_a)^{-1}\boldsymbol{X}_a'(\boldsymbol{C'W'VWC})^{-1}\boldsymbol{y}_a$$

and

$$\hat{\sigma}^2 = \frac{1}{N}(\boldsymbol{y}_a - \boldsymbol{CX\hat{\beta}})'(\boldsymbol{X}_a'(\boldsymbol{C'W'VWC})^{-1}\boldsymbol{X}_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{CX\hat{\beta}})$$

and the maximum value is

$$-\frac{N}{2}(log(2\pi) + 1) - \frac{N}{2}Log(\hat{\sigma}^2) - \frac{1}{2}Log(|\boldsymbol{C'CC}|)$$

The FIND command in RATS is then used to maximise this expression with respect to ρ.

4. Use the results at (3) to distribute $\boldsymbol{y}_a$ between quarters using equations 14 to 16 above with $\boldsymbol{\Omega}$ replaced by $\boldsymbol{W'VW}$

One can verify that this solution satisfies the constraints by premultiplying (9) by $\boldsymbol{C}$

References

[1] **Boot, J. C. G., W Feibes and J. H. C. Lisman** "Methods of deriving Quarterly Series from Annual Data", Applied Statistics, Vol 16, p 65-75.

[2] **Chow, Gregory C and Lin An-loh (1971)**, "Best Linear Unbiased Distribution and Interpolation of Time Series by Related Series", Review of Economics and Statistics, 53, 372-5.

[3] **Gómez, V. and A. Maravall** (1979), "TRAMO and SEATS : Instructions for the user", Banco de España, www.bde.es.

[4] **McGuire, M and M. Ryan**, (2000), "Macroeconomic Modelling Developments in the Central Bank", Central Bank of Ireland Quarterly Bulletin, Spring 2000.

[5] **O'Reilly, Liam** (1981), "Estimating Quarterly National Accounts", Journal of the Statistical and Social Inquiry Society of Ireland, Vol XXIV, Part IV, 1981/82, pp 1-39.

A   Listing of subroutine

```
/*

Procedure to convert annual series to quarterly using related indicators (Chow Lin (1971)).

John C Frain
Economics (ERP) Dept.
Central Bank of Ireland

Syntax

@y2q(options) yseries qseries start end
# <list of indicators in regression format>

     yseries =    series to interpolate. This should be set up as
                  a quarterly series.  Y2Q takes as the value to
                  interpolate the last quarter of each year

                  If you read an annual series from a RATS format
                  file into a RATS program with quarterly frequency
                  in the CALENDAR statement it will set up in this
                  way automatically

     qseries =    interpolated quarterly series

     start end    Range to use in interpolation.  start
                  and end should refer to quarters 1 and 4 respectively.

WARNING  WARNING WARNING WARNING  WARNING WARNING WARNING  WARNING WARNING

The list of indicators may include constant and trend terms.  Constant and trend
terms may also be  specified as options in the procedure call.  If they are included
in the list of indicators they will be used in the calculation of the Principal
Component.  If there is a single indicator do not use the PC option in conjunction
the CONST option.  In this case specify a constant in your indicators and use the
noconst option (if this is what you require).

The annual series should contain no #N/A over the interpolation range.
All indicators must contain no #N/A from the first period of the start year to
the final period of the end year of the interpolation.  If an #N/A is found the
procedure halts and returns an error message
```

```
Options;
     [pc] nopc        Heteroskedasticity adjustment required using sole indicator
                      or principal component if more than one indicator default
                      is to adjust

     const/[noconst]  noconst - default is to include constant with indicator
     trend/[notrend]  notrend - default is to exclude trend - can also be
                      included with indicators

     startrho         Initial value for rho in maximum liklihood estimation Default
                      is 0.5.

The procedure includes an amended version of the principal components procedure
distributed with RATS

References

Chow, Gregory C and Lin An-loh (1971). "Best Linear Unbisaed Interpolation, Distribution
and Interpolation of Time Series be Related Series", Review of Economics and Statistics,
53,372-5.

Frain, John (2004), "A RATS subroutine to to implement the Chow-Lin
distribution/interpolation procedure"  Central Bank and Financial Services
Regulatory Authority of Ireland Research Technical Paper 2/RT/04.




*/
/*
env noecho
env interactive nosubscripterrors
*/
********************************************************************************
*  PRINCOMP is a simple procedure for extracting principal components
*  from a set of series.
*
*  PRINCOMP  start  end   base
*  #   list of variables (regression format)
*
*  Options:
*    NCOMPS = number of components [number of variables]
*
*  start AND end should be provided explicitly
```

```
*  On output the principal components are put into series base+1,base+2,...
*
PROCEDURE PRINCOMP START END BASE
TYPE INTEGER START END
TYPE INTEGER *BASE
*
OPTIONS INTEGER NCOMPS  1 ;* changed JCF as one principal component only is required here
*
LOCAL INDEX VARS
LOCAL RECT XMATRIX EIGENVEC
LOCAL SYMM CMOM
LOCAL VECT EIGENVAL PRINCOEF COMPONNT
LOCAL INT  N
*
MAKE XMATRIX START END NOBSMAKE NVARS
MATRIX CMOM=TR(XMATRIX)*XMATRIX
EIGEN CMOM EIGENVAL EIGENVEC
IF .NOT.NCOMPS
   COMPUTE N = NVARS
ELSE
   COMPUTE N = NCOMPS
SCRATCH N START END BASE
DO ICOMP=1,N
   OVERLAY EIGENVEC(1,ICOMP) WITH PRINCOEF(NVARS)
   MATRIX COMPONNT=XMATRIX*PRINCOEF
   SET BASE+ICOMP START END = COMPONNT(T-START+1)
END DO ICOMP
END
*
*
*
*******************************************************************************
*
*  The procedure below calculates the auto correlation matrix of an AR(1) process
*  rho is the value of rho in e(t) = rho * e(t-1) + v(t)
*  var(v) = 1
*  n is the dimcnsion of the correlation matrix
*
Procedure corrmat rho n ar1corr
type real rho
type integer n
type symm *ar1corr
dim ar1corr(n,n)
ewise ar1corr(i,j) = rho**(i-j)/(1-rho**2)
```

```
end proc
*
*
********************************************************************************
*
*
PROCEDURE y2q yseries qseries start end

Option switch pc 0
option switch const 0
option switch trend 0
option real startrho 0.5

type series yseries ;*annual series to be interpolated
type series *qseries ;*interpolated series

type integer start end
local series dlqseries d4lqseries d4lyseries seruhat;* series for graphs


local integer startl endl

local vector[integer] list_indic ;* series numbers for indicators
local integer no_indic ;* number of indicators
local vector[integer] listvar ;* series nos of indicators and variable to be interpolated


local integer factor        ;* going from annual to quarterly if 4
local integer nyear             ;* number of years in analysis
local series ya              ;* annual series
local series temp
local vector[series] xa ;* series nos of annualised indicators

local integer w         ;*counter for principal component series (W+1)
local series het ;*principal component of indicator variables
local rect hetvec        ;*principal component in vector format
local rect hetmat        ;*contains het on diagonal

local series qtrend qintercept ;*quarterly intercept and trend series
local rect tempvec
local rect c                    ;*transforms data from quarterly to annual by summing quarterlies

local rect xamatrix
local rect yamatrix
```

20

```
local rect xqmatrix

local real loglik
local real sigmasq
local symm omega
local rect delta
local vect betahat
local vector uhat
local vector yqhat
local vector[string] varlabel
*
* read supplementary card
*
enter(varying,entries=no_indic) list_indic
* listvar is list of all input variables to be checked for #N/A
enter(varying) listvar
# yseries list_indic
*
*Set and check start date and end date for interpolation
*
inquire(regressorlist) startl>>start endl>>end
#listvar
if (%month(startl)<>1) .or. (%month(endl)<>4)
{
messagebox "There is a start/end date problem - Please restart"
return
}
*check for #N/A
dofor ii = listvar
do jj = startl, endl
if %valid(([series] ii)(jj)) <> 1
{
messagebox "N/A data in the specified or default range - Please restart"
return
}
end do jj
end dofor ii
*
* Extract Principal Component of Indicator Series
*
if pc
{
if no_indic==1
{
```

```
set het startl endl = abs([series]list_indic(1))
}
else
{
@princomp startl endl w
#list_indic
set het startl endl = abs([series] (w+1))
}
}
else
{
set het startl endl = 1
}
make hetvec startl endl
#het
compute hetmat = %diag(hetvec)
*print / het
*write hetmat
*
*WARNING WARNING WARNING
*
*adjust list of quarterly indicator variables to include constant and trend terms.  These are
*added if specified on options.  If a constant or trend are included on the supplementary card
*in the call they are included in the calculations of principal components.  If they are included
*in the options they are not included in the calculation of principal components.  This allows
*the user a wider choice but he must decide what is required.  I would think that they should
*be included in one place only.  Under certain circumstances the program will crash if used in
*both places
*
set qtrend startl endl = t-startl+1
set qintercept startl endl = 1
if trend
{
enter(varying,entries=no_indic) list_indic
#qtrend list_indic
}
*
*
*
if const
{
enter(varying,entries=no_indic) list_indic
#qintercept list_indic
}
```

```
*
*extract annual variable
*
compute factor=4;
compute nyear = (endl-startl +1)/factor
set ya 1 nyear = yseries(startl+4*t-1)
*
*extract annualised indicators
*
dim xa(no_indic)
do ii =1, no_indic
clear temp
filter list_indic(ii) startl+3 endl temp
# 0 1 2 3
# 1 1 1 1
set xa(ii) 1 nyear = temp(startl + 4*t -1) ;
end do ii
*
* xqmatrix is matrix of quarterly indicators
*
make xqmatrix startl endl
# list_indic

*print 1 nyear ya xa(1) xa(2)
*
* set up C matrix
*
compute tempvec = ||1.0, 1.0, 1.0, 1.0||
compute c = %kroneker(%identity(nyear),tempvec)

make yamatrix 1 nyear
#ya
dim xamatrix(nyear,no_indic)
ewise xamatrix(i,j) = ([series]xa(j))(i)
*
nonlin rho
compute rho = startrho
find maximum loglik

@corrmat rho 4*nyear omega

compute delta = c*hetmat*omega*hetmat*tr(c) ;*hetmat is symmetric

compute betahat = inv( tr(xamatrix) * inv(delta) * xamatrix) * tr(xamatrix) * inv(delta) * yamatrix
```

23

```
compute uhat = yamatrix-xamatrix*betahat

compute sigmasq = %scalar((1.0/nyear)*tr(uhat)*inv(delta)*uhat)

compute loglik=0

compute loglik  = -0.5*nyear*(log(2.0*%pi))$
  -  0.5 * nyear * log(sigmasq) $
  -  0.5 * log(%det(delta)) $
  - (1/(2.0*sigmasq)*(%scalar(tr(uhat) * inv(delta) * (uhat)) ))
disp "loglik = " loglik "rho = " rho "sigmasq = " sigmasq
end find
*
*
disp "coefficients"
write betahat
*
compute yqhat = xqmatrix*betahat + hetmat*omega*hetmat*tr(c) *inv(delta)*uhat
set qseries startl endl  = yqhat(t-startl+1)
* some pictures

set dlqseries = 100.0*log(qseries/qseries(t-1))
set d4lqseries = 100.0*log(qseries/qseries(t-4))
set d4lyseries = 100.0*log(yseries/yseries(t-4))

compute varlabel=||"interpolated Q on Q","interpolated Q on Q(-4)","annual Y on Y"||

graph(key=loright,Header="Rates of change of interpolated and annual series",klabel=varlabel,scale=both,patterns)
#dlqseries
#d4lqseries
#d4lyseries

disp " "
disp "Statistics on residual in GLS regression"
set seruhat = uhat(t)
statistics seruhat

disp " "
disp "Statistics of interpolated series"

statistics qseries

disp " "
```

```
disp "Printout of interpolated series, Q on Q and Q on Q(-4) %changes "
print / qseries dlqseries d4lqseries

end proc
```