Program GAP

Technical Description and User-manual

Christophe Planas and Alessandro Rossi [1]

European Commission, Joint Research Centre

Version 5.0, April 2018

[1]Emails: christophe.planas@ec.europa.eu, alessandro.rossi1@ec.europa.eu. We are grateful to Werner Roeger and the members of the Output Gap working group for their comments and suggestions.

Contents

Chapter I

Introduction

GAP is a program developed by the European Commission that implements the EU commonly agreed methodology to estimate the output gap, i.e. the deviation of GDP from its potential. The output gap is one main ingredient in the calculation of the cyclically-adjusted budget balance of the EU Member States. Since 2002 the European Commission applies the Cobb-Douglas production function approach to obtain the output gap from the short-term deviations of labour and total factor productivity (TFP) from their potential. The cycle in unemployment is considered as an unobserved dynamic factor which is common to a labour cost indicator in a Phillips curve relationship, while the cycle in productivity is shared by capacity utilization. The complement of the unemployment cycle makes up the NAWRU. GAP implements two bivariate dynamic factor models to decompose unemployment and productivity into equilibrium or potential plus short-term deviations. Interested readers can find an exhaustive description of the EU commonly agreed methodology in Havik et al. (2014).

This document describes the models, the statistical methodology, and the program interface. Chapter 2 contains three sections that review the model and the methodology. Section 1 focuses on the NAWRU, Section 2 on potential productivity, and

Section 3 describes several additional capabilities of GAP. Chapter 3 contains the user-manual for the Excel interface. Chapter 4 focuses on the Bayesian module.

Installation is made by unzipping the file GAP50.zip. This produces an Excel file, two executable files GAP50.exe file and BGAPINT50.exe, plus some Matlab DLL and figure files which are written into the BIN \WIN32 sub-directory. Users are invited to read the license agreement before running the program. Care must be taken that no blank characters are allowed in the host directory name.

GAP Excel interfaces for estimating the NAWRU and potential TFP in the 28 Member States can be found in the CIRCA web-site following the path European Commission - Economic and Financial Affairs - Output Gaps - Library.

Chapter II

Models and methodology

The two models offered for estimating the NAWRU and potential TFP belong to the class of dynamic factor models. Given the many specificities they are presented separately below.

1 The NAWRU

1.1 Model specification

Unemployment u_t is decomposed into a cycle or gap g_t plus the NAWRU n_t according to:

$$\begin{aligned} u_t &= n_t + g_t \\ \Delta n_t &= \eta_{t-1} + a_{nt} \\ \eta_t &= \eta_{t-1} + a_{\eta t} \\ g_t &= \phi_{g1} g_{t-1} + \phi_{g2} g_{t-2} + a_{gt} \end{aligned} \tag{1.1}$$

where Δ denotes first-difference. The NAWRU n_t follows either a first or second-order random walk process depending on whether the variance of the slope shocks $a_{\eta t}$ is null or strictly positive. The AR(2) fluctuations of g_t drive the evolution of a labour cost indicator w_t through the Phillips curve equation:

$$w_t = \mu_w + \phi_{w1} w_{t-1} + \beta_0 g_t + \beta_1 g_{t-1} + \gamma_w' z_t + a_{wt} \tag{1.2}$$

where the variables in the vector z_t contain exogenous information about labour productivity, terms-of-trade, and the wage share. All shocks, namely $a_{nt}, a_{\eta t}, a_{ct}$, and a_{wt} with variance respectively denoted V_n, V_η, V_c, and V_w, are independent and normally distributed white noises. The model for labour costs is stationary.

The Phillips curve (1.2) is compatible with both backward- and forward-looking expectations: with backward-looking expectations the labour cost indicator corresponds to the change in wage inflation whereas with forward-looking expectations the growth of real unit labour costs is the relevant indicator. The forward-looking assumption implies the restriction $\beta_1 = \beta_0(\phi_{w1} - 0.99)/(0.99\phi_{w1} - 1)\phi_{g2}$, where 0.99 is the assumed discount factor. The NAWRU model (1.1-1.2) is completed with an anchor at time $T + h$ say S_{T+h}, h denoting the horizon beyond the sample end T at which the NAWRU n_t is supposed to converge under the hypothesis of no-policy change; in short it is assumed that $n_{T+h} = S_{T+h}$ (see Atanas et al., 2017). The anchor S_{T+h} is built using a panel regression on structural indicators of the labour market (see Orlandi, 2012).

1.2 Econometric methodology

The likelihood associated to model (1.1)-(1.2) is calculated using state space methods. The model is first cast into state space representation (see Harvey, 1989, Durbin and Koopman, 2001). Since the two models are non-stationary, use is made of the diffuse Kalman filter described in de Jong (1991) and Koopman (1997) which provides the so-called diffuse likelihood.

The parameters of model (1.1)-(1.2) are estimated by maximum likelihood. The optimization is made using a sequential quadratic programming algorithm built upon Nocedal and Wright (2006). Standard deviations for parameter estimates are obtained inverting the Hessian matrix evaluated at the maximum likelihood parameters. Given the observations $u_1^T \equiv (u_1, \cdots, u_T)$, $w_1^T \equiv (w_1, \cdots, w_T)$, and the maximum likelihood estimate say $\hat{\theta}$, the NAWRU and the unemployment gap are estimated as the conditional expectation $n_{t|T} = E(n_t|u_1^T, w_1^T, \hat{\theta})$ and $g_{t|T} = E(g_t|u_1^T, w_1^T, \hat{\theta})$, $t = 1 \cdots, T$, which are calculated using the fixed-point smoother detailed in Harvey (1989). The smoother also gives the mean-squared-errors $V(x_t|u_1^T, w_1^T, \hat{\theta})$ for $x = n, g$. Mean square errors unconditional to the maximum likelihood parameters, i.e. $V(x_t|u_1^T, w_1^T)$, are also calculated using the approximation proposed by Ansley and Kohn (1986). Finally, the anchored estimates $n_{t|T}^a = E(n_t|u_1^T, w_1^T, \hat{\theta}, n_{T+h} = S_{T+h})$ are obtained as:

$$n_{t|T}^a = n_{t|T} + \frac{Cov(n_t, n_{T+h}|u_1^T, w_1^T, \hat{\theta})}{V(n_{T+h}|u_1^T, w_1^T, \hat{\theta})}(S_{T+h} - E(n_{T+h}|u_1^T, w_1^T, \hat{\theta})) \tag{1.3}$$

GAP must be supplied with the anchor S_{T+h} and the time-horizon h to produce the anchored NAWRU.

2 Potential TFP

2.1 Model specification

The logarithm of TFP say tfp_t is similarly decomposed into a trend p_t plus a cycle c_t assuming the model:

$$\begin{aligned} tfp_t &= p_t + c_t \\ \Delta p_t &= \eta_{t-1} + a_{pt} \end{aligned}$$

$$
\begin{aligned}
\eta_t &= \mu_p(1-\rho) + \rho\eta_{t-1} + a_{\eta t} \\
c_t &= \phi_{c1}c_{t-1} + \phi_{c2}c_{t-2} + a_{ct}
\end{aligned}
\tag{2.4}
$$

Potential TFP follows an I(1) damped trend model with average growth μ_p. Like for the unemployment gap, the cycle c_t is assumed to follow an AR(2) process. It is linked to the short-term developments of capacity utilization cu_t through the measurement equation (see Planas, Roeger, and Rossi, 2013):

$$
\begin{aligned}
cu_t &= \mu_{cu} + \beta_{cu}c_t + e_t \\
e_t &= \phi_{cu}e_{t-1} + a_{cut}
\end{aligned}
\tag{2.5}
$$

where e_t is an unobserved idiosyncratic component which can be autocorrelated. All shocks, namely $a_{pt}, a_{\eta t}, a_{ct}$, and a_{et} with variance respectively denoted V_p, V_η, V_c, and V_{cu}, are independent and normally distributed white noises.

2.2 Econometric methodology

The parameters of model (2.4)-(2.5) are estimated in the Bayesian framework. The advantage of the Bayesian approach is that information about model parameters available from macroeconomic theory and empirical studies can be incorporated into the analysis. Eliciting prior information however requires re-parameterizing the model governing the cyclical fluctuations in terms of amplitude A and periodicity τ as in:

$$
c_t = 2\ Acos(2\pi/\tau)c_{t-1} - A^2c_{t-2} + a_{ct} \tag{2.6}
$$

This specification imposes complex autoregressive roots. Obviously this re-parameterization does not apply if a white noise or AR(1) process is specified for the cycle.

GAP offers the following prior distributions:

- Cycle parameters:

$$\begin{aligned}
p(A) &= Beta(a_A, b_A) \\
p\left(\tfrac{\tau-\tau_l}{\tau_u-\tau_l}\right) &= Beta(a_\tau, b_\tau) \\
p(V_c) &= IG(s_{c0}, v_{c0})
\end{aligned}$$

- Damped trends:

$$\begin{aligned}
p(\mu_p) &= N(\mu_{p0}, M_{p0}^{-1}) \\
p(\rho) &= N(\delta_0, M_{\delta 0}^{-1}) \\
p(V_p) &= IG(s_{p0}, v_{p0}) \\
p(V_\mu) &= IG(s_{\mu 0}, v_{\mu 0})
\end{aligned}$$

- Random walk trends:

$$p(\mu_p, V_p) = NIG(\mu_{p0}, M_{p0}^{-1}, s_{p0}, v_{p0})$$

- Capacity utilization:

$$\begin{aligned}
p(\mu_{cu}, \beta_{cu}, V_{cu}) &= NIG(m_{cu0}, M_{cu0}^{-1}, s_{cu0}, v_{cu0}) \\
p(\phi_{cu1}) &= N(\phi_{cu1}, M_{\phi_c u}^{-1}) I_S
\end{aligned}$$

where $Beta(.,.)$ is the Beta distribution, τ_l and τ_u are the lower and upper bounds of τ's support, $IG(\cdot)$ is the inverted-2 Gamma distribution, $N(\cdot)$ is the Normal distribution, and $NIG(\cdot)$ the normal inverted-2-gamma distribution. The IG and the NIG distributions are parameterized as in Bauwens and Lubrano (1999, pp. 292 and 294). All parameters except variances are given a finite support. The hyperparameters of each prior distribution can be tuned using the menu **Priors** in the Bayesian module.

The likelihood function, still calculated using state space methods with diffuse Kalman filter initialization, re-weights the priors to give posterior distributions. These posterior distributions are generally not known in closed-form but samples can be obtained through a Markov Chain Monte Carlo scheme. As simulation routine GAP implements the slice sampler popularized by Neal (2003). Given an auxiliary variable the parameters are sampled one-at-a-time in a Gibbs loop (see for instance Chapter 8 in Robert and Casella, 2004). Posterior draws of the unobserved variables c_t and p_t, $t = 1, \cdots, T$ are then generated given the simulated parameters outside of the Gibbs loop using the simulation smoother devised by Durbin and Koopman (2002). Missing observations are allowed as capacity utilization in EU countries is typically available over short time-periods only.

As output, GAP delivers non-parametric estimates of the posterior distributions $p(c^T|tfp^T, cu^T)$, $p(p^T|tfp^T, cu^T)$, and $p(\eta^T|tfp^T, cu^T)$. Convergence is monitored using Geweke's convergence test (Geweke, 1992). Some statistics about the sampling efficiency are offered like the relative numerical efficiency - see Section 16.1. To infer about the strength of the link between the TFP cycle and capacity utilization, a 90% highest-posterior region on the parameter β_{cu} is also given.

3 Further GAP capabilities

Although mainly used for decomposing unemployment and TFP, GAP has some flexibility. Indeed it can process the following options:

- univariate models;
- white noise or AR(1) process for the cycle;
- random walk, damped trend, I(2), or second-order random walks models for the trend;
- Bayesian analysis of the NAWRU model (1.1)-(1.2);
- maximum likelihood estimation of the TFP model (2.4)-(2.5).

In addition, for one Member State it has appeared necessary to correct a level shift in the TFP series. To accommodate such a pattern GAP has been extended to allow exogenous regressors on the first endogenous series as in:

$$tfp_t \quad = \quad p_t + c_t + \sum_{i=1}^{r} \alpha_i z_{it} \tag{3.7}$$

The α-parameters are given a normal prior distribution. GAP then assigns the exogenous regressors either to the trend or to the cycle depending on the user-choice - see Section 8.2.

Chapter III

User-manual - Excel interface

GAP is made up of an Excel interface, a Matlab-based graphical interface for Bayesian analysis, and a Fortran program that operates all calculations. All code is available in open-source: the Excel settings are processed by the MAIN.BAS file, the Matlab interface is managed by the file BGAPINT50.m, and the Fortran routines are organized by the MAIN.FOR file.

The Excel interface serves to load the data, to specify the model, and to read the output. It is organized in five worksheets: **Data** and **Specifications** control the data and model inputs, **OutputML** and **OutputBayes** show the GAP output, and the fifth worksheet **Help& About** is purely informative. More details follow.

7 Worksheet Data

Users must enter data, i.e. 1st, 2nd, and exogenous series if any, in the following columns:

- First series: observations must be inserted in cells C4 to C(T+3), where T is the sample size.

- Second series: observations must be inserted in cells F4 to F(T+3), where T is the sample size.
- Exogenous series for the first equation must be inserted in cells G4 to G(T+3) until P4 to P(T+3).
- Exogenous series for the second measurement equation must be inserted in cells AK4 to AK(T+3) until AT4 to AT(T+3).

Missing observations at the beginning or at the end of the second series are allowed. Exogenous regressors must be extended with forecasts if GAP is asked to forecast the endogenous series.

In the upper left corner, information can be inserted about **Frequency**, **Starting month/quarter**, and **Starting year**. Once filled, pressing the button **Update time labels** produces dates in column B. These dates are then used in **Specification** for enabling users to select the range of data used for model estimation.

Warning: the second series and the regressors on the second equation should be covariance stationary.

8 Worksheet Specifications

8.1 Estimation and forecasting

- **First observation** Select the first observation for model estimation from a list that is automatically updated according to the time labels in Data.
- **Last observation** Select the last observation for model estimation like above.
- **Number of forecasts** Enter the number of forecasts.

- **Output location** Enter the directory where all output files are to be saved, for instance ”C:\GAP”. Care must be taken that GAP does not accept paths that include a space like for instance ”C:\MY GAP”.

- **Trend anchor / horizon** Enter the anchor value S_{T+h} together with the horizon of convergence h. GAP then delivers both the anchored and non-anchored estimates of the trend and cycle.

8.2 Model specification

Model specification is exclusively controlled by the Excel interface.

- **1st series** Tick the box first series to enter the model for series 1.

- **Trend model** Choose between second-order random walk, first-order random walk and damped trend.

- **Cycle AR order** Choose between 0, 1, 2, and 2 with complex roots.

- **Exogenous to trend/cycle** Enter the number of exogenous variables to be assigned to the trend and to the cycle. These cannot be greater than the number of exogenous series entered in the worksheet Data. The exogenous assigned to the trend must be entered in the first columns, then those assigned to the cycle. The maximum number of regressors is 10.

- **Phillips curve** Tick to fit a bivariate model that includes a second measurement equation such as the Phillips curve (1.2) or the equation (2.5).

- **ARMAX** Tick to insert a autoregressive lag in the second measurement equation as in (1.2).

- **RegAR** Tick to insert a autoregressive lag in the second measurement equation as in (2.5).

- **AR order** Choose between 0, 1 and 2.

- **MA order** Choose between 0 and 1.
- **Backward or forward** Impose a forward looking restriction as in (1.2).
- **# of exogenous** Enter number of exogenous variables on the Phillips curve. The maximum number is 10.
- **Endog 1st series - Intercept** Select Phillips curve to put μ_w in the Phillips curve equation like in (1.2) or Cycle to assign it to the first series cycle. In this case the first series cycle has a non-0 mean equal to μ_w/β_0 if c_t appears in the right-hand-side without lags.
- **Endog 1st series - Cycle** Select lag 0 for having the contemporaneous cycle c_t as regressor of the Phillips curve, lags 0-1 for having both c_t and c_{t-1}, and so on until lag 4 under backward-looking expectations. Only lags 0 and 1 are allowed with forward-looking expectations. Select "None" for no cyclical term in the second equation.

Trend-cycle decompositions are also possible for more than one series - see boxes 2nd series and 3rd series. This extension is not discussed here as it is currently disactivated.

8.3 Parameter constraints

For every parameter, the **Parameter constraints** cells offer the possibility to enter lower and upper bounds. GAP itself imposes a few constraints on variance parameters. In particular, for the variance of the trend and cyclical shocks, the upper bound must be less or equal to 1.2 times the variance of the differenced series. For the second equation residuals variance, the maximum upper bound allowed is 1.2 times the variance of the second series. The program automatically resets the upper bound to the maximum authorized if the value entered is out of the bounds.

The **Set default** button proposes default bounds for all parameters.

9 RUN Maximum likelihood

Estimation is performed when pressing **RUN** on the worksheet **Specifications**. Ticking the box "Maximum likelihood" produces the output described next.

10 Worksheet OutputML

Maximum likelihood estimation generates two output files, **SOL.TXT** and **OUTPUT.TXT**, which are written into the Output directory. Once estimation is completed, users are directed to the worksheet OutputML. Two buttons can be seen in the upper left corner:

- **Edit results** Press to read the model parameters estimates and diagnostics from the **SOL.TXT** file. The parameter estimates are reported together with their standard errors. The following diagnostics are produced:
  - : t-statistics of the parameter weighting the cycle in the second measurement equation.
  - : the first-four autocorrelations of the innovations in each endogenous series with their standard error.
  - : the Ljung-Box statistics computed on these first-four autocorrelations with its p-value.
  - : the "R-squared (one-step-ahead predictions)" which is equal to one minus the ratio of the variance of the innovations to the variance of the endogenous series.

– : the "R-squared (fitted values)" which is equal to one minus the ratio of the variance of the idiosyncratic part projected on contemporaneous information to the variance of the endogenous series.

– : "% of variance explained by common cycle" shows the amount of commonality in the second endogenous series.

– : "% reduction in cycle MSE due to 2nd series" shows the percentage reduction of MSE of the cycle estimate obtained from the bivariate model compared to a univariate decomposition.

- **Graph** Press to plot output series. The following graphs are offered:

  – : Series 1: there are six plots available which show the predicted values obtained as the series minus the innovation, the growth, the trend, the cycle, the anchored trend, and the anchored cycle.

  – : Phillips curve: predicted values - i.e. the series minus the innovation, fitted values - i.e. the series minus the estimated idiosyncratic component $a_{wt|T}$, the idiosyncratic component $a_{wt|T}$, and the Phillips curve cycle.

In the graphics, the unobserved components can be shown either as filtered quantities such as for instance $c_{t|t}$ or as smoothed values $c_{t|T}$ - see cell B11 on the worksheet OutputML.

For confidence bands, the box RMSE offers the possibility either to consider model parameters as given or to compute variances that take into account the uncertainty in model parameters like in Ansley and Kohn (AK, 1986). The confidence level can be set using the box "Conf. lev.".

When the Graph button is first pressed, the following series are loaded in the worksheet Output: the original series with forecasts if any, the root mean square error (RMSE) around these forecasts, the RMSE using AK procedure (AK-RMSE), the smoothed trend together with its RMSE and AK-RMSE, the filtered trend and the filtered cycle with associated RMSE, the first series innovations, the Phillips curve series with its forecasts, the RMSE and AK-RMSE around these forecasts if any, the Phillips curve innovations, the filtered and smoothed values of idiosyncratic component of the Phillips curve, the filtered and smoothed values of the Phillips curve cycle, the fitted values obtained as Phillips curve endogenous series minus the filtered idiosyncratic component, the anchored trend, and the anchored cycle. This output is loaded from the file **OUTPUT.TXT**.

Chapter IV

User-manual - Bayesian module

In the worksheet Specifications, selecting **Bayesian inference** and pressing RUN opens the Bayesian module where the main page shows the following information:

- Workspace either default or workspace filename including of path;
- Output location directory;
- Frequency, sample dates, sample size and number of missing observations;
- Number of forecasts;
- Model specification;
- Series including the exogenous regressors. One can plot each series by clicking on the series name and pressing the right-mouse button.

The model and the data can only be modified in Excel. The Bayesian module serves at eliciting prior distributions, configurating the MCMC sampler, and screening the posterior distributions. A menu bar guides the action. The entries are: File, Priors, MCMC design, RUN, Posterior, and Info that is purely informative. The contents are described in the next sections.

12 File

The File menu option contains five entries:

- **Load workspace** or Ctrl+L for changing workspace.
- **Save workspace** or Ctrl+S.
- **Update output location** to change the output location. This is useful when workspace files are exchanged. Care to save the workspace after the update if you want it to be permanent.
- **Close** or Ctrl+X to return to Excel.

13 Priors

The Priors sub-menu contains four entries:

- **Initialize using Max Likelihood** or Ctrl+A. When activated, the program estimates the model parameters by maximum likelihood and uses the results to propose prior distributions for all parameters.
- **Set** or Ctrl+P for tuning prior distributions - see below. If the option *Initialise using Max Likelihood* has been previously used, the program shows these priors. If *Set* is activated without previous use of *Initialise using Max Likelihood*, then the program proposes prior distributions on the basis of the bounds entered in the worksheet Specification of the Excel interface. In all cases the user can easily modify the prior distributions - see below.
- **Import** for loading priors from file. Either a specific prior file or a workspace file can be used.
- **Export** for saving priors into a file.

When **Set** priors or Ctrl+P is pressed, users are directed to a page with three tabs: Cycle, Trend and Second observational equation. Each tab displays the related equation together with the prior distribution proposed for each parameter appearing in this equation. Range sliders enable users to tune the mean and standard deviation (SD) of each distribution. Once the prior distributions are chosen, **Export hyperparameters** in the menu-bar saves the hyperparameters into the file **hyper.txt** in the output location directory. Notice that tuning is only permitted on the moments and not on the hyperparameters.

Once the prior distributions have been set, the menu-bar option **Save and close** returns to the main menu.

14 MCMC design

The sub-menu contains four entries:

- **Enter seed** Users must enter 0 if the random number generators are to be re-set to a starting point that is hold constant. Otherwise the starting point of the generators is random. Default value is 0.
- **Enter burn-in** This is the number of iterations that are run to initialize the MCMC algorithm. The output of these first simulations is discarded. Default value is 1,000.
- **Enter thinning** Thinning represents the number of simulations produced for one to be recorded. For instance, entering 10 makes GAP recording one output every ten iterations. The larger the thinning, the less correlations are expected in the chain output. Default value is 1.
- **Enter # of recorded simulations** This represents the number of points that are used for estimating the posterior distributions of all quantities of interest. Default is 10,000.

15 RUN

Once priors and the MCMC design have been selected, the Bayesian iterations are launched by pressing the **RUN** button. Run-time information can be seen in the DOS window behind the Main Menu page. When the iterations are ended, users are automatically led to the screening of posterior distributions.

Warning: closing the DOS window or interrupting the program execution closes the Bayesian module.

16 Output

Once GAP has run, the following files are created in the output location directory:

- **ACF.txt** $(4 \times \#\text{ of series}) \times 5$ the first four rows contain the 1%, 25%, the mode, the 75%, the 99% quantiles of the first four autocorrelations of the innovations of the first endogenous series; the fifth to eighth rows contain the same statistics calculated on the innovations of the second endogenous series if present;

- **Param.txt** $\#$ of parameters $\times$ 409 each row contain first 200 points of the support of the posterior distribution followed by the posterior distribution evaluated on these 200 points; the last nine columns contain the lower and upper bounds of the 90%-highest posterior region, the mode, the standard deviation, the numerical standard error, the relative numerical standard error, the p-value of the Geweke convergence test, and the first and fifth autocorrelations of the posterior draws. There are as many rows as estimated parameters.

- **Unobservable.txt** $(\#\text{of observations} + \#\text{of forecasts}) \times 26$ the first column contains the first series included the forecasts, the second to fifth columns the 2.5%, 5%, 95%, and 97.5% quantiles of the series forecasts, the sixth to tenth columns the posterior mean with the 2.5%, 5%, 95%, and 97.5% quantiles of the trend forecasts included, the eleventh to fifteenth columns the posterior mean with the 2.5%, 5%, 95%, and 97.5% quantiles of the cycle foreasts included, the sixteenth to twenty-first columns the second series included forecasts and estimated missing observations together with the 2.5%, 5%, 95%, and 97.5% quantiles plus the innovations posterior means, and the twenty-second to twenty-sixth columns the posterior mean with the 2.5%, 5%, 95%, and 97.5% quantiles of the trend slope extended with forecasts.

- **Marginal.txt** $(\#\text{of observations} + \#\text{of forecasts}) \times 1200$ the row refer to the time index. On each row the first 600 points gives the x-axis for the trend, cycle, and trend growth support, and the following 600 columns give the posterior distribution of the trend, cycle, and trend growth on their respective supports.
- **Marglik.txt** contains the marginal likelihood estimates.

These files are used for obtaining the posterior densities detailed in the next Section.

17 Posteriors

GAP offers graph facilities for visualizing posteriors. All graphs can be exported to postscript files that are written in the **Output Location** directory - see Sub-section 8.1. There are eight entries in the Posterior menu that we detail below.

17.1 Parameters

This page is automatically opened when the Bayesian iterations end or also when Ctrl+R is pressed. The plots are organized through tabs that refer to trend, cycle and Phillips curve equations. For each parameter, the prior (- -) and posterior (—) distributions are shown together with the following statistics:

Mode the posterior mode.

SD the posterior standard deviation.

NSE the numerical standard error of the posterior mean. It is computed as the spectrum at the zero frequency using a Parzen window of length equal to 4% of the size of the posterior sample.

RNE the relative numerical efficiency. It is computed as the ratio of the parameter posterior variance calculated without taking into account the chain correlations, i.e. SD above, to the NSE. The RNE thus indicates the number of drawn required to produce the same numerical accuracy if the draws had been made from an iid sample drawn directly from the posterior distribution. It corresponds to the square root of the inverse of the inefficiency factor. Close to 1 values indicate high efficiency.

rho(1) the first-lag autocorrelation of the recorded simulations;

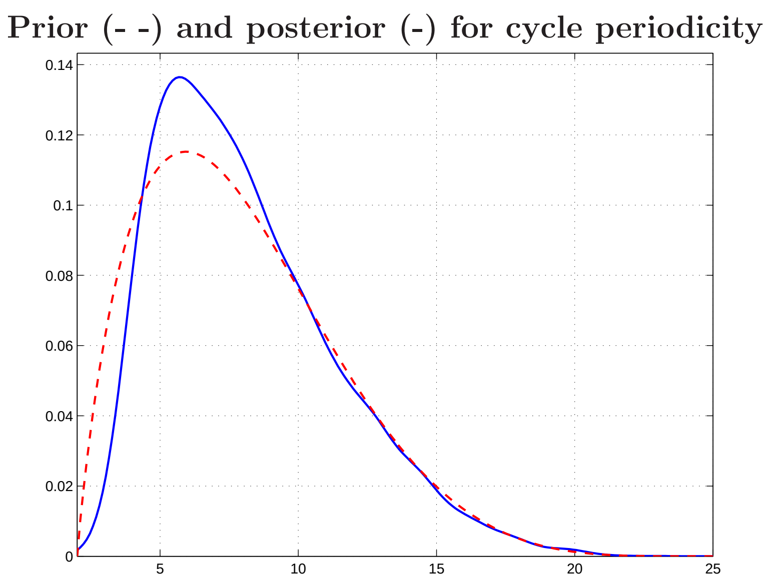
rho(5) the fifth-lag autocorrelation of the recorded simulations;

Geweke p-value the p-value of Geweke test for constancy of the chain mean. Geweke's convergence diagnostic (CD) tests for the constancy of the last 20% points mean compared to the mean of the first 50% points. A p-value lower than 0.05 is an indication of failure to converge.

More details about NSE, RNE and Geweke convergence test can be found in Geweke (1992).

Finally, 90% highest posterior region (HPR) are reported for the second equation parameter β_{cu}.

An **Export graphs** facility is available in the menu-bar of the Parameters window. It writes a post-script file for each subplot into the **Output Location** directory. The filenames correspond to the parameter name, i.e. A.ps for parameter A, tau.ps for τ, Vc.ps for V_c, etc... The graph below gives an example for the cycle periodicity τ:



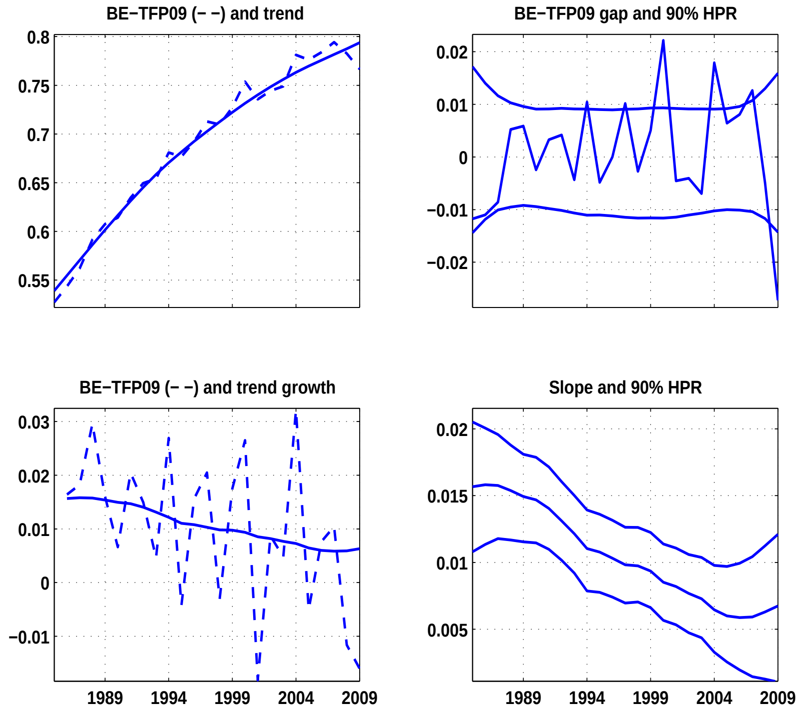
17.2 Unobservable posterior means

This option - also accessible via Ctrl+U - produces four subplots:

- upper left corner: the series and the trend posterior mean are displayed.
- upper right corner: the cycle posterior mean is shown. Confidence bands at 90% centered around 0 are also plotted. These bands are referred to as Highest Posterior Region, HPR in short.
- bottom left corner: the series growth and the trend growth.
- bottom right corner: the slope posterior mean together with 90% confidence bands.

An **Export graph** facility saves the graph in the **unobs.ps** file that is reproduced below for illustration.

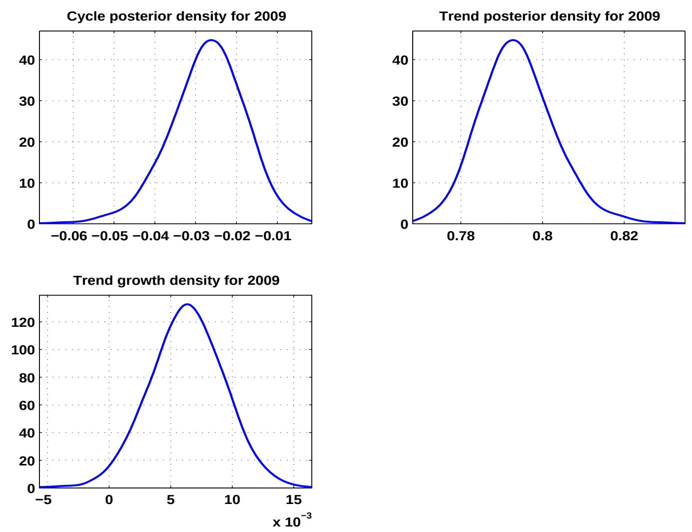
Unobservables posterior mean (-) with 90% confidence bands



17.3 Unobservables marginals

Unobservables marginals shows the posterior distribution of the cycle and of the trend at any given time period. It can be accessed also via Ctrl+M. Users are asked to select the date from a list; clicking on a date produces three subplots that show the posterior distribution of the cycle, trend and trend growth in this date. An **Export graph** facility saves the graph into a file with a name that is related to the time period selected, for instance **cycle2009.ps**, still in the **Output Location** directory. For instance:

Posterior distributions of trend & cycle in 2009



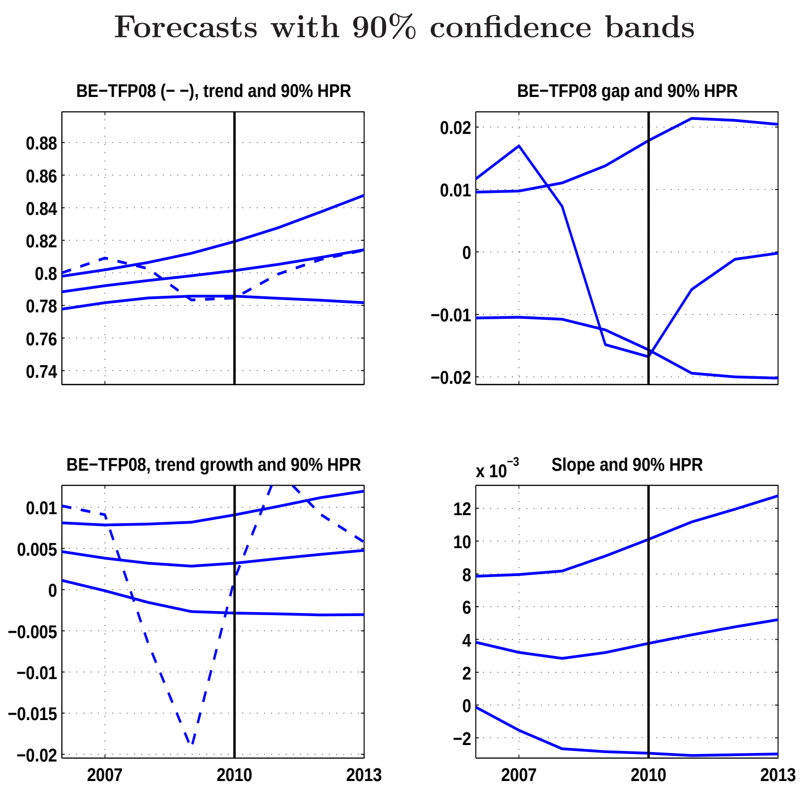
17.4 Forecasts

One figure with four subplots is displayed - also using Ctrl+F:

- upper left corner: the series (- -) and the trend posterior mean are displayed for the last four years and for the forecasting period. The 90% confidence band around the trend forecasts are reported.

- upper right corner: the cycle is shown for the last four years and for the forecasting period together with their 90% confidence band.
- bottom left corner: the series growth and the trend growth.
- bottom right corner: the slope posterior mean together with 90% confidence bands.

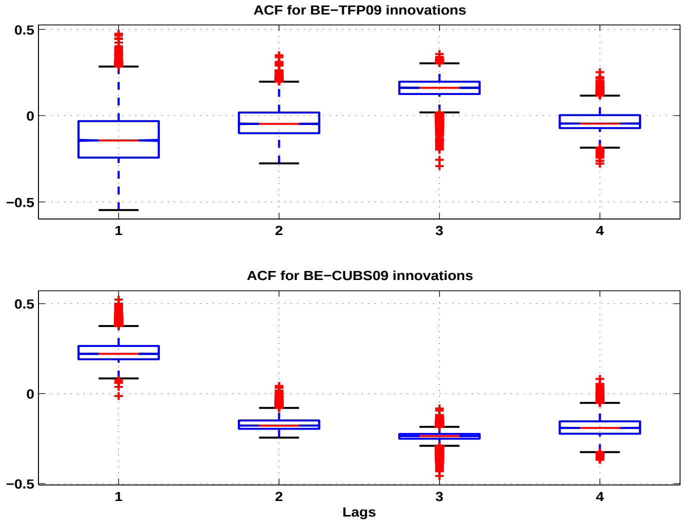
An **Export graph** facility saves the graph in the **fore.ps** file that is reproduced below:



17.5 Diagnostics

The ACF of the residuals is proposed for checking whether the model requires additional dynamics. The four autocorrelations are shown. These autocorrelations are obtained for residuals obtained all along the simulations, so the posterior distribution of the ACF is available. As illustrated in the graph below, GAP displays them as Box-plot. The box bounds represents the first and third quantiles, the mode, together with the 1% and 99% quantiles. A Ctrl-D shortcut to the figure is available.

Posterior distribution of the residuals autocorrelations



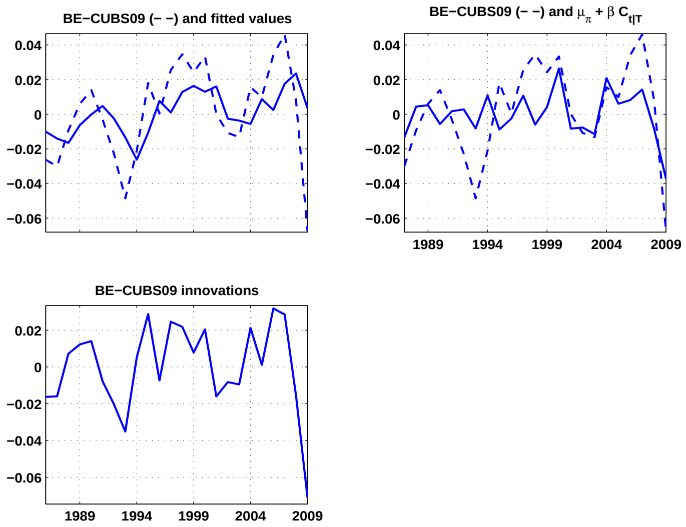
The export graph facility saves the graph as **diag.ps** into the output location directory.

17.6 Second equation fit

This facility - also accessible via Ctrl+C - shows the second series together with its fitted values and the corresponding innovations that are computed using of the parameter posterior mode. A third subplot compares the second series $\Delta\pi_t$ to the constant term plus the smoothed cyclical component, i.e. $\mu_\pi + \beta_0 c_{t|T}$, in Section 2's notations.

An **Export graph** facility saves the figure in the **pcfit.ps** file as illustrated below.

Second equation fit



17.7 Marginal likelihood

GAP implements two types of estimators of the marginal likelihood: the modified harmonic mean estimator (Gelfand and Dey, 1994, Geweke, 1999) and the bridge-sampling estimator (Meng and Wong, 1996). Marginal likelihood - also accessible via Ctrl+S - shows the estimates and the relative standard errors. The file is named Marglik.txt and it is saved in the output location directory.

18 Worksheet OutputBayes in the Excel interface

Once the Bayesian simulations are ended, the output can be also be read from the Excel worksheet **OutputBayes**. A graph facility is proposed with the following possible plots:

- Series 1 plus its forecasts if any;
- Series 1 with trend, both extended with forecasts if any;

- Series 1 cycle, with forecasts;
- Series 2 with forecasts and missing observations if any;
- Series 2 smoothed cycle, for instance $\beta_0 c_t + \beta_1 c_{t-1}$, including of forecasts;
- Series 2 innovations posterior mean.

All unobserved quantities are reported as posterior mean. Quantiles at 2.5%, 5%, 95% and 97.5% are also given so as to build 90% and 95% confidence intervals. When the Graph button is pressed, all relevant quantities are loaded from the **UNOBSERVABLE.TXT** file in the output directory.

19 Source code and the GAP50.nml file

The GAP source code is available in open-source. The Excel settings are managed by the `MAIN.BAS` file. The Matlab Bayesian interface is controlled by `BGAPINT50.m`. All calculations are made by Fortran routines which are organized by the `MAIN.FOR` file.

Pressing **RUN** either on Excel or on the Bayesian module creates the GAP50.nml file which contains the model specification, the parameters bounds, the prior distribution in Bayesian analysis, and the observations. The information is grouped in namelists as summarized below:

- `&ssm` contains the model equations for the endogenous and state variables together with:

  `maxintord` maximum order of integration of the state variables;
  `nonstatstate` number of non-stationary state elements;
  `estimation` ML or BA depending on whether maximum likelihood or Bayesian analysis is chosen.

- `&prior` contains the parameters lower and upper bounds as well as the density type and hyperparameters for Bayesian analysis like for instance:

```
lab(1) = A       pdf(1) = BE hyp(1, 1) = 1.2 1.2 0 0.99
lab(2) = Tau     pdf(2) = BE hyp(1, 2) = 2.0 2.0 2.1 40
lab(3) = Vuslope pdf(3) = IG hyp(1, 3) = 2.E-4 6 0 1.E-04
```

  where `BE` and `IG` stand for Beta and inverted-2 Gamma distribution. The first two numbers following `hyp(1, 1)` refer to the distribution hyper-parameters and the next two to the lower and upper bound of the distribution.

- `&mcmc` contains settings for the MCMC simulations;

- `&GAP` contains the output location string, the anchor `ANCH`, as well as several other parameters;

- `&dataset` contains the number of forecasts requested `nf` plus the observations after the keyword `obs =`. The matrix of observations includes the exogenous regressors and the constant if any.

20 References

Ansley C.F. and Kohn R. (1986), "Prediction mean squared error for state space models with estimated parameters", *Biometrika*, 73, 467-73.

Atanas H., Planas C., Roeger W. and Rossi A. (2017), " NAWRU estimation using structural labour market indicators", European Economy Discussion Papers, no. 69, European Commission: Brussels.

Bauwens L., Lubrano M. and Richard J. (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press.

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

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

Durbin, J. and Koopman, S.J. (2002), " A simple and efficient simulation smoother for state space time series analysis", *Biometrika*, 89, 3, pp. 603615.

Gelfand A. and Dey D., (1994). "Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B*, 56, 501-514.

Geweke J. (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments", in Bayesian Statistics, 4, by Bernardo J., Berger J., Dawid A. and Smith A. (eds), Oxford: Oxford University Press.

Geweke, J., (1999). "Using simulation methods for Bayesian econometric models: inference, development, and communication". *Econometric Reviews*, 18, 1, 1-73.

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

Havik, K., K. Mc Morrow, F. Orlandi, C. Planas, R. Raciborski, W. Roeger, A. Rossi, A. Thum-Thysen and V. Vandermeulen (2014), "The production function methodology for calculating potential growth rates & output gaps", European Economy, Economic Paper nr 535.

Koopman S.J. (1997), "Exact initial Kalman filtering and smoothing for non-stationary time series models", *Journal of the American Statistical Association*, 92, 440, 1630-1638.

Meng X.L. and Wong, W.H., (1996), "Simulating ratios of normalizing constants via a simple identity: a theoretical exploration", *Statistica Sinica*, 6, 831-860.

Neal R. (2003), "Slice sampling", *Annals of Statistics*, 31, 3, 705-767.

Nocedal J. and Wright, S.J. (2006), *Numerical Optimization*, 2nd ed., Springer: New York.

Orlandi F. (2012), "Structural unemployment and its determinants in the EU countries", European Economy Economic Paper, No 455, European Commission: Brussels.

Planas C., Roeger W. and A. Rossi (2013), "The information content of capacity utilization for detrending total factor productivity", *Journal of Economic Dynamic and Control*, 37, 577-590.

Robert C.P. and Casella G. (2004), *Monte Carlo Statistical Methods*, 2nd ed., Springer: New York.