Program GAP*- Version 2.2
Technical Annex
May 2003

Christophe Planas and Alessandro Rossi

Emails: christophe.planas@jrc.it, alessandro.rossi@jrc.it,

European Commission, Joint Research Centre,

Ispra, Italy.

---

*Please read disclaimer in worksheet "About " of the Excel interface.

Contents

1 Introduction and overview

Program GAP is a program developed by the Joint Research of European Commission on request of DG-ECFIN, following the guidelines of Werner Roeger that is gratefully acknowledged. The aim of the program is to decompose a macroeconomic quantity such that GDP or unemployment into long-term and short-term components, the short-term one being identified with output gap or unemployment gap, according to the input series. The program is based on Kuttner (1994) bivariate model that involves a Phillips-curve regression. Program GAP can be downloaded at www.jrc.cec.eu.int/uasa, following the links Area of Activities - Time Series Analysis - Projects - Gap - Further Information. The program is delivered together with an Excel interface and with a help file for the interface that can be found in the Help worksheet. We strongly suggest users to read the disclaimer in the "About" worksheet before running the program.

The model implemented in Program GAP is somewhat atypical. While most statistical decompositions are univariate, Kuttner's model associates to a classical decomposition a regression whose regressors include unobserved quantities such that the gap and its lags. This document describes the statistical procedures that have been used, including the references. The implementation is based on state-space models, with the model parameters estimated by exact maximum likelihood. It involves running the Kalman recursions with de Jong's diffuse initialisation (de Jong, 1991), and a smoother that produces the unobserved quantities.

Section 2 describes the general model specification. A summary table is given for a fast overview of the possibilities offered. Section 3 discusses model parameter estimation using state-space methods. The procedure used for obtaining standard deviations is also detailed together with the diagnostics that Program GAP reports. An example is developed for helping reader's comprehension. Section 4 presents the de Jong's diffuse Kalman filter that solves for initialisation problem for non-stationary systems. Finally, Section 5 explains the smoothing mechanism for unobserved component estimates. All references are grouped at the end of this document.

2 General model specification

The most general model that can be specified is a bivariate model like in Kuttner (1994) with an extension that allows for exogenous regressors and for autoregressive terms in the Phillips curve equation. Generally the variables considered are either unemployment or GDP and inflation but in the discussion here we refer to them simply as first and second variable. The first equation is specified similarly to regression models with ARIMA errors (see Fuller, 1996):

$$X_{1t} = \sum_{i=1}^{M_1} \alpha_{1i} Z_{1it} + \widetilde{X}_{1t}, \tag{1}$$

where $Z_{1.t}$ is a vector of $M_1 \leq 3$ exogenous variables. The remainder of this regression, $\widetilde{X}_{1t}$, is described as made up of a long term component or trend, X_{1t}^N, and of a short term component or cycle, X_{1t}^S, according to:

$$\widetilde{X}_{1t} = X_{1t}^N + X_{1t}^S. \tag{2}$$

The dynamic behavior of the short-term component is described by an AR(2) model as in:

$$\left(1 - \phi_1 L - \phi_2 L^2\right) X_{1t}^S = a_t^S, \tag{3}$$

where a_t^S is a white noise innovation with variance V_S. Trivially, if the AR parameters are restricted to be null, then the short-term component becomes a white noise. The trend or long-term component can be specified as a second order random walk:

$$(1 - L) X_{1t}^N = \mu_{1t-1} + a_t^N, \tag{4}$$

where the slope is such that:

$$(1 - L) \mu_{1t} = a_t^\mu. \tag{5}$$

Both a_t^N, and a_t^μ, are white noise innovations with variances V_N and V_μ. Notice that a null variance for the slope innovations, $V_\mu = 0$, implies that (4) and (5) reduces to a random walk plus drift and the trend becomes integrated of order 1 instead of 2. Also, in the second-order random walk plus noise model, setting to zero the variance of the innovations in the trend level, i.e. $V_N = 0$, and considering a simple white noise for the short term component, i.e. $\phi_1 = \phi_2 = 0$, yields the I(2)-trend plus noise model that is implicit in Hodrick-Prescott detrending (see Harvey and Jaeger,

1993), with inverse signal to noise ratio given by V_S/V_μ (see Hodrick and Prescott, 1997). Program Gap offers all these specifications. Discussions of univariate models for measuring gaps can be found for instance in Watson (1986) and in Clark (1987).

When exogenous variables are used, these are then assigned to the long-term component so that the final decomposition is:

$$X_{1t} = X_{1t}^{NF} + X_{1t}^S, \tag{6}$$

where the final trend is such that:

$$X_{1t}^{NF} = \sum_{i=1}^{M_1} \alpha_{1i} Z_{1it} + X_{1t}^N. \tag{7}$$

Similarly to the model put forward by Kuttner (1994), series 1 is related to a stationary transformation of the second series, typically change in inflation, according to:

$$
\begin{aligned}
X_{2t} &= \mu_2 + \sum_{i=1}^{M_2} \alpha_{2i} Z_{2it} + \gamma (1-L)^d X_{1t-1} + \sum_{i=0}^{r} \beta_i X_{1t-i}^S + \\
&\quad + \sum_{i=1}^{p} \phi_i^* X_{2t-i} + \sum_{i=0}^{q} \theta_i a_{t-i}^\pi,
\end{aligned}
\tag{8}
$$

where μ_2 is the intercept, $Z_{2\cdot t}$ is a vector of $M_2 \leq 10$ exogenous variables, d is the order of integration of the first series, r is the number of lags for which the short-term component of the series is supposed to have an impact on the change in inflation constrained to lie between 0 and 4, p is the number of autoregressive terms constrained to be less or equal to 2, q is the number of moving average terms with 3 as maximum and $\theta_0 = 1$. Finally the innovation a_t^π is a white noise with variance V_π that can be correlated with a_t^S but only when $\beta_0 = 0$. That correlation is denoted by ρ.

Program Gap allows users to enter restrictions on any of these parameters. Practitioners should be aware of several features:

1. In equation (8), the endogenous variable X_{2t} should be covariance stationary. Most often a preliminary stationary transformation is needed, so for instance changes in inflation are generally considered instead of inflation.

5

2. If X_{1t}^S is included in (8), the innovations a_t^τ and this regressor must be orthogonal. Thus Program Gap imposes $\rho = 0$. Not doing so would lead to an inconsistent parameter estimator.

3. If $V_S = 0$, Program Gap estimates no cycle, so the parameters in (8) are not identified. In this case they are automatically set to 0.

4. If a second-order random walk is specified for the first series trend, then in (8) the lagged first series is differenced twice, i.e. $d = 2$, in order to make the regressor stationary.

As mentioned in point 2, in some cases it can happen that the estimation yields a 0-innovation variance for the short-term component of the first series. In this case the cycle is estimated as 0, so no series decomposition is actually performed. In order to overcome this problem, Program Gap offers the possibility to specify a *canonical* model (see Pierce, 1978; Hillmer and Tiao, 1982) for the trend component in equation (4)-(5). Let us assume that a random walk plus drift is the model describing the long-term movements. It can be seen that the associated pseudo-spectrum, say $g_p(\omega)$, is such that

$$g_p(\omega) = \frac{V_N}{2 - 2\cos(\omega)},$$

where $\omega \in [0, \pi]$ is a frequency expressed in radians (see Harvey, 1989, pp.64-65). Since $g_p(\pi) = V_N/4$, movements at frequency π or with periodicity 2 do contribute to the overall variance of the trend component. Yet, since the decomposition is into long-term and short-term components, all movements with periodicity 2 should belong to the short-term component while the trend should be free of short-term fluctuations. This is imposed in the so-called canonical specification. That decomposition is achieved by removing from the trend spectrum its minimum, which is $g_p(\pi)$ in the case of a random walk, and by assigning it to the cycle spectrum. The models obtained for the new decomposition are then such that the trend model is non-invertible:

$$(1 - L) X_{1t}^N = \mu_1 + (1 + L) a_t^N,$$

and the AR(2) model for the cycle becomes an ARMA(2,2) model:

$$\left(1 - \phi_1 L - \phi_2 L^2\right) X_{1t}^S = \left(1 + \theta_1^S L + \theta_2^S L^2\right) a_t^S.$$

Because such a canonical specification maximises the cycle innovation variance in the series decomposition, it can be helpful for overcoming the problem of a 0-innovation variance. That specification is only allowed when the trend is specified as a random walk plus drift. ($V_\mu = 0$). If used, the order of the MA term in the second equation is restricted to be less or equal to 1. For a discussion of the properties of the canonical decomposition, see Maravall and Planas (1999).

Table 1 displayed in next page summarises all the specifications offered by Program Gap.

Table 1 Program GAP: options for model specification	
First series decomposition	▷ Bivariate
	▷ Univariate
	▷ Hodrick-Prescott filter
First equation	
Trend	▷ Second-order random walk
	▷ I(2)
	▷ Random walk plus drift
	▷ Canonical IMA(1,1) plus drift
Cycle	▷ AR(2)
	▷ AR(1)
	▷ White noise
	▷ ARMA(2,2) (if canonical IMA(1,1) is used for trend)
Exogenous regressors	▷ From 0 to 3
Second equation - Phillips curve	
Autoregressive terms	▷ AR(2)
	▷ AR(1)
	▷ No AR term
Moving average terms	▷ From 0 to 3
First series terms	▷ Diffrenced series with 1 lag
	▷ From 0 to 4 lags on cycle
Exogenous regressors	▷ Intercept
	▷ From 0 to 10
Innovatrions	▷ Orthogonal
	▷ Cross-correlated with first series innovations

3   Model parameter estimation

Let Λ denote the vector of all unknown parameters in model (1)-(8):

$$\Lambda = (\phi_1, \phi_2, \mu_N, \mu_\pi, \gamma, \beta_0, \beta_1, ..., \beta_4, \theta_1, \theta_2, \theta_3, \theta_1^S, \theta_2^S,$$
$$V_N, V_S, \rho, V_\pi, \alpha_{1,1}, ..., \alpha_{1,3}, \alpha_{2,1}, \alpha_{2,2}, ..., \alpha_{2,10}, \phi_1^*, \phi_2^*)'.$$

While the vector Λ has dimension 35, because of several mutual exclusions it never happens that 35 parameters need to be estimated. The exclusions are:

- either $\mu_N = 0$ or $V_\mu = 0$ or both, but they cannot be both different from zero;

- $\beta_j = 0$ for $j = 0, 1, ..., 4$ if $V_S = 0$;

- $\rho = 0$ if $\beta_0 \neq 0$ and $\beta_0 = 0$ if $\rho \neq 0$;

- $\theta_2 = \theta_3 = 0$ if canonical trend is considered.

The maximum number of parameters that can be estimated is thus 31.

Let $\mathbf{X_t}$ and $\mathbf{Z_t}$ denote the vectors of observations and of exogenous variables, respectively, so $\mathbf{X_t} = (\mathbf{X_{1t}}, \mathbf{X_{2t}})'$ and $\mathbf{Z_t} = (\mathbf{Z_{1t}}, \mathbf{Z_{2t}})'$, for the dates $t = 1, 2, ..., T$. We also denote by $\mathbf{X^t}$ and $\mathbf{Z^t}$ the vectors of observations and of exogenous up to time t. Because the first series is integrated of order $d = 1, 2$, the likelihood function is defined conditionally on the first d observations according to:

$$p\left(\mathbf{X_{d+1}}, ..., \mathbf{X_T} \,\middle|\, \Lambda, \mathbf{X^d}, \mathbf{Z^T}\right) = \prod_{t=d+1}^{T} f\left(\mathbf{X_t} \,\middle|\, \Lambda, \mathbf{X^{t-1}}, \mathbf{Z^T}\right), \qquad (9)$$

where $f(\cdot)$ is the normal pdf. Model parameters estimates are obtained by maximising the log of (9). For evaluating the log-likelihood, we cast the model into a state-space format (see Harvey, 1989, pp.100-166; Hamilton, 1994, pp 372-408). State-space models are defined by a measurement equation like

$$\mathbf{X_t} = \mathbf{H}\xi_\mathbf{t} + \mathbf{C}\mathbf{Z_t} + \mathbf{u_t}, \qquad (10)$$

where $\mathbf{X_t}$ is an observed vector of dimension $k \times 1$, $\xi_\mathbf{t}$ is the state vector with dimension $n \times 1$, $\mathbf{H}$ a matrix of dimension $k \times n$, $\mathbf{u_t}$ a vector of innovations with covariance

matrix R, $\mathbf{Z_t}$ a vector of r exogenous variables and C is a $k \times r$ matrix. The model is completed by a transition equation for the state vector such that

$$\xi_{\mathbf{t+1}} = \mathbf{F}\xi_{\mathbf{t}} + \mathbf{v_{t+1}}, \tag{11}$$

where $\mathbf{F}$ is a $n \times n$ transition matrix and $\mathbf{v_t}$ is a vector of innovations with covariance matrix Q. In the case of model (1)-(8) the matrices C, F, H, Q, and R are constant, but this is not a requirement in state space models.

Let us consider for instance model (1)-(7) with the equation (8) for the second series simplified as in:

$$X_{2t} = \beta_0 X_{1t}^S + a_t^\pi + \theta_1 a_{t-1}^\pi$$

and $\rho = 0$. If the state variable is chosen to be such that $\xi_{\mathbf{t}} = (\mathbf{X_{1t}^N}, \mu_{\mathbf{1t}}, \mathbf{X_{1t}^S}, \mathbf{X_{1t-1}^S}, \mathbf{a_t^\pi},$ $a_{t-1}^\pi)'$, then it is easily seen that , $\mathbf{v_t} = (\mathbf{a_t^N}, \mathbf{a_t^\mu}, \mathbf{a_t^S}, \mathbf{0}, \mathbf{a_t^\pi}, \mathbf{0})$, $\mathbf{u_t} = (\mathbf{0}, \mathbf{0})$, $R = 0_{2\times 2}$, and the matrices F, H, and Q, are given by:

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_1 & \phi_2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \beta_0 & 0 & 1 & \theta_1 \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} V_N & 0 & 0 & 0 & 0 & 0 \\ & V_\mu & 0 & 0 & 0 & 0 \\ & & V_S & 0 & 0 & 0 \\ & & & 0 & 0 & 0 \\ & & & & V_\pi & 0 \\ & & & & & 0 \end{bmatrix}$$

In the state space framework, evaluation of the log-likelihood can be computed as follows. Normality is assumed for convenience; shouldn't it hold the conditional expectation operators should be interpreted as linear projections and maximum likelihood estimator as quasi-maximum likelihood estimator. It can be seen that for all dates, the computation of $f\left(\mathbf{X_t} \middle| \mathbf{\Lambda}, \mathbf{X^{t-1}}, \mathbf{Z^T}\right)$, in (9) only involves the innovation vector $\mathbf{X_t} - \mathbf{E}\left[\mathbf{X_t} \middle| \mathbf{X^{t-1}}, \mathbf{Z^T}\right]$, where $\mathbf{X_{t|t-1}} = \mathbf{E}\left[\mathbf{X_t} \middle| \mathbf{X^{t-1}}, \mathbf{Z^T}\right]$, denotes expectation of $\mathbf{X_t}$ conditional on $\mathbf{X^{t-1}}$, and $\mathbf{Z^T}$, and the innovation covariance matrix

$Var\left[\mathbf{X_t} - \mathbf{X_{t|t-1}}\right]$. Writing $\xi_{t|t-1} = \mathbf{E}\left[\xi_t \left| \mathbf{X^{t-1}}, \mathbf{Z^T}\right.\right]$, from (10) it is easily seen that $\mathbf{X_{t|t-1}} = \mathbf{H}\xi_{t|t-1} + \mathbf{CZ_t}$, and that the innovation covariance matrix verifies

$$Var\left[\mathbf{X_t} - \mathbf{X_{t|t-1}}\right] = \mathbf{HP_{t|t-1}H'} + \mathbf{R}$$

where $\mathbf{P_{t|t-1}} = \mathbf{Var}\left[\xi_t - \xi_{t|t-1}\right]$. The problem is thus to find $\xi_{t|t-1}$ and $\mathbf{P_{t|t-1}}$ for all $t = d+1, ..., T$. By now, the initial condition $\xi_{d+1|d}$ and its covariance matrix $\mathbf{P_{d+1|d}}$ are taken as given; they will be discussed in the next section. From (10)-(11) and under the normality hypothesis, the joint distribution of the state vector ξ_t and of the observation $\mathbf{X_t}$ conditional on the past observations and on the exogenous variables is given by:

$$\begin{matrix} \xi_t \\ \mathbf{X_t} \end{matrix} \left| \mathbf{X^{t-1}}, \mathbf{Z^T} \sim \mathbf{N}\left(\begin{matrix} \xi_{t|t-1} \\ \mathbf{H}\xi_{t|t-1} + \mathbf{CZ_t} \end{matrix}, \left[\begin{matrix} \mathbf{P_{t|t-1}} & \mathbf{P_{t|t-1}H'} \\ & \mathbf{HP_{t|t-1}H'} + \mathbf{R} \end{matrix} \right] \right). \quad (12)$$

The properties of the normal distribution (see Harvey, 1989, p.165) then imply:

$$\xi_{t|t} = \xi_{t|t-1} + \mathbf{P_{t|t-1}H'}\left(\mathbf{HP_{t|t-1}H'} + \mathbf{R}\right)^{-1}\left(\mathbf{X_t} - \mathbf{H}\xi_{t|t-1} - \mathbf{CZ_t}\right) \quad (13)$$

$$\mathbf{P_{t|t}} = \mathbf{P_{t|t-1}} - \mathbf{P_{t|t-1}H'}\left(\mathbf{HP_{t|t-1}H'} + \mathbf{R}\right)^{-1}\mathbf{HP_{t|t-1}}. \quad (14)$$

Finally, equation (11) implies:

$$\xi_{t+1|t} = \mathbf{F}\xi_{t|t} \quad (15)$$

$$\mathbf{P_{t+1|t}} = \mathbf{FP_{t|t}F} + \mathbf{Q} \quad (16)$$

Equations (13)-(16) make up the well-known Kalman recursions (see Harvey, 1989, pp.104-112; Hamilton, 1994, pp.377-380). Running (13)-(16) for all $t = d+1, ..., T$ gives a straightforward evaluation of the log-likelihood. For instance, the contribution at time t to -2 times the log-likelihood is:

$$-2\log f\left(\mathbf{X_t} \left| \Lambda, \mathbf{X^{t-1}}, \mathbf{Z^T}\right.\right) = c + \log\left(\left|\mathbf{HP_{t|t-1}H'} + \mathbf{R}\right|\right) + $$
$$+ \left(\mathbf{X_t} - \mathbf{H}\xi_{t|t-1} - \mathbf{CZ_t}\right)'\left(\mathbf{HP_{t|t-1}H'} + \mathbf{R}\right)^{-1}\left(\mathbf{X_t} - \mathbf{H}\xi_{t|t-1} - \mathbf{CZ_t}\right)$$

All quantities in this last expression can produced by the Kalman recursions as previously described. Maximisation of the log-likelihood is performed by the E04UCF

routine of the Mark 19 Fortran NAG library. It uses a Newton-Raphson technique for isolating a maximum (see for instance Hamilton, 1994, p.138). Another routine based on a simulated annealing algorithm (see Kirkpatrick et al., 1983) is also proposed mainly for double-checking the results. Differently from Newton-Raphson techniques, simulated annealing does not use first derivative and hence it has the appealing feature of looking for a global optimum. Yet the price to pay is a much longer computing time, and for this reason we suggest its use only for performing checks in difficult cases.

Standard deviations for parameter estimates are obtained using the textbook result (see for instance Hamilton, p. 143) that the asymptotic distribution of the maximum likelihood estimates is well approximated by $\widehat{\Lambda} \sim \mathbf{N}\left(\mathbf{\Lambda}, (\mathbf{T}\Im)^{-1}\right)$, where $\Im$ the information matrix. Program Gap estimates the information matrix using the second derivatives of the log-likelihood at the maximum:

$$\widehat{\Im} = -\frac{1}{T}\frac{\partial^2 \log L}{\partial \Lambda \partial \Lambda'}\Big|_{\Lambda=\widehat{\Lambda}} \tag{17}$$

Sometimes it can happen that $\widehat{\Im}$ comes out as non-positive definite. In this case, Program Gap switches to the outer product of gradient estimate defined by:

$$\widehat{\Im} = \frac{1}{T}\sum_{t=d+1}^{T}\frac{\partial \log f\left(\mathbf{X_t}\Big|\mathbf{\Lambda}, \mathbf{X^{t-1}}, \mathbf{Z^T}\right)}{\partial \Lambda'}\Big|_{\Lambda=\widehat{\Lambda}} \times \frac{\partial \log f\left(\mathbf{X_t}\Big|\mathbf{\Lambda}, \mathbf{X^{t-1}}, \mathbf{Z^T}\right)}{\partial \Lambda}\Big|_{\Lambda=\widehat{\Lambda}} \tag{18}$$

Finally, for model (1)-(8) to give a statistically acceptable description of the series first two moments, the residuals $\mathbf{X_t} - \mathbf{X_{t|t-1}}$ must have white noise properties. This is checked by the Ljung-Box statistics (see Ljung and Box, 1979) that is computed according to:

$$Q\left(m\right) = \left(T-d\right)\left(T-d-2\right)\sum_{k=1}^{m}\frac{r\left(k\right)^2}{T-d-k},$$

where $r\left(k\right)$ is the lag-k residuals autocorrelation and $T-d$ is the number of effective residuals . Program Gap checks the first four autocorrelations, so $m = 4$ in the expression above.

4 Initialising the state covariance matrix with the diffuse Kalman filter

The diffuse Kalman filter serves at the exact initialisation of the Kalman filter when the state vector contains non-stationary elements. It has been introduced by de Jong (1991); a good description can be read in Durbin and Koopman (2001, pp 115-120). The description that we develop below is actually based on Durbin and Koopman's book. Let the initial state vector be

$$\xi_{\mathbf{1}} = \mathbf{a} + \mathbf{A}\delta + \mathbf{B}\eta_{\mathbf{0}}; \qquad \eta_{\mathbf{0}} \sim \mathbf{N}\left(\mathbf{0}, \mathbf{Q_0}\right), \tag{19}$$

where $\mathbf{a}$ is a $k \times n$ vector that is known, δ is a $d \times 1$ vector of unknowns and $\mathbf{A}$ and $\mathbf{B}$ are respectively $n \times d$ and $n \times (n-d)$ matrices with columns are those of the identity matrix. The matrix $\mathbf{A}$ is such that the i-th row of $\mathbf{A}$ loads if the i-th element of the state is non stationary and is made up of zero's otherwise. On the contrary, the i-th row of $\mathbf{B}$ loads the innovation η_0 if the i-th element of the state is stationary and is made up of zero's otherwise. Hence they verify $\mathbf{A'B} = \mathbf{0_{d \times (n-d)}}$. The vector $\mathbf{a}$ is made up of zero's except for the entries of the state vector that are constant, if any. Let us denote by $\mathbf{P}_{*,\delta}$ the matrix made up of the unconditional covariances of state assuming δ is known. It is easily seen that $\mathbf{P}_{*,\delta} = \mathbf{BQ_0B'}$. Conditional on δ the expectation of is:

$$\xi_{\mathbf{1}|\delta} = \mathbf{E}\left[\xi_{\mathbf{1}} \left|\delta\right.\right] = \mathbf{a} + \mathbf{A}\delta, \tag{20}$$

with associated covariance matrix $\mathbf{P_{1|\delta}} = \mathbf{Var}\left[\xi_{\mathbf{1}} \left|\delta\right.\right] = \mathbf{P}_{*,\delta}$. Running the Kalman recursion (13) it is easily seen that the output will be:

$$\xi_{\mathbf{1}|\mathbf{1},\delta} = \mathbf{a} + \mathbf{A}\delta + \mathbf{P_{1|\delta}H'}\left(\mathbf{HP_{1|\delta}H'} + \mathbf{R}\right)^{-1}\left(\mathbf{X_1} - \mathbf{CZ_1} - \mathbf{Ha} - \mathbf{HA}\delta\right), \tag{21}$$

and

$$\begin{aligned} \xi_{\mathbf{2}|\mathbf{1},\delta} &= \mathbf{F}\xi_{\mathbf{1}|\mathbf{1},\delta} \\ &= \mathbf{a_2} + \mathbf{A_2}\delta, \end{aligned} \tag{22}$$

where $\mathbf{a_2}$ and $\mathbf{A_2}$ are obtained by simple re-ordering of the terms in (21). In general, we will write:

$$\xi_{\mathbf{t}|\mathbf{t-1},\delta} = \mathbf{a_t} + \mathbf{A_t}\delta, \tag{23}$$

13

and similarly, if the starting condition would be known, the innovation would be obtained as:

$$v_{t|\delta} = \mathbf{X_t} - \mathbf{CZ_t} - \mathbf{H}\xi_{\mathbf{t}|\mathbf{t}-\mathbf{1},\delta} = v_{\mathbf{a},\mathbf{t}} + \mathbf{V_{A,t}}\delta. \tag{24}$$

The covariance matrix associated with (23) and its update that are obtained by running (14) and (16) with $\mathbf{P}_{*,\delta}$ as starting value will be denoted $\mathbf{P_{t|t-1,\delta}}$ and $\mathbf{P_{t|t,\delta}}$, respectively. Again, for a given δ, the variance of (24) will be denoted by $\mathbf{f_{t|\delta}}$. It is given by:

$$\mathbf{f_{t|\delta}} = \mathbf{HP_{t|t-1,\delta}H'} + \mathbf{R}. \tag{25}$$

Notice that all quantities involved in (23)-(24) can be obtained by running the Kalman filter (13)-(16) of the augmented state $(\mathbf{a_t}, \mathbf{A_t})$ that becomes a matrix with initial state $(\mathbf{a}, \mathbf{A})$ and initial covariance matrix $\mathbf{P}_{*,\delta}$ for each column of the state matrix, while at every time t the observation $\mathbf{X_t}$ in (13)-(16) becomes a vector $(\mathbf{X_t}, \mathbf{0_{1 \times d}})'$.

The diffuse Kalman filter initialisation states that δ is a random variable such that $\delta \sim \mathbf{N}(\mathbf{0}, \mathbf{kI_d})$, with $k \to \infty$. The problem is to find $\xi_{\mathbf{t}+\mathbf{1}|\mathbf{t}} = \mathbf{E}_\delta \left[\xi_{\mathbf{t}+\mathbf{1}|\mathbf{t},\delta} \right]$ and its covariance matrix in that case. Let $\widehat{\delta} = E\left[\delta \,|\mathbf{X^t} \right]$. Since

$$\xi_{\mathbf{t}+\mathbf{1}|\mathbf{t}} = \mathbf{E}_\delta \left[\xi_{\mathbf{t}+\mathbf{1}|\mathbf{t},\delta} \right] = \mathbf{a_{t+1}} + \mathbf{A_{t+1}}\widehat{\delta}, \tag{26}$$

and

$$
\begin{aligned}
\mathbf{P_{t+1|t}} &= E\left[\left(\xi_{\mathbf{t}+\mathbf{1}|\mathbf{t}} - \xi_{\mathbf{t}+\mathbf{1}} \right) \left(\xi_{\mathbf{t}+\mathbf{1}|\mathbf{t}} - \xi_{\mathbf{t}+\mathbf{1}} \right)' \right] \\
&= E\left[\left(\mathbf{a_{t+1}} + \mathbf{A_{t+1}}\widehat{\delta} - \xi_{\mathbf{t}+\mathbf{1}} \right) \left(\mathbf{a_{t+1}} + \mathbf{A_{t+1}}\widehat{\delta} - \xi_{\mathbf{t}+\mathbf{1}} \right)' \right] \\
&= \mathbf{P_{t+1|t,\delta}} + \mathbf{A_{t+1}}\mathbf{Var}\left(\widehat{\delta} \right) \mathbf{A'_{t+1}}.
\end{aligned} \tag{27}
$$

The problem turns out to be that of finding $\widehat{\delta}$ and its variance. Writing

$$
\begin{aligned}
\log p\left(\delta \,|\mathbf{X^t} \right) &= \log p\left(\delta \right) + \log \left(\mathbf{X^t} \,|\delta \right) - \log p\left(\mathbf{X^t} \right) \\
&= -\frac{1}{2}\delta\delta' + \sum_{j=1}^{t} \log p\left(v_{j|\delta} \right) - \log p\left(\mathbf{X^t} \right),
\end{aligned}
$$

and using (24) and (25) in this last expression yields:

$$\log p\left(\delta \,|\mathbf{X^t} \right) = -\frac{1}{2}\delta\delta' - \mathbf{b'_t}\delta - \frac{1}{2}\delta\mathbf{S_{A,t}} + \mathbf{c}, \tag{28}$$

14

where c denotes terms independent of δ and $\mathbf{b_t}$ and $\mathbf{S_{A,t}}$ are given by

$$
\begin{aligned}
\mathbf{b_t} &= \sum_{j=1}^{t} V_{A,j}' \mathbf{f_{j|\delta}^{-1}} v_{\mathbf{a,j}} \\
\mathbf{S_{A,t}} &= \sum_{j=1}^{t} V_{A,j}' \mathbf{f_{j|\delta}^{-1}} \mathbf{V_{A,j}}.
\end{aligned}
\tag{29}
$$

Normality implies that $\widehat{\delta}$ maximises $\log p\left(\delta \mid \mathbf{X^t}\right)$. Solving the first-order conditions on (28) yields

$$
\widehat{\delta} = -\left(\mathbf{S_{A,t}} + \frac{1}{\mathbf{k}}\mathbf{I_d}\right)^{-1}\mathbf{b_t}.
\tag{30}
$$

Evaluating minus the Hessian matrix at the maximum likelihood estimator and inverting gives the covariance matrix

$$
Var\left(\widehat{\delta}\right) = \left(\mathbf{S_{A,t}} + \frac{1}{\mathbf{k}}\mathbf{I_d}\right)^{-1}.
\tag{31}
$$

Finally, taking $k \to \infty$ in (30) and (31) and plugging the results into (26) and (27) yields

$$
\xi_{\mathbf{t+1|t}} = \mathbf{a_{t+1}} - \mathbf{A_{t+1}}\mathbf{S_{A,t}^{-1}}\mathbf{b_t},
\tag{32}
$$

and

$$
\mathbf{P_{t+1|t}} = \mathbf{P_{t+1|\delta}} + \mathbf{A_{t+1}}\mathbf{S_{A,t}^{-1}}\mathbf{A_{t+1}'}.
\tag{33}
$$

Equations (32)-(33) made up de Jong's diffuse Kalman Filter. All quantities can be computed using the augmented Kalman filter, as previously explained, and using (29) for intermediate values. Because $\mathbf{S_{A,t}}$ is singular as long as $t < d$, the expressions (32)-(33) can be evaluated only at $t = d$. The diffuse Kalman filter directly yields $\xi_{\mathbf{t+1|t}}$ and $\mathbf{P_{t+1|t}}$. This means for instance that the first d innovations do not exist. Once $\xi_{\mathbf{t+1|t}}$ and $\mathbf{P_{t+1|t}}$ are available, the augmented Kalman recursions that were necessary for computing (23)-(25) and (29) in the previous steps are collapsed and the usual recursions (13)-(16) are used.

15

Other procedures exist for initialising the Kalman filter when the state vector is not stationary. An alternative to de Jong's approach is to write analytically the initial state covariance as a function of k, to work out $\mathbf{P_{d|d}}(\mathbf{k})$ using the Kalman recursions and to derive its limit as k goes to infinity. It is the analytical counterpart of the algorithm described above. Yet, for a model with more than 30 parameters this analytical approach is untractable.

A third approach puts large numbers into the initial state covariance matrix for infinite variances and run the usual recursions. Although it is still commonly used, most textbook recommend against it because of the numerical instability it creates. For that reason we discarded it.

5 Smoothing the unobserved components

For casting model (1)-(8) into a state space format, it is necessary to include in the state vector the unobservable components X_{1t}^N and X_{1t}^S. For instance in the illustration given at the beginning of chapter 2, the state vector is $\xi_{\mathbf{t}} = (\mathbf{X_{1t}^N}, \mu_{\mathbf{1t}}, \mathbf{X_{1t}^S}, \mathbf{X_{1t-1}^S}, a_t^\pi, a_{t-1}^\pi)'$. Running the Kalman recursions yields the projections $X_{1t|t}^N$ and $X_{1t|t}^S$, but in general knowledge of the future helps in determining current unobserved quantities. Trivially, the current trend of an economy is better known if information about the future is available. Updating the state concurrent estimate $\xi_{\mathbf{t|t}}$ using the information available at time $t+1, \ldots, T$ is known as smoothing. How the incoming information affects the concurrent estimates can be understood from

$$\begin{matrix} \xi_{\mathbf{t}} \\ \xi_{\mathbf{t+1}} \end{matrix} \Big| \mathbf{X^t}, \mathbf{Z^T} \sim \mathbf{N} \left(\begin{matrix} \xi_{\mathbf{t|t}} \\ \xi_{\mathbf{t+1|t}} \end{matrix}, \begin{bmatrix} \mathbf{P_{t|t}} & \mathbf{FP_{t|t}} \\ & \mathbf{P_{t+1|t}} \end{bmatrix} \right),$$

where all the entries are produced by the Kalman recursions (13)-(16). The properties of normal distribution imply that

$$E\left[\xi_{\mathbf{t}} \Big| \xi_{\mathbf{t+1}}, \mathbf{X^t}, \mathbf{Z^T}\right] = \xi_{\mathbf{t|t}} + \mathbf{FP_{t|t}} \mathbf{P_{t+1|t}^{-1}} \left(\xi_{\mathbf{t+1}} - \xi_{\mathbf{t+1|t}}\right). \tag{34}$$

Because $\xi_{\mathbf{t}}$ is uncorrelated to the innovations $\mathbf{v_{t+2}}, \ldots, \mathbf{v_T}$ in the transition equation (11) and to the innovations $\mathbf{u_{t+1}}, \ldots, \mathbf{u_T}$ in the measurement equation (10), it fol-

lows that $E\left[\xi_{\mathbf{t}}|\xi_{\mathbf{t+1}}, \mathbf{X}^{\mathbf{t}}, \mathbf{Z}\right] = E\left[\xi_{\mathbf{t}}|\xi_{\mathbf{t+1}}, \mathbf{X}^{\mathbf{T}}, \mathbf{Z}\right]$. According to the law of iterated expectations

$$E\left[\xi_{\mathbf{t}}\left|\mathbf{X}^{\mathbf{T}}, \mathbf{Z}^{\mathbf{T}}\right.\right] = E\left[E\left[\xi_{\mathbf{t}}\left|\xi_{\mathbf{t+1}}, \mathbf{X}^{\mathbf{T}}, \mathbf{Z}^{\mathbf{T}}\right.\right]\right]$$

and

$$E\left[\xi_{\mathbf{t}}\left|\mathbf{X}^{\mathbf{T}}, \mathbf{Z}^{\mathbf{T}}\right.\right] = \xi_{\mathbf{t}|\mathbf{t}} + \mathbf{FP}_{\mathbf{t}|\mathbf{t}}\,\mathbf{P}_{\mathbf{t+1}|\mathbf{t}}^{-1}\left(\xi_{\mathbf{t+1}|\mathbf{T}} - \xi_{\mathbf{t+1}|\mathbf{t}}\right). \tag{35}$$

Because $\xi_{\mathbf{t}|\mathbf{t}}$ and $\xi_{\mathbf{t+1}|\mathbf{T}} - \xi_{\mathbf{t+1}|\mathbf{t}}$ belong to orthogonal information sets, the associated covariance matrix is:

$$Var\left[\xi_{\mathbf{t}}\left|\mathbf{X}^{\mathbf{T}}, \mathbf{Z}^{\mathbf{T}}\right.\right] = \mathbf{P}_{\mathbf{t}|\mathbf{t}} + \mathbf{FP}_{\mathbf{t}|\mathbf{t}}\,\mathbf{P}_{\mathbf{t+1}|\mathbf{t}}^{-1}\left(\mathbf{P}_{\mathbf{t+1}|\mathbf{T}} - \mathbf{P}_{\mathbf{t+1}|\mathbf{t}}\right)\mathbf{P}_{\mathbf{t+1}|\mathbf{t}}^{-1}\mathbf{P}_{\mathbf{t}|\mathbf{t}}\,\mathbf{F}' \tag{36}$$

The state vector can thus be smoothed at all dates by starting the recursions (35)-(36) at $t = T - 1$. Program Gap implements the fixed-point smoother algorithm described in Harvey (1989, pp 151-153).

17

References

[1] Clark, P.K. (1987), "The Cyclical Comonent of U.S. Economic Activity", *Quarterly Journal of Economics*, 797-814.

[2] de Jong, P. (1991), "The diffuse Kalman filter", *Annals of Statistics*, 19, 1073-1083.

[3] Durbin, J. and Koopman, S.J. (2001), Time Series Analysis by State Space Methods, Oxford University Press: Oxford.

[4] Fuller, W. (1996), Introduction to Statistical Time Series, 2nd edition, John Wiley & Sons: New York.

[5] Hamilton J. (1994), Time Series Analysis, Princeton: Jersey Hall.

[6] Harvey A.C. (1989), Forecasting Structural Time Series Models and the Kalman Filter, Cambridge University Press: Cambridge.

[7] Harvey, A.C. and Jaeger, A. (1993), "Detrending, Stylised Facts and the Business Cycles", *Journal of Applied Econometrics*, 8, 231-247.

[8] Hillmer S.C. , and Tiao G.C. (1982) "An ARIMA-Model-Based Approach to Seasonal Adjustment", *Journal of the American Statistical Association*, 77, 63-70.

[9] Hodrick R.J. and E.C. Prescott (1997), "Post-War U.S. Business Cycle: An Empirical Investigation", *Journal of Money, Credit and Banking*, 29, 1, 1-16.

[10] Kirkpatrick S., Gellat C.D.Jr., and Vecchi M.P. (1983), "Optimisation by Simulated Annealing", *Science*, 220, 671-680.

[11] Kuttner K.N. (1994), "Estimating Potential Output as a Latent Variable", *Journal of Business and Economic Statistics*, 12, 3, 361-368.

[12] Ljung, G. and Box G.E.P., (1978), "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, 65, 297-303.

[13] Maravall A. and Planas C. (1999), "Estimation Errors and the Specification of Unobserved Component Models", *Journal of Econometrics*, 92, 325-353.

[14] Pierce, D.A. (1978), "Seasonal Adjustment When Both Deterministic and Stochastic Seasonality Are Present", in Seasonal Analysis of Economic Time Series, ed. A.Zellner, Washington, U.S.Bureau of the Census, pp. 242-269.

[15] Watson M.W. (1986), "Univariate Detrending Methods with Stochastic Trends", *Journal of Monetary Economics*, 18,49-75.