NB toolbox
Theory, algorithms and implementations

Kenneth Sæterhagen Paulsen and Atle Loneland

January 2, 2023

Abstract

In this document we document the algorithms and implementation details that are base for the code that is available in NB toolbox (Paulsen, 2021).

Contents

1 Basic econometrics

The supported tests and methods will now be describe in turn.

1.1 Autocorrelation

The theoretical sample autocorrelation is given by

$$\rho_k = \frac{\sum_{t=k+1}^{T}(Y_t - \overline{Y})(Y_{t-k} - \overline{Y})}{\sum_{t=1}^{T}(Y_t - \overline{Y})^2}. \tag{1}$$

It is estimated by utilizing the fast Fourier transform (FFT) and the inverse fast Fourier transform ($IFFT$)

$$f = FFT(Y) \tag{2}$$

$$F = f * conj(f) \tag{3}$$

$$acf = real(normalized(IFFT(F))) \tag{4}$$

$$\rho_k = acf(k). \tag{5}$$

If the sample autocorrelation decays geometrically it is a indication of a low order ARIMA(p,x,0) process, while if the sample autocorrelation decay fast to zero it is a sign of a low order ARIMA(0,x,q) process.

The partial autocorrelation is the coefficient on Y_{t-k} when Y_t is estimated on Y_{t-1}, Y_{t-2}, ... ,Y_{t-k}, see equation 6, where ϱ_k will be the partial autocorrelation at lag k. This is a partial correlation as the sample autocorrelation is cleaned by the first $k-1$ autocorrelations. The partial autocorrelation will for an ARIMA(p,x,0) process cut off at lag p, while it will decay slowly to zero for an ARIMA(0,x,q) process.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \varrho_k Y_{t-k} \tag{6}$$

1.1.1 Error bands

There are different options on how to calculate the error bands. The first option is asymptotic error bands. For sample autocorrelation these are calculated as

$$\rho_1 \pm 1/\sqrt{T} * \Phi(1 - \alpha/2) \tag{7}$$

$$\rho_k \pm \sqrt{(1 + 2 * acf(1:k)' * acf(1:k))/T} * \Phi(1 - \alpha/2) \quad for \quad k = 2, ..., p. \tag{8}$$

While for the partial autocorrelation these are calculated by

$$\varrho_k \pm 1/\sqrt{T - k - 1} * \Phi(1 - \alpha/2), \tag{9}$$

where α is the selected significance level.

An alternative way is to estimate the error bands in two steps. The first step will be to estimate an ARIMA model on the series of the data set. At this stage you can select the maximum order of the AR and MA terms to be used for the selection routine, and you can choose the estimation method. By selecting a

criteria a model is selected. See section 1.4 for more on the information criterion. The selected model is then used to simulate 1000 bootstrapped series. For each of these draws the sample autocorrelation or partial autocorrelation is recalculated. Then the $\alpha/2$*100 and the (1-$\alpha/2$)*100 percentiles of this distribution is used as the error band. Be aware that this method assumes a stationary series. The error bands with this procedure should be taken with care, as the data generating process may not be estimated well.

It is also possible to use different non-parametric bootstrap methods. These are only added so it is possible to analyze the different bootstrapping methods when used to bootstrap parameters of models. I.e. are the residuals autocorrelated, if so does the bootstrapping method take care of this? The answer is yes, if the calculated autocorrelation is centered inside the error bands.

1.2 Unit root tests

To test for non-stationarity of time-series is an important task in economics, as non-stationary variables may produce spurious results when not taking into account properly. There are many suggested ways to test for a non-stationary time-series, but let's start with some theory. Assume we have the following time-series

$$y_t = \rho y_{t-1} + BX_t + u_t, \tag{10}$$

where y_t is the time-series to be tested for non-stationarity and has size 1 x 1, y_{t-1} is its lagged value, X_t is included exogenous regressors, as constant and/or time-trend, and has size N x 1, u_t is the residual of the regression and has size 1 x 1 and B is a 1 x N vector of the parameters to be estimated. ρ is the parameter of interest here. If $|\rho| < 1$ the series y_t is stationary, otherwise not. To test for a unit root means to test the null hypothesis $\rho = 1$ against the alternative hypothesis $\rho < 1$.[1] The methods for unit root testing is described in the following section.

1.2.1 Augmented Dickey-Fuller (ADF) test

Let us define $\alpha = 1 - \rho$ then the ADF test amounts to test the null hypothesis $\alpha = 0$ against the alternative hypothesis $\alpha < 0$ in the test regression

$$\triangle y_t = \rho y_{t-1} + BX_t + \beta_1 \triangle y_{t-1} + \beta_2 \triangle y_{t-2} + \ldots + \beta_1 \triangle y_{t-p} + u_t. \tag{11}$$

It is assumed that y_t follows a AR(p) process.[2] The test statistic is

$$adf = \frac{\tilde{\alpha}}{se(\tilde{\alpha})}, \tag{12}$$

where $\tilde{\alpha}$ is the OLS estimate of α. The ADF test does not follow the normal t-distribution, but can be compared against the critical values given in MacKinnon (1996). The ADF test is robust to ignored MA terms in 11, as long as sufficient lagged difference terms are included. There are two problems with the ADF test:

1. You need to choose the exogenous variables to include in you test. To select the wrong model will reduce the power of the test.
2. You need to find the number of lagged variables. Either manually or use a automatic model selection algorithm using an information criteria.

[1] If the null hypotheses is $\rho = -1$ vs $\rho > -1$ test $-y_t$ with the normal hypothesis instead.
[2] See section 6 for more on ARIMA models.

1.2.2 Phillips-Perron test

This test uses the regression

$$y_t = \rho y_{t-1} + BX_t + u_t. \tag{13}$$

The Phillips-Perron test statistic is adjusting for possible autocorrelation in the residual, and is given by

$$pp = \sqrt{\frac{\gamma_0}{\lambda}} \frac{\tilde{\rho}}{se(\tilde{\rho})} - T\frac{\lambda - \gamma_0}{2\tilde{\sigma}\sqrt{\lambda}}, \tag{14}$$

where $\tilde{\sigma} = \sqrt{\frac{u'u}{T-K}}$, K is the number of estimated parameter, T is the sample length and u is constructed by stacking u_t' over time. λ is an estimator of the residual spectrum at frequency zero, see section 1.5, and $\gamma_0 = \frac{u'u}{T}$. MacKinnon (1996) lower-tail critical values are provided.

1.3 Cointegration tests

Two tests for cointegration is available: the Engle-Granger test and the Johansen test.

1.3.1 Engle-Granger test

The Engle-Granger test is a two step procedure to test for a single cointegrated vector. The first stage amounts to estimate

$$Y_t = BX_t + u_t, \tag{15}$$

where Y_t has size 1 x 1 and X_t has size N x 1. B has size 1 x N, while u_t has size 1 x 1. Then the second stage is to test for a unit root in the residual u_t by an ADF test, see section 1.2. The null hypotheses is that the vector $\tilde{Y}_t = \begin{bmatrix} Y_t \\ X_t \end{bmatrix}$ is not cointegrated. For more on this approach see section 19.2 of Hamilton(1994).

1.3.2 Johansen test

Let Y_t be a M x 1 vector of the dependent variables of the model. Then if the hypothesis is that Y_t follow a VAR(p) in levels, then it can be written equivalently as

$$\Delta Y_t = A_1 \Delta Y_{t-1} + A_2 \Delta Y_{t-2} + \ldots + A_{p-1}\Delta Y_{t-p+1} + \zeta Y_{t-1} + BX_t + u_t. \tag{16}$$

Then A_i for all i and ζ has size 1 x M. X_t is the exogenous variables of the model and has size N x 1, while B has size 1 x N. u_t is the residual of the regression, and has size M x 1.

Step 1: Estimate the following equations by OLS

$$\Delta Y_t = \pi_1 \Delta Y_{t-1} + \pi_2 \Delta Y_{t-2} + \ldots + \pi_{p-1}\Delta Y_{t-p+1} + \beta X_t + \varepsilon_t \tag{17}$$

$$Y_{t-1} = \varrho_1 \Delta Y_{t-1} + \varrho_2 \Delta Y_{t-2} + \ldots + \varrho_{p-1}\Delta Y_{t-p+1} + \alpha X_t + \nu_t. \tag{18}$$

Step 2: Calculate canonical correlations:

$$\Sigma_{\nu\nu} = \frac{1}{T}\sum_{t=1}^{T} \nu_t' \nu_t \tag{19}$$

$$\Sigma_{\varepsilon\varepsilon} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t' \varepsilon_t \tag{20}$$

$$\Sigma_{\nu\varepsilon} = \frac{1}{T}\sum_{t=1}^{T} \nu_t' \varepsilon_t \tag{21}$$

$$\Sigma_{\varepsilon\nu} = \Sigma_{\nu\varepsilon}'. \tag{22}$$

Let us define

$$\Sigma = \Sigma_{\nu\nu}^{-1}\Sigma_{\nu\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{\varepsilon\nu}. \tag{23}$$

Step 3: Eigenvalues

Find the eigenvalues of Σ and order them as $\lambda_1 > \lambda_2 > \ldots > \lambda_M$.

Step 4: Test statistic

To test the null hypothesis of exactly h cointegrated relations vs the alternative hypothesis of M cointegrated relations you can use a likelihood ration test

$$traceTest = -T\sum_{i=h+1}^{M} log(1-\lambda_i). \tag{24}$$

This is the trace statistic. Critical values for this test can be found inMacKinnon *et al.* (1999).

To test the null hypothesis of exactly h cointegrated relations vs the alternative hypothesis of h+1 cointegrated relations you can use a likelihood ration test

$$maxEigenvalueTest = -Tlog(1-\lambda_{h+1}). \tag{25}$$

This is the maximum eigenvalue statistic. Critical values for this test can be found in MacKinnon *et al.* (1999).

To learn more about the Johansen test see the sections 20.2 and 20.3 of Hamilton and Press (1994).

1.4 Information criterion

Information criterions are used to select the best model out of a set of models. You can choose between the information criterion:

Akaike information criterion

$infoCrit = -2 * (\frac{logLikelihood}{T}) + 2 * \frac{K}{T}$

Corrected Akaike information criterion

$infoCrit = -2 * (\frac{logLikelihood}{T}) + 2 * \frac{K}{T} * \frac{T}{T-K-1}$

Modified Akaike information criterion

$infoCrit = -2 * (\frac{logLikelihood}{T}) + 2 * \frac{K+\kappa}{T}$

Schwarz information criterion

$infoCrit = -2 * (\frac{logLikelihood}{T}) + 2 * \frac{Klog(T)}{T}$

Modified Schwarz information criterion

$$infoCrit = -2 * (\frac{logLikelihood}{T}) + 2 * \frac{(K+\kappa)log(T)}{T}$$

Hannan and Quinn information criterion

$$infoCrit = -2 * (\frac{logLikelihood}{T}) + 2 * \frac{Klog(log(T))}{T}$$

Modified Hannan and Quinn information criterion

$$infoCrit = -2 * (\frac{logLikelihood}{T}) + 2 * \frac{(K+\kappa)log(log(T))}{T}$$

Where T is the number of periods of the estimation sample, K is the number of estimated parameters of the equation or the system of equations. The log likelihood is calculated as

$$logLikelihood = \begin{cases} -\frac{T}{2}(1 + log(2\pi) + log(\tilde{\sigma}^2)) & \text{for single equations} \\ -\frac{T}{2}(M(1 + log(2\pi)) + log(|\Omega|)) & \text{for systems of equations ,} \end{cases} \tag{26}$$

where $\tilde{\sigma}^2 = \frac{u'u}{T}$, $\Omega = \frac{u'u}{T-K}$, u_t is the residual of the regression(s) and has size M x 1, u is constructed by stacking u_t' vertically over time and M is the number of equations of the system.

The modified information criteria will only differ from the normal when selecting between different lags of the equation or system of equations. κ is then given by

$$\kappa = |\rho|^2 \frac{|\underline{Y}_{t-1}'\underline{Y}_{t-1}|}{|u'u|}. \tag{27}$$

Y_t is the dependent variables of the model and has size M x 1, $\underline{Y}_{t-1}$ is constructed by stacking Y_t' over time and lag it one period. ρ is a diagonal matrix with the coefficient in front of the lagged dependent variable of each equation along the main diagonal.

1.5 Frequency zero spectrum estimation

To estimate the frequency zero spectrum a kernel sum-of-covariances estimator is used

$$\lambda = \gamma_0 + 2\sum_{j=1}^{l-1} \gamma_j K(j/l), \tag{28}$$

where T is the sample length, l is the bandwidth parameter, which truncates lags in the covariance weighting and K is the kernel function. See section 1.6 for the supported kernel functions.

$$\gamma_j = \frac{\sum_{t=j+1}^{T} u_t u_{t-j}}{T}, \tag{29}$$

where u_t is the series of interest. The bandwidth parameter is of considerable importance. There are three ways of finding/setting this. The first option is to set it manually. The second option is to use Newey-West (1994) data-based automatic bandwidth selection method, see section 1.5.1. While the third option is to use Andrews (1991) data-based automatic bandwidth selection method, see section 1.5.2.

1.5.1 Newey-West

First step is to calculate the autocovariances

$$\gamma_j = \frac{\sum_{t=max(1-j,1)}^{min(T,T-j)} u_t u_{t+j}}{T} \text{ for } j \in [-n, n], \tag{30}$$

where $n = \lceil \frac{4T}{100}^{2/9} \rceil$.[3] Then the optimal bandwidth is given by

$$bandwidth = \left\lceil \varphi \left(T \left(\frac{s_q}{s_0} \right)^2 \right)^{\frac{1}{2q+1}} \right\rceil, \tag{31}$$

where $s_0 = \sum_{-n}^{n} \gamma(j)$, $s_q = \sum_{-n}^{n} j^q \gamma(j)$. q and φ will depend on the used kernel:

$$q = \begin{cases} 1 & \text{Barlett kernel} \\ 2 & \text{Parzen kernel} \\ 2 & \text{Quadratic spectral kernel} \end{cases} \tag{32}$$

$$\varphi = \begin{cases} 1.1447 & \text{Barlett kernel} \\ 2.6614 & \text{Parzen kernel} \\ 1.3221 & \text{Quadratic spectral kernel} \end{cases} \tag{33}$$

1.5.2 Andrews

Andrews(1991) method uses the regression

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{34}$$

where u_t is the series of interest. Then the bandwidth can be found by the formula

$$bandwidth = \left\lceil \varphi \left(T \left(\alpha \right)^2 \right)^{\frac{1}{2q+1}} \right\rceil, \tag{35}$$

where

$$\alpha = \begin{cases} \frac{4\tilde{\rho}^2}{(1-\tilde{\rho})^2(1+\tilde{\rho})^2} & \text{If } q = 1 \\ \frac{4\tilde{\rho}^2}{(1-\tilde{\rho})^4} & \text{otherwise.} \end{cases} \tag{36}$$

See section 1.5.1 for the values of φ and q, which will depend on the kernel function used. $\tilde{\rho}$ is the OLS estimate of ρ.

1.6 Kernel functions

You can choose between the Bartlett, Parzen and quadratic spectral kernel:

Bartlett	$f(x) = \begin{cases} 1 - \lvert x \rvert & \text{if } \lvert x \rvert \leq 1 \\ 0 & \text{otherwise} \end{cases}$ (37)
Parzen	$f(x) = \begin{cases} 1 - \lvert x \rvert & \text{if } 0 \leq \lvert x \rvert \leq \frac{1}{2} \\ 2(1 - \lvert x \rvert)^3 & \text{if } \frac{1}{2} < \lvert x \rvert \leq 1 \\ 0 & \text{otherwise} \end{cases}$ (38)
Quadratic spectral	$f(x) = \frac{25}{12\pi^2 x^2 \left(\frac{sin(6\pi x/5)}{6\pi/5} - cos(6\pi/5) \right)}$ (39)

[3] $\lceil \rceil$ is the ceil operator.

1.7 Bootstrapping

Sometimes it is needed to do classical inference on tests or analysis where asymptotic properties are not known. Examples are impulse responses, forecast error variance decomposition, shock decomposition, inference on factor models etc. To construct confidence intervals on these objects we need to draw from the unknown distribution of the parameters. One way to do this is by bootstrapping. We will first go through the basic case. Lets start with a description of a general single equation model

$$y_t = \beta X_t + u_t, \tag{40}$$

where y_t is the dependent variable and has size 1 x 1, β is the parameters to estimate and has size 1 x N, X_t is the regressors and has size N x 1 and u_t is the residual of the regression with size 1 x 1. Lets first define

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} \quad X = \begin{bmatrix} X_1' \\ \vdots \\ X_T' \end{bmatrix}, \tag{41}$$

where T is the end of the estimation sample. Then the basic bootstrapping approach is to do the following:

1. Estimate 40.
2. Re-sample the residual with replacement to construct Q simulated series of u. Abbreviate each simulated series by $\tilde{u}^i$.
3. For each simulated series of u construct a simulated series of y by $\tilde{y}_t^i = \hat{\beta} X_t + \tilde{u}_t^i$, where i is the simulation index and $\hat{\beta}$ is the estimate of β.
4. Re-estimate equation 40 where y_t is substituted with $\tilde{y}_t^i$ for all i. You will then end up with Q different estimates of β. I.e. $\hat{\beta}^i$ for $i \in [1:Q]$. This means that you have an estimate of the distributions of all the parameters of the vector β.
5. Use the distributions of the parameters to construct simulations of the object of interest, and use the $\alpha/2$ and $1-\alpha/2$ percentiles to calculate two-sided α confidence intervals.

When asked for a bootstrap method use the Bootstrap option to choose the above algorithm. When is this method valid? Only when the classical assumption of OLS is satisfied. I.e. that the residual is distributed

$$u_t \sim N(0, \sigma) \text{ and } E[u_t u_{t-j}] = 0 \text{ for } j \neq 0, \tag{42}$$

i.e. the residual does not display any heteroscedasticity and/or autocorrelation. In the next sections we will discuss valid bootstrap method when one or both of these assumption is violated.

1.7.1 Wild bootstrap

When the residuals show heteroscedasticity it means that

$$u_t \sim N(0, \sigma_t) \text{ and } E[u_t u_{t-j}] = 0 \text{ for } j \neq 0, \tag{43}$$

i.e. the variance of the residual are no longer constant through time. The solution to this problem is to use Wild bootstrap. The algorithm is as before, but instead of just re-sampling with replacement, the residuals are multiplied by a mean zero random variable

$$\tilde{u}_t = u_t \varepsilon_t. \tag{44}$$

We assume that ε_t follows the Rademacher distribution

$$\varepsilon_t = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2}. \end{cases} \tag{45}$$

For a more general discussion read Davidson and Monticini (2014).

1.7.2 Block bootstrap

When the residuals show autocorrelation it means that

$$E[u_t u_{t-j}] \neq 0 \text{ for some } j \neq 0, \tag{46}$$

i.e. the residual are correlated across time. One solution to deal with such a problem is to use random length block bootstrap. Again it is only point 2 in the above algorithm that change. The (approximate) mean block length is selected by the following criteria[4]

$$l_m = T^{\frac{1}{3}}. \tag{47}$$

Then p random length intervals are selected such that

$$T = \sum_{j=1}^{p} l_j, \tag{48}$$

where the l_j is drawn as follows

$$l_j = \lceil 2 l_m \tau_j \rceil \tag{49}$$

$$\tau_j \sim U(0,1), \tag{50}$$

where the operator $\lceil \rceil$ is the ceil operator and U(0,1) is the uniform distribution on 0 to 1. Then the random starting point of the intervals are drawn from the $U(0,T)$ distribution. Again as many draws to be sure that condition 48 is satisfied. This is repeated Q times to get Q bootstrapped residual series, $\tilde{u}$. The algorithm used here is a slightly simplified version of the one presented in Hounyo (2014). I.e. no tapering and it does not correct for heteroscedasticity.

1.7.3 Wild block bootstrap

When both heteroscedasticity and autocorrelation is present the bootstrapped residuals are constructed as in section 1.7.2, with only one difference. Each selected interval is multiplied by a mean zero random variable, as in the case of the Wild bootstrap. Let u_{l_i} be one of the selected blocks of the algorithm given in section 1.7.2. Then the final draw for that block is constructed as follows

$$\tilde{u}_{l_i} = u_{l_i} \varepsilon_t. \tag{51}$$

It is assumed that ε_t follows the Rademacher distribution

$$\varepsilon_t = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2}. \end{cases} \tag{52}$$

[4] Because of the use of the ceil function later on, the mean block length is an under-predictor of the mean of the block length actually used

The algorithm used here is a slightly simplified version of the one presented in Hounyo (2014). I.e. no tapering.

2 State-space representation

Many models used in economics can be formed into a state-space representation. It consists of two matrix equations. The observation equation linking the observables to the state variables is assumed to be on the form

$$Y_t = d + H_t X_t + G Z_t + v_t. \tag{53}$$

Y_t is the observable variables with size M x 1, X_t is the state variables with size N x 1, H_t is the potentially time-varying observation matrix with size M x N, d is the constant with size M x 1 and v_t is the measurement error with size M x 1, Z_t is the exogenous variables of the observation equation with size W x 1. The measurement error is assumed to be normally distributed with covariance matrix R, where R has size M x M,[5] i.e.

$$v_t \sim N(0, R). \tag{54}$$

The state equation linking the current state of the state variables with its own lags and some exogenous disturbances is on the form

$$X_t = ss_t + A_t(X_{t-1} - ss_t) + C_t u_t. \tag{55}$$

A_t is the state transition matrix, that may vary over time, with size N x N, C_t is the input matrix, that may vary over time, with size N x S, ss_t is the potentially time varying steady-state of the state variables, u_t is the exogenous disturbances with size S x 1. The disturbances is assumed to be normally distributed with covariance matrix I,[6] i.e

$$u_t \sim N(0, I). \tag{56}$$

If some elements of u_t does not have an impact on the uncertainty of the model, then the corresponding elements of C_{t-1} is zero. Please see Hamilton and Press (1994) section 13.1 for a more thorough description of the state-space representation.

2.1 Kalman filter

The Kalman filter can be used to get estimates of state variables, or the unobservable variables, given the observable variables and the parameter values of the model. The filter assumes that the model is stationary.

Let us start out with some definitions. $X_{t|t-1} = E_{t-1}[X_t]$ is the expectation of X_t given information on the observables up until time $t-1$, while $X_{t|t} = E_t[X_t]$ is the expectation of X_t given information on the observables up until time t.

First we can use equation 53 to predict Y_t using $X_{t|t-1}$

$$Y_{t|t-1} = d + H_t X_{t|t-1} + G Z_t. \tag{57}$$

As $E_{t-1}[v_t] = 0$. To get a measure of the forecast error we can use equations 53 and 57 to get

[5] It is also assumed that the measurement error is uncorrelated across time.

[6] It is also assumed that the exogenous disturbances is uncorrelated across time.

$$\begin{aligned} F_t &= E[(Y_t - Y_{t|t-1})(Y_t - Y_{t|t-1})'] \\ &= E[H_t(X_t - X_{t|t-1})(X_t - X_{t|t-1})'H_t'] + E[v_t v_t'] \\ &= HP_{t|t-1}H' + R, \end{aligned} \tag{58}$$

where $P_{t|t-1}$ is the error when forecasting X_t given information on the observables up until time $t-1$. We need F_t as we want to update the projection of $X_{t|t-1}$ given the new information on Y_t

$$\begin{aligned} X_{t|t} &= X_{t|t-1} + E[(X_t - X_{t|t-1})(Y_t - Y_{t|t-1})']F_t^{-1}(Y_t - Y_{t|t-1}) \\ &= X_{t|t-1} + P_{t|t-1}H_t'F_t^{-1}(Y_t - d - H_t X_{t|t-1}), \end{aligned} \tag{59}$$

where we in line 2 have used equation 57 and

$$\begin{aligned} E[(X_t - X_{t|t-1})(Y_t - Y_{t|t-1})'] &= E[(X_t - X_{t|t-1})(H_t(X_t - X_{t|t-1}) + v_t)'] \\ &= P_{t|t-1}H_t', \end{aligned} \tag{60}$$

The associated error is

$$\begin{aligned} P_{t|t} &= E[(X_t - X_{t|t})(X_t - X_{t|t})'] \\ &= P_{t|t-1} - P_{t|t-1}H_t'F_t^{-1}H_t P_{t|t-1}. \end{aligned} \tag{61}$$

But we are interested in $X_{t+1|t}$ and $P_{t+1|t}$. These we can find by first noting that

$$X_{t+1|t} = ss_{t+1} + A_{t+1}(X_{t|t} - ss_{t+1}). \tag{62}$$

By substituting 59 into 62 we get

$$X_{t+1|t} = ss_{t+1} + A_{t+1}(X_{t|t-1} - ss_{t+1}) + K_t\nu_t, \tag{63}$$

where we have defined $\nu_t = Y_t - d - H_t X_{t|t-1}$ and $K_t = A_{t+1}P_{t|t-1}H_t'F_t^{-1}$. K_t is called the Kalman gain. With the associated error

$$\begin{aligned} P_{t+1|t} &= A_{t+1}P_{t|t}A_{t+1}' + C_{t+1}C_{t+1} \\ &= A_{t+1}(P_{t|t-1} - P_{t|t-1}H_t'F_t^{-1}H_t P_{t|t-1})A_{t+1}' + C_{t+1}C_{t+1}' \\ &= (A_{t+1}P_{t|t-1} - A_{t+1}P_{t|t-1}H_t'F_t^{-1}H_t P_{t|t-1})A_{t+1}' + C_{t+1}C_{t+1}' \\ &= (A_{t+1}P_{t|t-1}A_{t+1}' - A_{t+1}P_{t|t-1}H_t'K_t') + C_{t+1}C_{t+1}' \\ &= A_{t+1}P_{t|t-1}(A_{t+1}' - H_t'K_t') + C_{t+1}C_{t+1}'. \end{aligned} \tag{64}$$

The starting values of the filter is set to the unconditional mean of X_1 and P_1

$$X_{1|0} = E[X_1] = ss_1 \tag{65}$$

$$vec(P_{1|0}) = E[X_1X_1'] = (I_{M^2} - A_1 \otimes A_1)^{-1}vec(C_1C_1'). \tag{66}$$

See Hamilton and Press (1994) section 13.2 for a more thorough description of the Kalman filter and its

properties.

2.2 Kalman smoother

In contrast to the Kalman filter the Kalman smoother uses all the information in the observable variables to estimate the unobservable variables. The first part of the Kalman smoother is to run through the Kalman Filter. Then by a backward recursion on the following equation you can get the smoothed estimates

$$X_{t-1|T} = X_{t-1|t-1} + J_{t-1}(X_{t|T} - X_{t|t-1}) \text{ for } t = T, \ldots, 2 \tag{67}$$

$$J_{t-1} = P_{t-1|t-1}A_t' P_{t|t-1}^{-1}. \tag{68}$$

See Hamilton and Press (1994) section 13.6 for a more thorough description of the Kalman smoother and its properties. The algorithm implemented in the code follows the smoothing steps:

$$r_{t-1} = A_t' r_t, \tag{69}$$

$$r_{t-1}^i = r_t^i + F_t^{-1}\nu_t - K_t' r_t, \tag{70}$$

$$X_{t|T} = X_{t|t-1} + P_t r_{t-1}, \tag{71}$$

where $r_T = 0$ and r_t^i refers to the elements of r_t that is restricted to the observed variables only. Smoothed estimate of u_t can be found from using

$$u_{t|T} = C_t' r_t. \tag{72}$$

2.3 Evaluate the likelihood

The likelihood can be calculated by use of the Kalman filter. From section 2.1 we know that the distribution of Y_t given all information up until $t-1$ is given by

$$Y_t | I(t-1) \sim N(\tilde{Y}_t, F_t). \tag{73}$$

That means that the likelihood of observation t is given by

$$f_{Y_t|I(t-1)}(Y_t|I(t-1)) = (2\pi)^{-N/2}|F_t|^{-1/2} exp(-\frac{1}{2}\nu_t'(F_t)^{-1}\nu_t). \tag{74}$$

And by summing over the number of observations (T) we get

$$likelihood = \sum_{t=1}^{T} f_{Y_t|I(t-1)}(Y_t|I(t-1)). \tag{75}$$

Then to estimate the parameters of the model you need to use a parameter grid or another optimizing algorithm to find the maximum likelihood over the parameter space. See Hamilton and Press (1994) section 13.4 for a more thorough description on how to use the Kalman filter to find the maximum likelihood estimates.

2.4 Diffuse filter

If the model in equation 55 contains unit roots, i.e. has eigenvalues outside the unit circle, then you must start the Kalman filter with a few diffuse steps. Let TD be the number of periods the filter is in the diffuse state. Then as long as $t \leq TD$, the filter follows the steps:

$$F_t^{inf} = HP_{t|t-1}^{inf}H', \tag{76}$$

$$F_t^{*} = HP_{t|t-1}^{*}H', \tag{77}$$

$$K_t^{inf} = A_{t+1}P_{t|t-1}^{inf}H'(F_t^{inf})^{-1}, \tag{78}$$

$$K_t^{*} = A_{t+1}P_{t|t-1}^{*}H'(F_t^{inf})^{-1} - K_t^{inf}F_t^{*}(F_t^{inf})^{-1}, \tag{79}$$

$$X_{t+1|t} = ss_{t+1} + A_{t+1}(X_{t|t-1} - ss_{t+1}) + K_t^{inf}\nu_{t+1}, \tag{80}$$

$$P_{t+1|t}^{inf} = A_{t+1}P_{t|t-1}^{inf}(A_{t+1}' - H'K_t^{inf\prime}. \tag{81}$$

$$P_{t+1|t}^{*} = A_{t+1}P_{t|t-1}^{*}(A_{t+1}' - H'K_t^{inf\prime}) - A_{t+1}P_{t|t-1}^{inf}K_t^{*\prime}). \tag{82}$$

The diffuse stage continue until $rank(P_{t|t-1}^{inf}) = 0$.[7]

The smoothing steps for $t \leq TD$ are:

$$r_{t-1}^0 = A_t'r_t^0, \tag{83}$$

$$r_{t-1}^1 = A_t'r_t^1, \tag{84}$$

$$r_{t-1}^{1,i} = r_t^{1,i} + (F_t^{inf})^{-1}\nu_t - K_t^{*\prime}r_t^0 - K_t^{inf\prime}r_t^1, \tag{85}$$

$$r_{t-1}^{0,i} = r_t^{0,i} - K_t^{inf\prime}r_t^0, \tag{86}$$

$$X_{t|T} = X_{t|t-1} + P_tr_{t-1}^0 + P_tr_{t-1}^1, \tag{87}$$

where $r_{TD}^0 = 0$, $r_{TD}^1 = r_{TD}$ and $r_t^{X,i}$ refers to the elements of r_t^X that is restricted to the observed variables only (for $X \in \{0, 1\}$).

The likelihood of observation t is calculated as

$$f_{Y_t|I(t-1)}(Y_t|I(t-1)) = (2\pi)^{-N/2}|F_t^{inf}|^{-1/2}. \tag{88}$$

3 Single equation, time-series

A general single equation model can be written as

[7] Which is true in the algorithm as long as all singular values of $P_{t|t-1}^{inf}$ is less than a selected tolerance level.

$$Y_t = B_1 X_t + B_2 Z_t + u_t, \tag{89}$$

where Y_t is the dependent variables of the model with size M x 1, X_t is the exogenous variables of the model with size N x 1, Z_t is the endogenous variables of the model with size Q x 1, u_t is the residuals of the regression with size M x 1, B_1 has size M x N and B_2 has size M x Q.

3.1 OLS

The OLS estimator assumes that no endogenous variables is included in the model. For the OLS estimator to be valid we need the following to be true:

1. The conditional distribution of u_t given X_t has mean zero. I.e. $E[u_t|X_t] = 0$
2. Y_t and X_t are independently and identically distributed (IID).
3. Large outliers are unlikely. I.e. X_t and u_t have finite fourth moment
4. $var[u_t|X_t] = \sigma^2$. I.e. homoscedastic standard errors. Can be relaxed if robust standard errors are used. See section 3.13.
5. The conditional distribution of u_t given X_t is normal. I.e. $u_t|X_t \sim N(0, \sigma^2)$.
6. X_t have full rank. I.e. there is no multicollinearity.

Let T be the number of observations, then the OLS estimator is given by

$$B_1^{OLS} = (X'X)^{-1}XY, \tag{90}$$

where $X = \begin{bmatrix} X_1 \\ \vdots \\ X_T \end{bmatrix}$ and $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix}$.

3.2 TSLS

TSLS stands for two stage least squares. To get rid of the endogeneity problem it is possible to use instruments. TSLS is a way to combine many instruments in an optimal way. The first stage amounts to regress the endogenous variables on its instruments.

$$Z_{q,t} = \alpha + GW_{q,t} + \varepsilon_{q,t}, \tag{91}$$

where $q \in [1, Q]$, $W_{q,t}$ is the selected instruments for the endogenous variable $Z_{q,t}$ with size R x 1, $\varepsilon_{i,t}$ is the residual with the classical OLS residual properties, α is the constant and G has size 1 x R.

$W_{q,t}$ must satisfy two conditions to be valid instruments for $Z_{q,t}$

1. $corr(Z_{q,t}, W_{q,t}) \neq 0$
1. $corr(u_t, W_{q,t}) = 0$.

The first condition means that the instrument is relevant, while the second assumption secures that the instrument is exogenous in relation to the dependent variable(s) of the model. From the first stage estimation you can form the instrumented part of the endogenous variable by:

$$\tilde{Z}_{q,t} = \tilde{\alpha} + \tilde{G}W_{q,t}, \tag{92}$$

where $\tilde{\alpha}$ and $\tilde{G}$ is the OLS estimates of α and G. By construction we know that $corr(u_t, \tilde{Z}_{q,t}) = 0$ and $corr(Z_{q,t}, \tilde{Z}_{q,t}) \neq 0$, so $\tilde{Z}_{q,t}$ is a valid instrument for $Z_{q,t}$, and we can in the second stage estimate (by OLS)

$$Y_t = B_1 X_t + B_2 \tilde{Z}_t + u_t, \tag{93}$$

where $\tilde{Z}_t$ is formed by stacking the $Z_{q,t}$s vertically for $q \in [1, Q]$.

3.3 Ridge

In this section we assume that Y_t only consist of one variable, i.e. $M = 1$. Ridge estimation solves the constrained optimization problem

$$min_{B_1} \left\{ \sum_{t=1}^{T} (Y_t - B_1 X_t)^2 \right\}, \tag{94}$$

s.t.

$$\sum_{j=1}^{N} \left(B_1^j \right)^2 < \lambda. \tag{95}$$

λ is a prespecified free parameter that determines the degree of regularization. Let T be the number of observations, then the Ridge estimator is given by

$$B_1^{Ridge} = (X'X + \lambda I)^{-1} XY, \tag{96}$$

where $X = \begin{bmatrix} X_1 \\ \vdots \\ X_T \end{bmatrix}$ and $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix}$.

3.4 LASSO

In this section we assume that Y_t only consist of one variable, i.e. $M = 1$. LASSO (least absolute shrinkage and selection operator) has the objective to solve

$$min_{B_1} \left\{ \sum_{t=1}^{T} (Y_t - B_1 X_t)^2 \right\}, \tag{97}$$

s.t.

$$\sum_{j=1}^{N} |B_1^j| < \lambda, \tag{98}$$

where T be the number of observations and λ is a prespecified free parameter that determines the degree of regularization. We use the local linearization and active set method proposed in Osborne *et al.* (2000) to solve the LASSO objective.

3.5 Estimation results

After estimation there exist many statistics and tests. In this section some of them are described.

3.6 R-squared

Calculated as $R^2 = 1 - \frac{\widehat{\sigma}^2}{var(Y)}$, where $\widehat{\sigma}^2$ is the estimated residual variance.

3.7 Adjusted R-squared

Calculated as $R^2 = 1 - \frac{\widehat{\sigma}^2/(T-N-Q)}{var(Y)/(T-1)}$, where $\widehat{\sigma}^2$ is the estimated residual variance.

3.8 S.E. of regression

The standard error of the regression is calculated as $\sqrt{\frac{\sum_{t=1}^{T} u_t^2}{T-N-Q}}$.

3.9 Sum squared residuals

Calculated as $\sum_{t=1}^{T} u_t^2$.

3.10 Log likelihood

For a single equation the log likelihood is calculated as $-\frac{T}{2}(1 + log(2\pi) + log(\tilde{\sigma}^2))$. Where $\tilde{\sigma}^2 = \frac{u'u}{T}$ and u is constructed by stacking u_t' over time.

For a system of multiple equations it is calculated as $-\frac{T}{2}(M(1 + log(2\pi)) + log(|\Omega|))$. Where $\Omega = \frac{u'u}{T-N-Q}$. u is constructed by stacking u_t' over time, and has size T x M.

3.11 F-statistics

Tests the null hypotheses that all parameters except the constant is 0. Will return nan with models without a constant. See section 3.17.8 for a more general description of the F-statistics.

3.12 P-value(F-statistics)

See section 3.17.8.

3.13 Standard errors of regression coefficients

The discussion refers to the regression 89. Lets define u by stacking u_t' vertically over time, Y by stacking Y_t' vertically over time and $\tilde{X}$ by stacking $\begin{bmatrix} X_t' & Z_t' \end{bmatrix}$ vertically over time. The parameters of interest is $\beta = \begin{bmatrix} B_1' \\ B_2' \end{bmatrix}$ and has size K x M, where $K = N + Q$. The regression 89 can then be written as

$$Y = \tilde{X}\beta + u. \tag{99}$$

3.13.1 Homoscedasticity only

$$SE(\beta) = \frac{u'u}{T-K}(\tilde{X}'\tilde{X})^{-1}. \tag{100}$$

The standard error calculated in this way can only be used when the residual of the regression does not show heteroscedasticity and/or autocorrelation.

3.13.2 White heteroscedasticity robust

$$SE_W(\beta) = \frac{T}{T-K}(\tilde{X}'\tilde{X})^{-1}(u^2 * X)'X(\tilde{X}'\tilde{X})^{-1}, \tag{101}$$

where $*$ is the element-wise multiplication operator. The standard error calculated in this way can be used when the residuals of the regression show heteroscedasticity, but no autocorrelation. If the residuals do not exhibit heteroscedasticity the standard error constructed in this way reduce the strength of the t-statistic compared to the homoscedasticity only standard errors.

3.13.3 Newey-West heteroscedasticity and autocorrelation robust

$$SE_{NW}(\beta) = \frac{T}{T-K}(\tilde{X}'\tilde{X})^{-1}\tilde{\Sigma}(\tilde{X}'\tilde{X})^{-1} \tag{102}$$

$$\tilde{\Sigma}_i = \frac{T}{T-K}\left\{\sum_{t=1}^{T} u_{i,t}^2 X_t' X_t + \sum_{v=1}^{l-1}\left(2K(v/l)\sum_{t=v+1}^{T}(u_{i,t}u_{i,t-v}X_t'X_t)\right)\right\}. \tag{103}$$

i is the equation index, and $i \in [1, M]$. We then have:

$$\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \tilde{\Sigma}_M \end{bmatrix} \tag{104}$$

l is the bandwidth, and is found by $l = \lfloor 4(T/100)^{\frac{2}{9}} \rfloor$.[8] The standard error calculated in this way is robust to both heteroscedasticity and autocorrelation. On the other hand, if the residuals don't exhibit autocorrelation the standard errors constructed in this way reduce the strength of the t-statistic compared to the homoscedasticity only standard errors or the white heteroscedasticity robust standard errors.

3.14 T-statistic

Calculated by $t = \frac{\widehat{\beta}}{SE_X(\widehat{\beta})}$, where $\widehat{\beta}$ is the estimated parameter of β.

3.15 P-value

Calculated as $p\_value = 2\Phi(-|t|)$, where Φ is the t-distribution with T-N-Q degrees of freedom and t is calculated as in section 3.14.

3.16 Automatic model selection

See the papers:

1. Sucarrat, G. and Escribano, A. (2011): Automated Model Selection in Finance: General-to-Specific Modelling of the Mean and Volatility Specifications
2. Hendry, D. F. and Krolzig, H.-M. (2005): The Properties of Automatic Gets Modeling. Economic Journal 115, C32 to C61.
3. Hendry, D. F. and Krolzig, H.-M. (2001): Automatic Econometric Model Selection using PcGets. London: Timberlake Consultants Press.

[8] $\lfloor\rfloor$ is the floor operator

3.17 Test statistics

The supported test statistics will be described in the following subsection. Some are automatically reported in the estimation printout in NB toolbox (Paulsen, 2021).

3.17.1 Arch test

The Arch test will test for autoregressive conditional heteroscedasticity (Arch) in the residual. The test statistic is

$$archTest = (T-p)\frac{\sum_{t=p+1}^{T}\widehat{v}_t^2}{\sum_{t=p+1}^{T}v_t^2}, \tag{105}$$

where p is the number of lags to include in the regression below, $v_t = u_t^2$ and $\widehat{v}_t$ is formed by the predicted values of the regression

$$v_t = c + a_1 v_{t-1} + a_2 v_{t-2} + \ldots + a_p v_{t-p} + \varepsilon_t. \tag{106}$$

The p-value of the test can be found by $p\_value = 1 - \chi^2(archTest, p)$.[9] The null hypothesis is no Arch effects.

3.17.2 Autocorrelation test

The Godfrey autocorrelation test will test for autocorrelation in the residuals. The test statistic is

$$autoTest = (T-p)\frac{\sum_{t=p+1}^{T}\widehat{u}_t^2}{\sum_{t=p+1}^{T}u_t^2}, \tag{107}$$

where p is the number of lags to include in the regression below and $\widehat{u}_t$ is formed by the predicted values of the regression

$$u_t = c + a_1 u_{t-1} + a_2 u_{t-2} + \ldots + a_p u_{t-p} + \varepsilon_t. \tag{108}$$

The p-value of the test can be found by $p\_value = 1 - \chi^2(autoTest, p)$.[10] The null hypothesis is no autocorrelated residuals.

3.17.3 Breusch-Godfrey test

The Breusch-Godfrey autocorrelation test will test for autocorrelation in the residuals. The test statistic is

$$BGTest = (T-p)\frac{\sum_{t=p+1}^{T}\widehat{u}_t^2}{\sum_{t=p+1}^{T}u_t^2}, \tag{109}$$

where p is the number of lags to include in the regression below and $\widehat{u}_t$ is formed by the predicted values of the regression

$$u_t = c + a_1 u_{t-1} + a_2 u_{t-2} + \ldots + a_p u_{t-p} + BX_t + \varepsilon_t. \tag{110}$$

X_t is the exogenous regressors from the original regression, and has size K x 1. The p-value of the test can be found by $p\_value = 1 - \chi^2(BGTest, p)$.[11] The null hypothesis is no autocorrelated residuals.

[9] This holds only asymptotically.
[10] This holds only asymptotically.
[11] This holds only asymptotically.

3.17.4 Breusch-Pagan test

As the White test this test can be used to test the null hypothesis of no heteroscedasticity. There are two main differences. The first is that the Breusch-Pagan test only tests for heteroscedasticity which is linear in the regressors, so failing to find evidence of heteroscedasticity with the Breusch-Pagan test doesn't rule out a nonlinear relationship between the independent variable(s) and the error variance. The second difference is that the Breusch-Pagan test implemented corrects for non-normality of the residuals.[12]

Let us define W as a T x K matrix of the observation of X_t'[13], $u = [u_1; \cdots ; u_T]$ of size T x 1, $\varepsilon = [u_1^2; \cdots ; u_T^2]$ of size T x 1, $\iota = [1; \cdots ; 1]$ of size T x 1 and $\bar{u} = \frac{u' * u}{T}$. Then the test statistic is given by

$$BPTest = \frac{1}{V} \left[g'W(W'W)^{-1}W'g \right], \tag{111}$$

where $g = \varepsilon - \bar{u}\iota$. And

$$V = \frac{1}{T} \sum_{t=1}^{T} \left[u_t^2 - \bar{u} \right]^2. \tag{112}$$

This test is asymptotically distributed as χ^2 with $K - 1$ degrees of freedom.

3.17.5 Chow breakpoint test

To test for breaks in the parameter estimates a Chow test can be used. The test statistic is given as

$$chowTest = \frac{(SSR - SSR1 - SSR2)/(N + Q)}{(SSR1 + SSR2)/(T - N - Q)}, \tag{113}$$

where SSR is the sum of squared residual from the full regression, $SSR1$ is the sum of squared residual from estimating the model on the first subsample and $SSR2$ is the sum of squared residual from estimating the model on the second subsample. The p-value is given by $p\_value = \Theta(chowTest, N + Q, T - 2(N + Q))$, where Θ is the F-distribution.

3.17.6 Durbin-Watson statistic

The Durbin-Watson (DW) statistic can be used to find out if the residuals of a regression are autocorrelated. The test statistics is

$$dw = \frac{\sum_{t=2}^{T}(u_t - u_{t-1})}{\sum_{t=1}^{T} u_t^2}. \tag{114}$$

To find the tests critical value just search in google for; durbin watson critical values. There are three main problems with the DW test statistic:

1. The distribution of the DW test depends on the included regressors.
2. If lagged dependent variables are included in the regressors the test statistic is no longer valid.
3. The DW-statistic can only be used to the test the null hypothesis of no autocorrelation against the alternative hypothesis of first-order autocorrelation.

[12] See Koenker(1981) and Koenker and Bassett (1982).

[13] If there is no constant included in X_t it will be added.

3.17.7 Durbin-Wu-Hausman test

This test can be used to test if a regressor is exogenous or if more regressors are jointly exogenous, i.e. not correlated with the residual. We folow Wu (1973). The equation of interest is

$$Y_t = B_1 X_t + B_2 Z_t + u_t, \tag{115}$$

where Y_t is the dependent variable of the model with size 1 x 1, X_t is the exogenous variables of the model with size N x 1, Z_t is the possibly endogenous variables of the model with size Q x 1, u_t is the residuals of the regression with size 1 x 1, B_1 has size 1 x N and B_2 has size 1 x Q.

Let's say we have a set of instruments $W_{q,t}$ for each of the Q possibly endogenous variables. See section 3.2 for more on the conditions the instruments must satisfy. Let then $\tilde{Z}_t$ be defined as in section 3.2, i.e. the instrumented possibly endogenous variables. Then the approach suggested by Wu uses an auxiliary regression on the form

$$Y_t = B_1 X_t + B_2 Z_t + D\tilde{Z}_t + v_t. \tag{116}$$

If the coefficients in D are insignificant by a F-test, then you cannot reject the null hypothesis that the variables in Z_t are exogenous. The test statistic is assumed to be distributed as $F(Q, T-N-Q)$.

3.17.8 F-test

The joint F-test that all regression coefficients equal to 0, except the coefficient of the constant, can be simplified as

$$F = \frac{\frac{RSS_1 - RSS_2}{N+Q-1}}{\frac{RSS_2}{T-N-Q}}, \tag{117}$$

where RSS_1 is the restricted sum of squares, RSS_2 is the unrestricted sum of squares and $\widehat{\sigma}^2 = \frac{u'u}{T}$. The p-value reported is given by $p\_value = \Theta(F, N+Q-1, T-N-Q)$, where Θ is the F-distribution.

A general F-statistic testing any linear restriction on the form $Ab = c$ can be calculated as

$$F = \frac{1}{S}(Ab_i - c)'\tilde{\sigma}_i^2 A(X'X)^{-1}A'(Ab - c), \tag{118}$$

where S is the number of restrictions and A has size S x N + Q. $b = \begin{bmatrix} B_1' \\ B_2' \end{bmatrix}$ is the estimated parameters of the model and has size N + Q x M and b_i is the parameters of the equation i for some $i \in [1, M]$, which means the ith column of b. c is the tested value of the linear restriction. X is the regressors of estimated equation stacked vertically over time[14] and $\tilde{\sigma}_i^2$ is the residual variance of equation i. The p-value of the F-test can be found by $p\_value = \Theta(F, S, T-N-Q)$, where Θ is the F-distribution.

3.17.9 Normality test

One of the assumption of OLS estimation is normality of the standard errors. There are many types of normality test, but one such test is

$$normTest = \frac{T-k+1}{6}\left(skewness(u)\right)^2 + \frac{T-k+1}{24}\left(kurtosis(u) - 3\right)^2, \tag{119}$$

where u is the time-series to test with length T and k is the number of estimated parameters of the model if we are dealing with residuals of a regression, otherwise $k = 0$. The p-value of the test can be found by

[14] All equations has the same regressors, so there is no need for a subscript i here.

$p\_value = 1 - \chi^2(normTest, 2)$. The null hypothesis is normality of the residuals. This is the Jarque-Bera test.

3.17.10 Sargan-Hansen test

The Sargan-Hansen test is a test for valid instruments. I.e. the instruments are not correlated with the residual. Let's say we have a system of equations, but we are only interested in estimating the first one

$$Y_t = B_1 X_t + B_2 Z_t + u_t \tag{120}$$

$$Z_{q,t} = B_{q,1}\tilde{X}_{q,t} + B_{q,2}Y_t + B_{q,3}\tilde{Z}_{q,t} + v_{q,t}, \tag{121}$$

where Y_t is the dependent variable of equation 120 with size 1 x 1, X_t is the exogenous variables of equation 120 with size N x 1, Z_t is the endogenous variables in relation to equation 120 with size Q x 1, u_t is the residuals of equation 120 with size 1 x 1. $\tilde{X}_{q,t}$ is the exogenous variables related to the qth endogenous variable with size R_q x 1, while $\tilde{Z}_{q,t}$ is Z_t where $Z_{q,t}$ is removed. B_1 has size 1 x N, B_2 has size 1 x Q, $B_{q,1}$ has size 1 x R_q, $B_{q,1}$ has size 1 x 1 and $B_{q,3}$ has size 1 x $Q-1$. Finally $v_{q,t}$ is the residuals of qth equation with size 1 x 1.

To perform the Sargan-Hansen test we need to estimate an auxiliary regression

$$\hat{u}_t = D\bar{X}_t + \varepsilon_t, \tag{122}$$

where $\hat{u}_t$ is the estimated residual from estimation of 120 with TSLS, and $\bar{X}_t$ is a vector with size S x 1 including all the unique exogenous variables from equation 120 and the selected instruments for all the variables in Z_t. Let the number of variables included in $\tilde{X}_{q,t}$ but not included in X_t be abbreviated by o_q for each $q \in [1, Q]$.[15] Then as long as $o_q > 1$ for all q the system is said to be overidentified. If the system is not overidentified this test is not valid!

The test statistic is constructed as $R^2_{IV}T$. Where R^2_{IV} is the R-squared calculated from the estimation of equation 122 by OLS. The test statistic is distributed as χ^2 with $O = \sum_{q=1}^{Q} o_q$ degrees of freedom. The null hypothesis is that the instruments are valid.

3.17.11 White test

The White (1980) statistic can be used to test for heteroscedasticity. Lets say we estimate (by OLS)[16]

$$Y_t = c + BX_t + u_t. \tag{123}$$

X_t has size K x 1, and B has size 1 x K. Let $\hat{u}_t$ be the OLS residual from this regression. Then to test for heteroscedasticity an auxiliary regression is formed

$$\hat{u}_t^2 = C_0 + C_1 X_t + C_2 \tilde{X}_t + \varepsilon_t, \tag{124}$$

where $\tilde{X}_t$ contain all the squared and cross terms of the regressors. Then the White test is

$$whiteTest = R^2 T, \tag{125}$$

where R^2 is the R-squared from the auxiliary regression. The white test is asymptotically distributed as χ^2 with K degrees of freedom. The null hypothesis is no heteroscedasticity.

[15] These are the potential instruments.

[16] Here X_t does not contain the constant term as before, and is separated out for clarity (c).

4 Step ahead model

A general step ahead model can be written as

$$Y_{t+h} = B_h X_t + u_t^h, \tag{126}$$

where $h > 0$, Y_{t+h} is the dependent variables at different lead of the model with size M x 1, X_t is the exogenous variables of the model with size N x 1, u_t^h is the residuals of the regression with size M x 1 and B_h has size M x N.

The reason we call this a step ahead model is that we can produce direct forecast using this model the number of selected periods ahead.

4.1 Estimation

The model is estimated with OLS.

5 MIDAS

A general mixed data sampling (MIDAS) model can be written as

$$Y_{t+h} = A_h Y_{t+h-1} + \sum_{i=1}^{K} \sum_{j=0}^{L_i} B_{h,j}^i X_{tm_i-j}^i + u_t^h, \tag{127}$$

where $h \in \{1, ..., H\}$, Y_{t+h} is the dependent variables at different lead h of the model, and A_h is the autoregressive parameter to estimate. $X_{tm_i}^i \forall i \in \{1, ..., K\}$ is the exogenous variables of the model, observed at the observations tm_i at higher or equal frequency as the dependent variable. u_t^h is the residuals of the regression with the normal assumptions and $B_{h,j}^i$ is the $H \times \sum_{i=1}^{K} L_i$ parameters of the exogenous variables to estimate.

In NB toolbox you can therefore add as many exogenous variables as right-hand side variables, and they can have any frequency lower than or equal to the dependent variable. The lowest frequency supported is weekly.

5.1 Estimation

Let us assume we have T observation on the dependent variable, and let us assume that each of the K exogenous variables has T_i^m observations, It is useful to transform equation 127 to

$$Y_h = A_h Y_{h-1} + XQ(\theta)B + u^h, \tag{128}$$

where Y_h has size T, A_h is as before, X is a matrix with size $T \times \sum_{i=1}^{K} L_i$, $Q(\theta)$ is a mapping matrix with size $\sum_{i=1}^{K} L_i \times \sum_{i=1}^{K} q_i$, θ is a set of parameters that decides the elements of $Q(\theta)$, q_i will be described later and B is a matrix of unrestricted parameters with size $\sum_{i=1}^{K} q_i \times H$. Finally we have, where we have for illustration purposes assumed that $L_i = L \; \forall \; i$

$$X = \begin{bmatrix} X_{um_1^1}^1 & \cdots & X_{um_K^1}^K & \cdots & X_{um_1^1-L}^1 & \cdots & X_{um_K^1-L}^K \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{um_1^T}^1 & \cdots & X_{um_K^T}^K & \cdots & X_{um_1^T-L}^1 & \cdots & X_{um_K^T-L}^K \end{bmatrix} \tag{129}$$

where um_i^t is the last high frequency observation in the low frequency period t. E.g. if the dependent variable is observed at yearly frequency, and the first observation is 2000, and the first exogenous variable is observed at monthly frequency, then um_1^1 will represent the date 2000M12.

5.1.1 Unbalanced option

If the exogenous regressors have leading observations NB toolbox has an options to allow you to explore this ragged edge. Let say one of the monthly exogenous variables have 3 more months of observations then the yearly dependent variable, and the last observation of the dependent variable is 2000, we get that um_1^T represents 2001M3 instead of 2000M12.

5.1.2 Unrestricted

$Q(\theta)$ is the identity matrix, while B is estimated with OLS. In this case $q_i = L_i$. Be careful not to estimate too many parameters, as this will lead to overfitting!

5.1.3 Beta lag

Here we again assume $L_i = L \; \forall \; i$ for a ease of the disposition. For the Beta lag model we must have that $q_i = q = 1 \; \forall \; i$. Let G be a row vector with L elements. Let $G_{1,j}$ be the jth element of this vector, then it is given by

$$G_{1,j} = \frac{f(\frac{k}{k^{max}}, \theta)}{\sum_{i=1}^{k^{max}} f(\frac{i}{k^{max}}, \theta)} \tag{130}$$

where $f(z, \theta) = (1-z)^{\theta-1}/\beta(1, \theta)$.

$$Q(\theta) = G \otimes I \tag{131}$$

As $Q(\theta)$ depend non-linearly on the single parameter θ we must estimate this model with either NLS or profiling. In NB toolbox this model is estimated by profiling θ over the grid 1 to 40, using 80 grid points. The value in this grid that minimizes the sum of squared residuals from the estimation of equation 128 by OLS given the value of θ is the value selected. And the OLS estimates of B at this value is the estimated values of B.

5.1.4 Almon lag

Here we again assume $L_i = L \; \forall \; i$ and $q_i = q = 3 \; \forall \; i$ for a ease of the disposition. For the Almon lag model $q \leq 3$.

First we define a base grid

$$g_{base} = \{1, 2, \ldots, L-1, L\} \tag{132}$$

Then the jth row (G_j) of the matrix G is given by

$$G_j = g_{base}^j \tag{133}$$

$$Q(\theta) = G \otimes I \tag{134}$$

B is then estimated by OLS given this value of $Q(\theta)$.

5.1.5 Legendre lag

Here we again assume $L_i = L \; \forall \; i$ and $q_i = q = 3 \; \forall \; i$ for a ease of the disposition. For the Legendre lag model $q \leq 3$.

First we define a base grid

$$g_{base} = \{-1, -1 + \frac{2}{L-1}, \ldots, 1 - \frac{2}{L-1}, 1\} \tag{135}$$

$$G = \begin{bmatrix} 1 \\ g_{base} \\ \frac{3}{2}g_{base}^2 - \frac{1}{2} \end{bmatrix} \tag{136}$$

$$Q(\theta) = G \otimes I \tag{137}$$

B is then estimated by OLS given this value of $Q(\theta)$.

6 ARIMA

A general ARIMA(p,0,q) can be written as

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + u_t + b_1 u_{t-1} + b_2 u_{t-2} + \ldots + b_q u_{t-q}. \tag{138}$$

To estimate such a model we need to be sure that the series y is stationary. Say if the series y is integrated at level 1, I(1), then by taking the difference the series y is I(0). I.e. it is stationary. To check that a series is I(1) you can use a unit root test. See section 1.2. If the test cannot reject that the series is I(1) you must difference the data before estimating equation 138. To estimate an ARIMA(p,0,q) is in general not so easy. There are two estimation methods supported; Maximum likelihood and Hannan-Rissan. These two estimators will be discussed in sections 6.1 and 6.2 respectively.

6.1 Maximum likelihood

First we write the state-space representation of the ARIMA(p,0,q) model, see section 2 for a general description of the state-space representation.

State transition matrix

$$A = \begin{bmatrix} \rho_1 & 1 & \cdots & 0 \\ \rho_2 & 0 & \ddots & 0 \\ \vdots & \vdots & \cdots & 1 \\ \rho_k & 0 & \cdots & 0 \end{bmatrix}. \tag{139}$$

$k = max(p, g)$. If $q > p$ the $q - p$ last elements of the first column is zero.

Input Matrix:

$$C = \begin{bmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_q \end{bmatrix}. \tag{140}$$

State equation residual assumption:

$$Q = Eu_t' u_t = \sigma, \tag{141}$$

where σ has size 1 x 1, and is the residual variance.

Observation matrix:

$$H = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}. \tag{142}$$

The constant in the state equation is set to zero. The constant in the observation equation is zero if the constant option is not used, otherwise it is estimated. There is no measurement error, so $R = 0$.

The Kalman filter is then used to evaluate the likelihood, and maximized over a grid or by other methods. Initial values of the parameter vector will come from the Hannan-Rissan algorithm.[17] You can read about this estimation approach in general terms in section 2.3.

6.2 Hannan-Rissan

Lets say that you want to estimate an ARIMA(p,0,q) model. The Hannan-Rissan method uses a 3 step method:

1. Estimate an ARIMA(pq,0,0) model with OLS.
2. Lag the residual from the regression in step 1 to form the q residual terms.
3. Estimate the ARIMA(p,0,q) model by OLS, where the residual terms formed in step 2 is used to estimate the MA coefficients.

This approach is fast, but often unprecise. Another problem is that the residual of the regression in step 3 is not equal to the residual used in step 2. This will most likely make inference using the standard test statistics invalid.

6.3 Seasonal ARMA terms

Box and Jenkins (1976) recommended the use of seasonal autoregressive (SAR) and seasonal moving average (SMA) terms for monthly and quarterly data that contains systematic seasonal movements.

A second order ARMA process without seasonality is given by

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_1 u_{t-1} + b_2 u_{t-2} + u_t, \tag{143}$$

which can easily be represented using lag operator L as

$$(1 - a_1 L - a_2 L^2) Y_t = (1 + b_1 L + b_2 L^2) u_t. \tag{144}$$

You can add SAR(sp) and SMA(sq) terms to you expression to take account of seasonality. With quarterly data, you might want to add SAR(4) or a SMA(4) to the expression. Now the estimated error structure for the ARMA representation would be

$$(1 - a_1 L - a_2 L^2)(1 - \vartheta L^4) y_t = (1 + b_1 L + b_2 L^2)(1 + \omega L^4) u_t. \tag{145}$$

This expression is equivalent to

$$\begin{aligned} y_t =& a_1 y_{t-1} + a_2 y_{t-2} + \vartheta y_{t-4} - \vartheta a_1 y_{t-5} - \vartheta a_2 y_{t-6} + u_t \\ &+ b_1 u_{t-1} + b_2 u_{t-2} + \omega u_{t-4} + \omega b_1 u_{t-5} + \omega b_2 u_{t-6}, \end{aligned} \tag{146}$$

where the parameters ϑ and ω are associated with the seasonal part of the processes. More generally we can write a seasonal ARMA(p,q) as

[17] If this method returns non-stable estimates, the initial guess will be an ARIMA(p,0,q) with its first AR coefficient set to 0.8, and the rest of the coefficients set to 0.

$$(1 - a_1L - a_2L^2 - ... - a_pL^p)(1 - \vartheta L^{sp})y_t = (1 + b_1L + b_2L^2 + ... + b_qL^q)(1 + \omega L^{sq})u_t. \tag{147}$$

Note that in the example above we used sp and sq equal to 4 since we assumed quarterly data. If you on the other hand are using monthly data, please set ps and sq to 12.

6.4 ARIMAX

We can also append exogenous variables to the ARIMA model. When doing this we add the exogenous variables to the measurement equation

$$y_t = c + \beta Z_t + e_t, \tag{148}$$

where e_t can be expressed as

$$e_t = a_1e_{t-1} + ... + a_pe_{t-p} + \theta_1u_{t-1} + ... + \theta_qu_{t-q}. \tag{149}$$

β is a 1xW coefficient vector, and Z_t is a Wx1 vector with the exogenous variables. It can be represented as

$$\begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_W \end{bmatrix} \begin{bmatrix} z_{t,1} \\ z_{t,2} \\ \vdots \\ z_{t,W} \end{bmatrix}, \tag{150}$$

where W is the amount of exogenous variables.

6.4.1 Maximum Likelihood

Similar to 6.1, the only extension is that we now include the exogenous variables and a constant in the estimation. This means that $G = \beta$ and $d = c$ in equation 53 in section 2.

6.4.2 Hannan-Rissan

Two step procedure:

1. Estimate
$$y_t = c + \beta Z_t + e_t, \tag{151}$$
obtain $\hat{c}$ and $\hat{\beta}$. From these estimates you can get
$$\hat{e}_t = y_t - \hat{c} - \hat{\beta} Z_t. \tag{152}$$

2. Estimate
$$\hat{e}_t = a_1\hat{e}_{t-1} + ... + a_p\hat{e}_{t-p} + \theta_1u_{t-1} + ... + \theta_qu_{t-q}, \tag{153}$$
like in section 6.2.

7 Error-correction model (ECM)

A general ECM can be written as, where Δ is the difference operator

$$\begin{aligned}\Delta Y_t =& B_0 X_t + B_1(Y_{t-1} - \rho Z_{t-1}) + D_1 \Delta Y_{t-1} + ... + D_i \Delta Y_{t-i} \\ &+ G_1 \Delta Z_{t-1} + ... + G_j \Delta Z_{t-j} + u_t,\end{aligned} \tag{154}$$

where Y_t is the dependent variable of the model with size 1 x 1, X_t is the exogenous variables of the model with size N x 1 (including the constant), Z_t is the endogenous variables of the model with size Q x 1, u_t is the residuals of the regression with size 1 x 1, B_0 has size 1 x N, B_1 has size 1 x 1, ρ has size 1 x Q, D_i has size 1 x 1 and G_i has size 1 x Q.

7.1 Estimation

The model is estimated with OLS on the following specification of equation 155

$$\begin{aligned}\Delta Y_t =& B_0 X_t + B_Y Y_{t-1} + B_Z Z_{t-1} + D_1 \Delta Y_{t-1} + ... + D_i \Delta Y_{t-i} \\ &+ G_1 \Delta Z_{t-1} + ... + G_j \Delta Z_{t-j} + u_t.\end{aligned} \tag{155}$$

8 VAR

Let start with some basic VAR theory. Any VAR can be written on the form (reduced form)

$$Y_t = AY_{t-1} + BX_t + u_t. \tag{156}$$

$$u_t \sim \Phi(0, \Sigma). \tag{157}$$

Y_t has size M x 1, X_t has size N x 1, u_t has size M x 1, A has size M x M, B has size M x N, Σ is the covariance matrix of the residuals and has size M x M and Φ is some multivariate distribution, often assumed be multivariate normal.

For now we only study the case that the number of lags of the VAR is 1, for an easy discussion.[18]. The dependent variables of the VAR is the variables included in Y_t, i.e. the variables we want to explain by its own lags and some possibly exogenous variables. The exogenous variables of the VAR is the variables included in X_t. Here will also variables like the constant and time trend be included.

Block exogenous variables can also be declared. These are variables that can be included in $\tilde{Y}_t$ when the VAR is written in the following form:

$$\begin{bmatrix} Y_t \\ \tilde{Y}_t \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ \tilde{Y}_{t-1} \end{bmatrix} + BX_t + u_t \tag{158}$$

I.e. the dependent variables in Y_t can be affected by the block exogenous variables in $\tilde{Y}_t$, but not the other way around. For example can foreign variables be seen as block exogenous to domestic variables by the small open economy hypotheses.

Estimation methods will be discussed in the next two section.

8.1 OLS

Estimation of equation 156 is straight forward with OLS. It will be an efficient and consistent estimator under the standard assumption on the residuals. Estimation of equation 158 with OLS is on the other hand

[18] Any VAR of order P can anyway be written in the way $Y_t = AY_{t-1} + BX_t + u_t$ by redefining Y_t.

not efficient, but it is still a consistent estimator.

8.2 Ridge

Estimation of equation 156 is done equation by equation using the estimator described in section 3.3. Estimation of equation 158 is not supported.

8.3 LASSO

Estimation of equation 156 is done equation by equation using the estimator described in section 3.4. Estimation of equation 158 is not supported.

8.4 B-VAR

When it comes to bayesian estimation you need to select a prior. Each of the supported priors will now be described in the next sections.

8.4.1 Standard Minnesota prior

Lets first rewrite the equation 156 by defining

$$\tilde{X}_t = \begin{bmatrix} Y_{t-1} \\ X_t \end{bmatrix} \tag{159}$$

$$\tilde{A} = \begin{bmatrix} A & B \end{bmatrix}. \tag{160}$$

And then transpose all elements to get

$$Y_t' = \tilde{X}_t' \tilde{A}' + u_t' \tag{161}$$

$$Y_t' = (I \otimes \tilde{X}_t')a + u_t', \tag{162}$$

where $a = vec(A')$. A Minnesota prior then assumes that a is distributed following equation 163 and that Σ can be approximated by an estimate. It is also assumed that Σ is diagonal, and that the diagonal elements can be estimated by the standard deviation of OLS residual of each individual equation of the VAR estimated separately. We abbreviate the diagonal elements of Σ by σ_{ii}.

$$a \sim N(\underline{a}^m, \underline{V}) \tag{163}$$

The Minnesota prior involves setting most of the elements of the prior mean to zero, thus ensuring shrinkage of the VAR coefficients towards zero to prevent over-fitting. To be precise the Minnesota prior used here will set the prior mean of all elements of a to zero except the first lag of the dependent variable of each equation in the VAR, which can be set by a hyperparameter.

The Minnesota prior covariance matrix $\underline{V}^m$ is also assumed to be diagonal. If we let $\underline{V}_i^m$ be the block of $\underline{V}^m$ associated with the coefficients of equation i and let $\underline{V}_{i,jj}^m$ be its diagonal elements. Then we can express the priors on the diagonal elements used as

$$\underline{V}_{i,jj}^m = \begin{cases} \frac{\alpha_1}{p^2} & \text{for own lags} \\ \frac{\alpha_2 \sigma_i}{p^2 \sigma_{jj}} & \text{for lags of variable } j \neq i \\ \alpha_3 \sigma_{ii} & \text{for exogenous variables.} \end{cases} \tag{164}$$

p is the lag index, and $p \in [1, P]$, where P is the number of lags of the VAR. The default tightness priors in NB toolbox (Paulsen, 2021) are $\alpha_1 = 0.5$, $\alpha_2 = 0.5$ and $\alpha_3 = 100$.

Under these priors the posterior distribution of a is given by

$$a|y, \Sigma \sim N(\overline{a}, \overline{V}^m) \tag{165}$$

$$\overline{V} = \left[[\underline{V}^m]^{-1} + \left(\widehat{\Sigma}_{-1} \otimes (\tilde{X}'\tilde{X}) \right) \right]^{-1} \tag{166}$$

$$\overline{a} = \overline{V} \left[[\underline{V}^m]^{-1} \underline{a}^m + (\widehat{\Sigma}_{-1} \otimes \tilde{X})' vec(Y') \right], \tag{167}$$

where $\tilde{X}$ is constructed by stacking the $\tilde{X}_t'$ over all t. For more on the Minnesota prior used see Koop and Korobilis (2009).

8.4.2 Jeffrey prior

The VAR can be written as (See section 8.4.1)

$$Y_t' = (I \otimes \tilde{X}_t')a + u_t', \tag{168}$$

where $a = vec(A)$. The likelihood function can then be derived and shown to be of the form

$$a|\Sigma, Y \sim N\left(\widehat{a}, \Sigma \otimes (\tilde{X}'\tilde{X})^{-1} \right) \tag{169}$$

$$\Sigma^{-1}|Y \sim W\left(S^{-1}, T - K - M - 1 \right), \tag{170}$$

where $\widehat{a}$ is the OLS estimate of a, $K = N + M$, W is the Wishart distribution and S can be found by

$$S = (Y - \tilde{X}\widehat{A}')'(Y - \tilde{X}\widehat{A}'), \tag{171}$$

where $\widehat{A}$ is the OLS estimate of the matrix A. $\tilde{X}$ and Y is constructed by stacking the $\tilde{X}_t'$ and Y_t' respectively over all t.

Equations 169 and 170 is then the posterior distribution given the diffuse Jeffrey prior. To draw from this distribution the OLS estimate of Σ is used as the initial value.

8.4.3 Normal-Wishart

The Normal-Wishart prior uses the natural conjugate prior, and has the form

$$a|\Sigma \sim N\left(\underline{a}, \Sigma \otimes \underline{V} \right) \tag{172}$$

$$\Sigma^{-1} \sim W\left(\underline{S}^{-1}, \underline{v} \right), \tag{173}$$

where $\underline{a}$, $\underline{V}$, $\underline{S}$ and $\underline{v}$ are priors chosen as follows

$$\underline{a} = \mathbf{0} \tag{174}$$

$$\underline{V} = V_{scale} I \tag{175}$$

$$\underline{S} = S_{scale} I \tag{176}$$

$$\underline{v} = M + 1. \tag{177}$$

With this prior the posterior becomes

$$a|\Sigma, Y \sim N\left(\overline{a}, \Sigma \otimes \overline{V}\right) \tag{178}$$

$$\Sigma^{-1}|Y \sim W\left(\overline{S}^{-1}, \overline{v}\right), \tag{179}$$

where

$$\overline{V} = \left[\underline{V}^{-1} + \tilde{X}'\tilde{X}\right]^{-1} \tag{180}$$

$$\overline{A} = \overline{V}\left[\underline{V}^{-1}\underline{A} + \tilde{X}'\tilde{X}\widehat{A}\right] \tag{181}$$

$$\overline{a} = vec(\overline{A}) \tag{182}$$

$$\overline{S} = \widehat{S} + \underline{S} + \widehat{A}'\tilde{X}'\tilde{X}\widehat{A} + \underline{A}'\underline{V}^{-1}\underline{A} - \overline{A}'(\underline{V}^{-1} + \tilde{X}'\tilde{X})\overline{A}. \tag{183}$$

$\widehat{S}$ is the standard error and $\widehat{A}$ is the parameter estimates from the OLS estimator. And

$$\overline{v} = \underline{v} + T, \tag{184}$$

where T is the number of observations corrected for the number of lags of the VAR. See Koop and Korobilis (2009) for more thorough discussion of this prior. This prior does not support block exogenous variables.

8.4.4 Independent Normal-Wishart

The independent Normal-Whishart prior takes the form

$$p(\alpha, \Sigma^{-1}) = p(\alpha)p(\Sigma^{-1}), \tag{185}$$

where

$$\alpha \sim N(\underline{a}, \underline{V}) \tag{186}$$

$$\Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{v}). \tag{187}$$

Using this prior we don't get a closed form solution for the joint posterior of α and Σ^{-1}, but the conditional posterior distributions has closed form solutions

$$\alpha|Y, \Sigma^{-1} \sim N(\overline{a}, \overline{V}) \tag{188}$$

$$\Sigma^{-1}|Y, \alpha \sim W(\overline{S}^{-1}, \overline{v}). \tag{189}$$

Let I_T be the identity matrix with size T, then

$$\overline{V} = \left(\underline{V}^{-1} + X'(\Sigma^{-1} \otimes I_T)X\right)^{-1} \tag{190}$$

$$\overline{a} = \overline{V}\left(\underline{V}^{-1}\underline{\alpha} + X'(\Sigma^{-1} \otimes I_T)vec(Y)\right) \tag{191}$$

$$\overline{v} = T + \underline{v} \tag{192}$$

$$\overline{S} = \underline{S} + (Y - XA)'(Y - XA), \tag{193}$$

where A, Y and X are defined in the same way as section 8.4.2. By knowing the conditional posterior a Gibbs sampler algorithm can be used to draw sequentially from the conditional distributions, which will then results in draws from the joint posterior. A burn in sample is required for this algorithm to work. Default is 500 draws.

For more on this prior see Koop and Korobilis (2009).

The priors $\underline{a}$, $\underline{V}$, $\underline{S}$ and $\underline{v}$ are chosen as follows

$$\underline{a} = \mathbf{0} \tag{194}$$

$$\underline{V} = V_{scale} I \tag{195}$$

$$\underline{S} = S_{scale} I \tag{196}$$

$$\underline{v} = M + 1. \tag{197}$$

8.4.5 Combined Minnesota and Independent Normal-Wishart prior

This prior is formed by setting $\underline{a} = \underline{a}^m$, $\underline{V} = \underline{V}^m$ and $\underline{S} = S_{scale} S^{OLS}$, i.e. use the standard Minnesota prior for the coefficients and the OLS estimate of the residual covariance matrix as the priors for the Independent Normal-Wishart prior.

8.4.6 Giannone, Lenza and Primiceri

This is the prior specified by Giannone *et al.* (2012). It is a combined Minnesota and Normal-Wishart prior. So the prior is on the form

$$a|\Sigma \sim N\left(\underline{a}, \Sigma \otimes \underline{V}\right) \tag{198}$$

$$\Sigma^{-1} \sim W\left(\underline{S}^{-1}, \underline{v}\right), \tag{199}$$

This prior is formed by setting $\underline{a} = \underline{a}^m$, i.e. same prior on a as was the case for the Minnesota prior. We define as before tha $\underline{a} = vec(\underline{A})$.

The prior covariance matrix $\underline{V}$ is assumed to be diagonal. If we let $\underline{V}_i$ be the block of $\underline{V}$ associated with the coefficients of equation i and let $\underline{V}^m_{i,jj}$ be its diagonal elements. Then we can express the priors on the diagonal elements used as

$$\underline{V}_{i,jj} = \begin{cases} \frac{\lambda^2}{p^2 \sigma_i} & \text{for lags on endogenous , i.e. for } j > N \\ V_c & \text{for exogenous variables, i.e. for } j \leq N. \end{cases} \tag{200}$$

p is the lag index, and $p \in [1, P]$, where P is the number of lags of the VAR. The default tightness priors are $\lambda = 0.2$ and $V_c = 10^7$. In contrast to the standard Minnesota prior, σ_i is estimated based on an AR(1) model for each dependent variable i instead of an AR(p) model.

The prior on the covariance matrix is given by

$$\underline{S} = S_{scale} \Sigma \tag{201}$$

$$\underline{v} = M + 1. \tag{202}$$

With this prior the posterior becomes

$$\Sigma^{-1} | Y \sim W \left(\overline{S}^{-1}, \overline{v} \right), \tag{203}$$

$$a | \Sigma, Y \sim N \left(\overline{a}, \Sigma \otimes \overline{V} \right) \tag{204}$$

where

$$\overline{V} = \left[\underline{V}^{-1} + \tilde{X}' \tilde{X} \right]^{-1} \tag{205}$$

$$\overline{A} = \left(X'X + \underline{V}^{-1} \right)^{-1} \left(X'Y + \underline{V}^{-1} \underline{A} \right) \tag{206}$$

$$\overline{a} = vec(\overline{A}) \tag{207}$$

$$\overline{\epsilon} = y - X \overline{A} \tag{208}$$

$$\overline{S} = \underline{S} + \overline{\epsilon}' \overline{\epsilon} + \left(\overline{A} - \underline{A} \right)' \underline{V}^{-1} \left(\overline{A} - \underline{A} \right). \tag{209}$$

$$\overline{v} = \underline{v} + T. \tag{210}$$

where T is the number of observations corrected for the number of lags of the VAR.

8.4.7 Priors for the long run

The prior for the long run is implemented as in Giannone *et al.* (2016). They append some dummy observations to the end of the sample of the data to apply the prior. First you need to specify the a prior for the long run relation. This is done with a matrix H of size M x M, where M is the number of endogenous variable in the VAR. The i-th row of H, abbreviated by $H_{i.}$, contains the coefficients of one linear combination of Y_t that you think are likely to be mean reverting. If we rewrite 156 as follows

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_P \Delta Y_{t-P} + u_t, \tag{211}$$

where we have dropped the exogenous variables for simplicity and expanded the lag structure. We want to put a prior on Π given H by

$$\Pi_{.i}|H_{.i},\Sigma \sim N(0,\frac{\phi_i^2}{(H_{.i}\overline{y}_0)^2}), \tag{212}$$

where the i-th column of Π is abbreviated by $\Pi_{.i}$, i-th column of H is abbreviated by $H_{i.}$, $\overline{y}_0 = \frac{1}{P}\sum_{t=1}^{P} Y_t$ and ϕ_i is a hyper parameter to be chosen by the user. This prior can, as shown by Giannone et. al (2014), be implemented by adding the dummy observations

$$Y_{t_i^*} = Y_{t_i^*-1} = \cdots = Y_{t_i^*-P} = \frac{H_{i.}\overline{y}_0}{\phi_i}\left[H^{-}1\right]_{.i}, i \in [1,n]. \tag{213}$$

This prior is a conjugate prior, and can therefore be combined with the other priors by adding these artificial observations to the end of the original sample, and then apply the standard steps of the prior of your choice. For more on this prior see Giannone *et al.* (2016).

8.4.8 Sum-of-coefficient prior

The sum-of-coefficients prior of Litterman *et al.* (1986) and Sims and Zha (1996) corresponds to a special case of the Priors for the long, with a mechanical choice of $H = I$, and hyperparameters $\phi_i = \mu \; \forall \; i \in [1,n]$.

8.4.9 Dummy-initial-observation prior

We implements the Dummy-initial-observation prior by **?** by appending the following dummy observation to the sample

$$Y^{++} = \frac{\overline{y}_0}{\delta}, \tag{214}$$

$$X^{++} = \begin{bmatrix}\frac{1}{\delta} & Y^{++} & \cdots & Y^{++}\end{bmatrix}, \tag{215}$$

where we have defined $\overline{y}_0 = \frac{1}{P}\sum_{t=1}^{P} Y_t$. The first element of X^{++} is due to the constant term. If more exogenous variables than the constant term are added to the VAR, they will be added the dummy observations $X_E^{++} = \frac{1}{P}\sum_{t=1}^{P} X_t$.

8.4.10 Stochastic-volatility-dummy prior

We implements the Stochastic-volatility-dummy prior by Lenza and Primiceri (2020) by adjusting the sample data. The correction to the model (equation 156) is as follows

$$Y_t = AY_{t-1} + BX_t + s_t u_t. \tag{216}$$

where s_t is given by

$$s(t) = \begin{cases} 1 & \text{for } t < T^* \\ eta_{t-T^*-1} & \text{for } t \in \{T^*, T^{**}\} \\ 1 + (eta_{T^{**}-T^*-1} - 1)\rho^{t-T^{**}} & \text{for } t > T^{**} \end{cases}$$

where eta_{t-T^*-1} and ρ are hyper-priors. Default is that $\rho = 0.5$ and

$$eta_{t-T^*-1} = \frac{\frac{1}{M}\sum_{i=1}^{M}\left(Y_t^i - Y_{t-1}^i\right)}{\frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{T^*-1}\sum_{s=1}^{T^*-1}\left(Y_s^i - Y_{s-1}^i\right)\right)} \tag{217}$$

T^* must be set by the user, and the same applies for $N^{MAX} = T^{**} - T^* + 1$. You can do empirical Bayes to estimate eta_{t-T^*-1} and ρ, as is described in section 8.5.

8.4.11 Laplace prior

In Bayesian statistics, the Laplace prior is a continuous probability distribution that can be used as prior belief about the unknown parameters of a model. It is a type of double exponential distribution, also known as the double-sided exponential distribution with probability density function given by:

$$p(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \tag{218}$$

where x is the value of the unknown parameter, μ is the location parameter, and b is the scale parameter.

The Laplace prior is a symmetric distribution, meaning that it is symmetrical about the location parameter μ and is often used in situations where there is a belief that the unknown parameter is likely to be close to a particular value. For example, if the location parameter μ is set to 0, the Laplace prior is often used to model a belief that the unknown parameter is likely to be close to 0. In this case, the posterior mode is the solution of the LASSO method introduced in section 3.4. Note that, in general the posterior mean is not the LASSO solution, and does in fact not yield a sparse solution for the parameters of the problem.

From here on out we follow the setup in Park and Casella (2008) and assume that Y_t only consist of one variable, i.e. $M = 1$. Extension to the full VAR case 156 can then be done equation-by-equation. Consider a model of the form

$$\begin{aligned} Y_t \mid B_1, \lambda, \sigma &\sim \mathcal{N}\left(B_1 X_t, \sigma^2 \mathbf{I}\right) && (219) \\ \beta \mid \lambda, \sigma &\sim \prod_{j=1}^{N} \frac{\lambda}{2\sigma} e^{-\frac{\lambda}{\sigma}\left|B_1^j\right|}, && (220) \end{aligned}$$

where the last equation is the i.i.d. Laplace prior for each of the elements of B_1. Using this, one can easily show that the posterior density for B_1 is

$$p(B_1 \mid Y_t, X_t, \sigma^2) \propto e^{-\frac{1}{2\sigma}\left[(Y_t - B_1 X_t)^2 + \frac{1}{\lambda^2}\sum_{j=1}^{N}\left|B_1^j\right|\right]}, \tag{221}$$

where the expresion inside the brackets is the same objective function as of the Lasso estimator in section 3.4 and, as mentioned above, the mode of this posterior equals the LASSO estimate.

The Laplace prior is closely related to the Gaussian prior, which is the Bayesian version of Ridge regression, but unfortunately in the former case no closed form solution is available for the posterior distribution and MCMC methods are therefore needed in order to sample from this distribution. One popular choice for carrying out posterior inference is Gibbs sampling. This is done by first rewriting the Laplace prior as a mixture of Gaussians and then recursively sample from the conditional distributions. I.e., for each of the B_1^j's rewrite

$$B_1^j \mid \sigma^2 \sim \mathcal{L}(0, 2\sigma^2\lambda^2) \tag{222}$$

as

$$\begin{aligned} B_1^j \mid \sigma^2, w_i &\sim \mathcal{N}(0, \sigma^2 w_i) && (223) \\ w_i \mid \sigma^2 &\sim \mathrm{e}^{\frac{1}{8\sigma^2\lambda^4}}. && (224) \end{aligned}$$

We are then able sample from the posterior by sequentially drawing from

- $p\left(B_1 \mid Y_t, X_t, w, \sigma^2\right)$, which is Gaussian.

- $p\left(w \mid Y_t, X_t, B_1, \sigma^2\right)$, which is inverse of an Inverse-Gaussian.
- $p\left(\sigma^2 \mid Y_t, X_t, B_1, w\right)$, which is Inverse-Gamma.

In order to reduce the dependency on the initial sample and to prevent autocorrelation between the samples, we use a default burn-in sample of 500 and thinning of every second draw, i.e., the first 500 samples are discarded and only every second draw is kept.

8.4.12 DSGE-VAR prior

We follow the setup in Negro and Schorfheide (2004) and assume a Normal - Inverse wishart prior on the form

$$A|\Sigma \sim N\left(\underline{A}, \Sigma \otimes \underline{V}\right) \tag{225}$$

$$\Sigma^{-1} \sim W\left(\underline{S}^{-1}, \underline{v}\right), \tag{226}$$

where $\underline{A}$, $\underline{V}$, $\underline{S}$ and $\underline{v}$ are priors chosen as follows

$$\underline{A} = \Gamma_{xx}^{-1}(\theta)\Gamma_{xy}(\theta), \tag{227}$$

$$\underline{V} = \left(\lambda T \Gamma_{xx}(\theta)\right)^{-1}, \tag{228}$$

$$\underline{S} = \lambda T\left(\Gamma_{yy}(\theta) - \Gamma_{yx}(\theta)\Gamma_{xx}^{-1}(\theta)\Gamma_{xy}(\theta)\right), \tag{229}$$

$$\underline{v} = \lambda T - M - N, \tag{230}$$

where $\Gamma_{zw}(\theta) = E\left[Z'W\right]$ is the theoretical covariance between the variables Z and W from the DSGE model (or any other model for that mather), and λ is the ratio of dummy over actual observations, i.e. measures the weight of the prior relative to the sample. We must have that $\lambda T \geq M+N$. With this prior the posterior becomes

$$A|\Sigma, Y \sim N\left(\overline{A}, \Sigma \otimes \overline{V}\right) \tag{231}$$

$$\Sigma^{-1}|Y \sim W\left(\overline{S}^{-1}, \overline{v}\right), \tag{232}$$

where

$$\overline{V} = \left(\underline{V}^{-1} + X'X\right)^{-1}, \tag{233}$$

$$\overline{A} = \overline{V}\left(\lambda T \Gamma_{xy}(\theta) + X'Y\right), \tag{234}$$

$$\overline{S} = \frac{1}{(\lambda+1)}\left[\left(\lambda T \Gamma_{yy}(\theta) + Y'Y\right) - \left(\lambda T \Gamma_{yx}(\theta) + Y'X\right)\overline{A}\right], \tag{235}$$

and

$$\overline{v} = \underline{v} + T, \tag{236}$$

where T is the number of observations corrected for the number of lags of the VAR.

8.5 Empirical Bayesian

All prior given above uses a set of hyperparameters (γ), i.e. the prior is specified as $\pi(\theta|\gamma)$. So when doing Bayesian estimation we maximize the posterior

$$p(\theta|y,\gamma) = \frac{f(y|\theta,\gamma)\pi(\theta|\gamma)}{m(y|\gamma)}, \tag{237}$$

where $f(y|\theta,\gamma)$ is the likelihood and $m(y|\gamma)$ is the marginal density of the data (MDD), and is given by

$$m(y|\gamma) = \int f(y|\theta|\gamma)\pi(\theta|\gamma)d\theta. \tag{238}$$

In the empirical Bayesian we maximize $m(y|\gamma)$ w.r.t. γ.

8.5.1 Normal-Whishart and GLP

For the Normal-Whishart priors, see sections 8.4.3 and 8.4.6, Giannone *et al.* (2012) shows that the MDD has a closed form solution

$$\begin{aligned} m(y|\gamma) = &\left(\frac{1}{\pi}\right)^{\frac{M\cdot T}{2}} \frac{\Gamma_M\left(\frac{T+\underline{v}}{2}\right)}{\Gamma_M\left(\frac{\underline{v}}{2}\right)} \cdot \\ &\left|\underline{V}\right|^{-\frac{M}{2}} \cdot \left|\underline{S}\right|^{\frac{\underline{v}}{2}} \cdot \left|X'X + \underline{V}^{-1}\right|^{-\frac{M}{2}} \cdot \\ &\left|\underline{S} + \overline{\epsilon}'\overline{\epsilon} + \left(\overline{A}-\underline{A}\right)'\underline{V}^{-1}\left(\overline{A}-\underline{A}\right)\right|^{-\frac{T+\underline{v}}{2}}, \end{aligned} \tag{239}$$

where T is the number of observations corrected for the number of lags of the VAR, Γ_M is the M-variate Gamma function, while the rest of the terms are defined in section 8.4.6. Remember here $\underline{v}$, $\underline{A}$, $\underline{V}$, $\underline{S}$, $\overline{A}$ and $\overline{\epsilon}$ all will be functions of the hyperparameters γ.

8.5.2 DSGE-VAR

For the DSGE-VAR prior it is most useful to write the MDD in the following closed form solution

$$m(y|\gamma) = \left(\frac{1}{\pi}\right)^{\frac{M\cdot T}{2}} \frac{\Gamma_M\left(\frac{T+\underline{v}}{2}\right)}{\Gamma_M\left(\frac{\underline{v}}{2}\right)} \left|\underline{V}\right|^{-\frac{M}{2}} \cdot \left|\underline{S}\right|^{\frac{\underline{v}}{2}} \cdot \left|\overline{V}\right|^{-\frac{M}{2}} \left|\underline{S}\right|^{-\frac{\overline{v}}{2}}. \tag{240}$$

8.6 Time varying parameter bayesian estimation

Here we use the algorithm described in section 11.2. In this case $O_t = \hat{Y}_t$, where $\hat{Y}_t$ is Y_t, but where the lags are removed. $\lambda_t = H$ from section 9.1, which implies that $F_t = Y_t$.

Caution: This estimator does not include the exogenous variables in the state equation. The exogenous variables are instead used to pre-filter the data, as described in algorithm of section 11.2 point 2. In this case the variables are not standardized!

8.7 Adding measurement restrictions

See section 9.7.2.

8.8 Identification

Estimation of VAR models are normally done on the reduced form

$$Y_t = AY_{t-1} + BX_t + u_t, \tag{241}$$

where

$$Eu_t' u_t = \Omega. \tag{242}$$

The residuals u_t in 157 have no structural meaning, as the different residuals in u_t may be correlated, i.e. Ω is not assumed to be diagonal. This is what identification will help us with. There are different approaches for how to identify the underlying orthogonal shocks. But let us start with a general set up. We are looking for a matrix C and orthogonal shocks such that

$$u_t = C\varepsilon_t,\ E\varepsilon_t'\varepsilon_t = I,\ C'C = \Omega. \tag{243}$$

But typically there are many matrices, C, that satisfies 243, so additional information must be used to pin down C, or Cs in the case of under-identification (e.g. sign restrictions).

One approach is to use Cholesky decomposition of Ω. This will produce a C that is lower triangular. So if for example we have a VAR with 3 dependent variables, and we use a Cholesky decomposition it will result in the equation (for now assume only one lag and no exogenous variables)

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \end{bmatrix} = A \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \end{bmatrix} + \begin{bmatrix} C_{1,1} & 0 & 0 \\ C_{2,1} & C_{2,2} & 0 \\ C_{3,1} & C_{3,2} & C_{3,3} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}. \tag{244}$$

We see that the shock $\varepsilon_{1,t}$ will only affect the variable $Y_{1,t}$ contemporaneously, and as we know it is not correlated with the other shocks we can interpret it as a shock that only affects variable $Y_{1,t}$ contemporaneously. So what about the second equation? We see that $Y_{2,t}$ will depend on both $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$, but as we already have identified $\varepsilon_{1,t}$ from the first equation, we can back out $\varepsilon_{2,t}$. Along this way we can identify the third as well. So what are the economical assumption we are making here? Say that the we have a VAR with GDP, inflation and the key policy rate, in that order. Then the shock $\varepsilon_{1,t}$, call it a supply shock, may affect inflation and the key policy rate contemporaneously, while the shock $\varepsilon_{3,t}$, call it a monetary policy shock, will not affect GDP and inflation contemporaneously. The restriction used here is the so called zero short term restriction.

It is also possible to combine zero and sign restrictions on all periods, as well as long term restrictions. Long term restriction will mean restriction on the cumulative sum over the infinite horizon. A sign restriction is a restriction on the sign of the response of one shock on one variable at a given period. This documentation will not go into a further technical description of this algorithm and the economical reasoning behind sign restriction, but instead refer to Binning (2013). One thing to note, is that under sign restriction you are not finding a unique C matrix, i.e. the model is under-identified, and many C matrices may be found to satisfy 243 and the added restrictions. Therefore it is important when doing for example IRFs at a later stage to include this uncertainty. Therefore it makes no sense to look at a point IRF for one identified C matrix. You must instead identify many Cs and produce density IRFs based on those. Either with or without parameter uncertainty.

8.9 Test the VAR

Each of the supported tests for a VAR model will be described in the next sub sections.

8.9.1 Test of stationarity

Stationarity of the VAR is only dependent on the matrix in front of the lagged dependent variables. For the VAR to be stationary the eigenvalues of the matrix A must be inside the unit circle. I.e. all eigenvalues must have modulus less than one.

8.9.2 LM-test for autocorrelation

Let $\tilde{u}_t$ be the residual from the estimated VAR. Then the LM-test for autocorrelation uses the auxiliary regression

$$\tilde{u}_t = a_1 y_{t-1} + \cdots + a_P y_{t-P} + BX_t + D\tilde{u}_{t-j} + v_t, \tag{245}$$

where y_{t-i} has size K, i.e. the number of dependent variables of the VAR, and P is the number of lags of the VAR. The elements $\tilde{u}_{-j}$, $\cdots$, $\tilde{u}_{-1}$ is set to zero. Let $\tilde{v}_t$ be the OLS residual from this regression, then the test statistic is given by

$$LM_j = -(T - K(p+1) - \frac{1}{2}) ln \left(\frac{|\Omega_j|}{|\Omega|} \right), \tag{246}$$

where[19]

$$\Omega = \frac{\tilde{u}'\tilde{u}}{T} \tag{247}$$

$$\Omega_j = \frac{\tilde{v}'\tilde{v}}{T}. \tag{248}$$

This test statistic is distributed as χ^2 with K degrees of freedom. The null is no autocorrelation of degree j.

8.9.3 Ljung-Box test for autocorrelation

The Ljung-Box test for autocorrelated residuals is constructed as

$$LB = T(T-2) \sum_{h=1}^{\lceil T/4 \rceil} \frac{trace(\Omega_h' \Omega_0^{-1} \Omega_h' \Omega_0^{-1})}{T-h}, \tag{249}$$

where

$$\Omega_h = \frac{\tilde{u_0}' \tilde{u_h}}{T}, \tag{250}$$

and $\tilde{u}_h$ is constructed by concatenating $\tilde{u}_{t-h}$ vertically over time. The elements $\tilde{u}_{-h}$, $\cdots$, $\tilde{u}_{-1}$ are set to zero.

This test statistic is distributed as χ^2 with $K^2(\lceil T/4 \rceil - P + 1) - K^2$ degrees of freedom. K is the number of dependent variables and P is the number of lags of the VAR. The null hypothesis is no autocorrelation in the VAR residuals.

8.9.4 Granger causality

Granger causality or predictive causality can be tested using a VAR. It amounts to test if one variable x (of potentially many) has predicting power when forecasting another variable y. In the VAR this can be tested by a joint F-test that all coefficient on the lags of x in the equation where y is the dependent variable is

[19] $\tilde{u}$ and $\tilde{v}$ is constructed by concatenating $\tilde{u}_t$ and $\tilde{v}_t$ vertically over time.

equal to zero. The null hypothesis is that x does not Granger-cause y. The test statistic is distributed as F(M,T-M-N).

9 VAR with missing observations

First we set up a state space model which handles missing observations

$$J_t Y_t = H_t X_t + v_t, \tag{251}$$

$$v_t \sim N(0, R), \tag{252}$$

$$X_t = AX_{t-1} + BZ_t + Du_t, \tag{253}$$

$$u_t \sim N(0, \Sigma). \tag{254}$$

Y_t is the observed variables of the model and it has size M x 1 (may have missing observation at some point in time), X_t is the unobserved state vector and it has size S x 1, v_t is the vector with measurement errors and it has size M x 1, Z_t is a vector with the observations on the exogenous variables of the model and it has size N x 1 (cannot have missing observations), u_t is a vector with the reduced form residuals of the model and it has size S x 1, A has size S x S, B has size S x N, D has size S x M and Σ has size M x M. The matrices J_t and H_t are time varying due to fact that we are dealing with missing observations. In the case we have no missing observations the matrix J_t will be the identity matrix with size M x M, while the H_t will be a matrix with size M x S. We come back to how we find H_t later on. When we have missing observations in the vector Y_t, we remove the corresponding rows of J_t and H_t, i.e. we only let the non-missing values in Y_t, inform the state vector X_t in period t.

To allow for measurement errors is in general not yet supported, so $R = 0$, except in special cases.

9.1 MO-VAR

In the case we are dealing with a normal VAR, as discussed in section 8, but some observation on the dependent variables of the model are missing. The dependent variables are the variables included in the vector Y_t. Let H be the value of H_t if there are no missing observations, then H will be partitioned matrix on the form

$$H = \begin{bmatrix} I & 0 \end{bmatrix}, \tag{255}$$

where I is an identity matrix with size M x M, while H is filled in with zeros to have the size M x S. In this case $S = M(L-1)$, where L is the number of lags included in the VAR. So H_t will be constructed by removing the rows of H that corresponds to the missing observations.

9.2 MF-VAR

Mixed frequency vector autoregressive (MF-VAR) model is just a special case of a VAR model with missing observations. Let H be the value of H_t if there are no missing observations on any frequency. From now on we will use the example of combining quarterly and monthly data in a VAR, but any combination of yearly, quarterly, monthly and weekly data is supported, and corresponding adjustment of the H matrix is done automatically. In our special case the state equation 253 will be on monthly frequency, while some of the

dependent (observed) variables, the variables included in the vector Y_t, are on both monthly and quarterly frequency. What we do is to expand the vector Y_t with missing data when it is not in a month that ends the quarter. In this way we reduced the problem to a missing observation VAR. Let Q be the number of variables on quarterly frequency and order them first in the Y_t vector, then we can partition H as

$$H = \begin{bmatrix} MQ \\ MM \end{bmatrix}, \tag{256}$$

where MM is a matrix with size (M - Q) x S, MQ is a matrix with size Q x S and I is an identity matrix with size M-Q x M-Q. The MQ matrix will depend on how the monthly observations aggregates to the quarterly observation. The supported aggregations are

Level summed

$Y_t^i = X_t^i + X_{t-1}^i + X_{t-2}^i$

Diff summed

$Y_t^i = X_t^i + 2X_{t-1}^i + 3X_{t-2}^i + 2X_{t-3}^i + X_{t-4}^i$

Level average

$Y_t^i = \frac{1}{3}(X_t^i + X_{t-1}^i + X_{t-2}^i)$

Diff average

$Y_t^i = \frac{1}{3}X_t^i + \frac{2}{3}X_{t-1}^i + X_{t-2}^i + \frac{2}{3}X_{t-3}^i + \frac{1}{3}X_{t-4}^i$

End

$Y_t^i = X_t^i$

To illustrate, we set up a bivariate VAR model with one quarterly observed series and one monthly observed series. The quarterly observed series is in this example the log difference of GDP, and monthly observed series is the log difference of CPI. The aggregation of the quarterly series must then use the Diff average method found in the table above. We include 5 lags in our VAR. This means that the state vector X_t will be a 10 x 1, i.e. $S = 10$. The observation vector Y_t, has size 2 x 1, i.e. $M = 2$ and $Q = 1$. We then get the following

$$MQ = \begin{bmatrix} \frac{1}{3} & 0 & \frac{1}{2} & 0 & 1 & 0 & \frac{1}{2} & 0 & \frac{1}{3} & 0 \end{bmatrix}, \tag{257}$$

$$MM = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{258}$$

If we where to add more lags to our model we can expand the MQ and MM matrix with zeros at the end, and if we where to reduce the number of lags we still need to keep the state vector with the same size, but then we can append zeros to the upper right part of the A, i.e. restrict the parameters on the lags we do not include in the VAR to zero. Again H_t will be constructed by removing the rows of H that corresponds to the missing observation.

9.3 Maximum likelihood

To estimate the model using maximum likelihood the Kalman filter is used. For more on this see section 2.3. The estimation of the monthly observation will be done using the Kalman smoother, for more on this see section 2.2.

9.4 Bayesian

The supported priors for bayesian estimation can be found in section 8.4, but the Jeffrey and Laplace priors are not supported in this case. Neither is the different dummy observations priors. The approach will only be described here in general terms. As we are dealing with missing observations we need to add an additional step to filter out the missing observations of the state variables conditional on the parameters, i.e. $X|a, \Sigma, Y$. Where we have defined $a = vec(A)$ to be the parameters of the state equation, Σ to be the covariance matrix of the residuals of the state equation, X and Y to be the state variables and observed variables stacked over time. A Kalman smoother is used at this step, see section 2.2 for more on the Kalman smoother.

After the missing observation are filtered out we draw from the conditional distribution of the parameters of the model $a, \Sigma|X$. See section 8.4 for more on this step.

The approach used here is the so called Gibbs sampler. The gibbs sampler draws iteratively the parameters and missing observations from the conditional posterior distributions. In this case the steps of the Gibbs sampler are

1. Sample from $X|a, \Sigma, Y$ using the Kalman smoother.
2. Sample from $a, \Sigma|X$, this will depend on the prior used. See corresponding subsection of the prior used in section 8.4.

To start the Gibbs sampler we need some starting values of a and Σ. They are found as follows

Minnesota

The initial value of a is just the prior $\underline{a}$. For the Minnesota prior we need to find an estimate for σ_{ii}, from there the rest of the algorithm presented in section 8.4.1 applies as before. For a normal VAR with missing observations σ_{ii} is estimated by a AR model using OLS on each series, where we fill in the missing observations with a spline method. This is also done for the high frequency variables in the MF-VAR case, but for the lower frequencies we use a MF-AR(1) model to estimate the variance of the residual.[20]

Normal-Whishart

First we use individual MF-AR(1) models for each low frequency series to get the smoothed estimates of the observations of that series on the high frequency.[20] Then these smoothed series plus the high frequency data is used to estimate A and Σ of the state-equation by OLS. These are then used as the initial values, remembering that $a = vec(A)$.

Independent Normal-Whishart

Same as for Normal-Whishart.

As we use a Gibbs sampler, we need to have a burn in sample. Default is 500. This is to remove the dependency on the initial values. To prevent autocorrelated draws from the Gibbs sampler we must also do thinning. Default is to only keep every second draw.

9.5 Time varying parameter bayesian estimation

Here we use the algorithm described in section 11.2. In this case $O_t = \hat{Y}_t$, where $\hat{Y}_t$ is Y_t, but where the lags are removed. $\lambda_t = H$ from section 9.2, which implies that $F_t = Y_t$.

Caution: This estimator does not include the exogenous variables in the state equation. The exogenous variables are instead used to pre-filter the data, as described in algorithm of section 11.2 point 2. In this case the variables are not standardized!

[20] See section 9.8 for more on the MF-AR(1) approach.

9.6 Variables observed at different frequencies

9.6.1 Change in frequency

A variable that now is observed on a high frequency may back in time only be observed at a low frequency. This can be handle by changing H according to the switch in frequency. Otherwise it will be just like another case of a variable with missing observations.

9.6.2 Observed at different frequencies at the same time

This chapter should be taken with care, as we have not yet any evidence that this approach will work in practise in real-time applications. The meaning of working, is that it will improve forecast performance.

Sometimes you have observed data on the same variable at different frequencies, but the highest frequency observations (or both) exhibits a lot of measurement errors. In this case it may be good for forecast performance to introduce measurement errors in the observation equation, i.e. allow $R \neq 0$. This amount to estimate the noise components in the series and the VAR jointly.

9.6.2.1 Maximum likelihood For the maximum likelihood estimator this means that we optimize over an extended set of parameters. We restrict us to the case that R is diagonal, but we allow for zero restrictions on any elements along that diagonal.

9.6.2.2 Bayesian In the case of Bayesian estimation we suggest a dogmatic prior, where the diagonal elements of R is set in the following way for the high frequency variables[21]:

$$\underline{R_{i+Q}} = \sigma_{ii} \cdot R_{i+Q}^{scale}, \tag{259}$$

where $\underline{R_w}$ is the prior on the w^{th} element along the diagonal of R, σ_{ii} is the i^{th} element of covariance matrix of the residuals, i.e. Σ, and R_{i+Q}^{scale} is a hyperparameter set by the researcher.

For the low frequency variables we allow for the prior on the following type:

$$\underline{R_i} = \sigma_i^L \cdot R_i^{scale}, \tag{260}$$

where σ_i^L is the residual variance from an AR(1) model on the low frequency only observations of the low frequency variable number $i \in [0, Q]$, and again R_i^{scale} is a hyperparameter set by the researcher.

As we use a dogmatic prior on R the algorithms presented in section 9.4 will not change in any other way than that the conditional distribution now also are conditioned on the dogmatic prior on R. This is not any stricter than assuming $R = 0$, which is just another dogmatic prior on R.

9.7 Adding measurement restrictions

It is also possible to add additional measurement restrictions to the MO-VAR or MF-VAR models.

9.7.1 Bayesian estimation

All the priors that are supported for MO-VAR and MF-VAR models, except the prior described in section 9.5, can deal with additional measurement restrictions during estimation. In this case we are expanding the equation 251 with M^R additional observations equations. The diagonal elements of R associated with the appended measurement equations are set by the dogmatic prior:

$$\underline{R_i} = \sigma_i^R \cdot R_i^{scale}, \tag{261}$$

[21] Remember that the low frequency variables are ordered first in the measurement equations.

where σ_i^R is the variance of the restricted variable number $i \in [0, M^R]$, and R_i^{scale} is a hyperparameter set by the researcher. The elements of the H_t can be calibrated to any numbers, or be set to observed time-varying parameters. The last can be handy if you want to use time-varying aggregation weights.

9.7.2 Conditional forecast

For all other estimators it is only possible to expand the measurement equation for use during conditional forecast. The equations are set up in the same way as described in section 9.7.1, and the same dogmatic prior for R is used in this case.

9.8 MF-AR(1)

To set up a MF-AR(1) model we use the same state-space representation as 251 to 254. In this case $H = MQ^i$, where MQ^i is the i'th row of MQ from equation **??**. A is a S x S matrix with the AR coefficient located a the upper left corner, while the rest of the matrix is zeros. D is a S x 1 matrix, where the first element is 1 and the rest is zeros. Σ is a 1 x 1 matrix with the variance of the residual of the state equation. To estimate the model we use a maximum likelihood approach, which is further described in section 9.3. To get the smoothed estimates if the high frequency series a Kalman smoother is used with the estimated parameters substituted into the state-space model, see section 2.2 for more on the Kalman smoother.

9.9 Note on forecasting

9.9.1 Maximum likelihood

9.9.1.1 Point forecast The initial value used for forecasting is the maximum likelihood estimates of the smoothed state variables, otherwise forecasting follows the normal steps provided in section 15.

9.9.1.2 Density forecast Not yet implemented.[22]

9.9.2 Bayesian

9.9.2.1 Point forecast The initial value used for forecasting is the posterior mean estimates of the smoothed state variables, while the posterior mean of the parameters is used to solve for the companion form of the VAR model. Given these assumptions the forecasting follows the normal steps provided in section 15.

9.9.2.2 Density forecast A density forecast is made by drawing from the posterior distribution of both the initial observation and the parameters. These where both produced during the estimation stage. This means that the only difference compared to section 15, is that we may not know the initial observations (nowcast) of some variables.

10 Factor models using principal component

The idea behind factor models is to summarize a huge data set into a few factors that can explain most of the variation in the huge data set, and therefore reduce the parameter to identify in the final model. These factors are unobservable, and have to be estimated. Let the data set consist of J number of variables, then as $J \to \infty$ the uncertainty in estimates of the L factors will shrink, as long as $L << J$. We might think of these factors as diffuse concepts such as "economic activity" or a "financial indicator". In this section

[22] Bootstrapping will be too time consuming.

we will use a two step procedure to estimate the different factor models. The first step is to use principal component analysis to estimate the unobservable factors, which is described in section 10.1. The second step is to use the estimated factors in the wanted model. We support three classes of models of this type; single equation factor model, the step ahead factor model and factor augmented VAR model (FA-VAR), see sections 10.2, 10.3 and 10.4 respectively. Another class of factor models is found in section 11.

10.1 Principal component

The problem is to find a map from the factors to the huge data set on the form (observation equation):

$$Z_t = GF_t + B\chi_t + \nu_t \tag{262}$$

Where Z_t are the variables of the huge data set and has size J x 1, F_t are the unobservable factors to estimate, χ_t are other variables of the observation equation and has size N x 1 and ν_t is the residual of size J x 1.[23] G is a matrix of parameters to estimate and has size J x L. Finally B is a matrix of parameters to estimate and has size J x N.

It is from here assumed that χ_t only consists of a constant term or is empty. You have three options to transform Z_t before using the principal component algorithm, i refers to the observed variable i:

None

$B^i = 0$, $\sigma^i = 1$ and $\tilde{Z}_t^i = Z_t^i$.

Demean

$B^i = \frac{1}{T}\sum_{t=1}^{T} Z_t^i$, $\sigma^i = 1$ and $\tilde{Z}_t^i = Z_t^i - B^i$.

Standardise

$B^i = \frac{1}{T}\sum_{t=1}^{T} Z_t^i$, $\sigma^i = \frac{1}{T-1}\sum_{t=1}^{T}(Z_t^i - B^i)^2$ and $\tilde{Z}_t^i = \frac{Z_t^i - B^i}{\sigma^i}$.

Where $G = \lambda * \Sigma$[24] and λ is the principal component coefficients based on the variables in $\tilde{Z}_t$.

Let first stack all the $\tilde{Z}_t'$ over time to get $\tilde{Z}$. Then the first step is to find the principal components by singular value decomposition of $\tilde{Z}$

$$\tilde{Z} = U\Omega V'. \tag{263}$$

Here $\tilde{Z}$ has size T x J, where T is time dimension of the data set. U is a orthogonal matrix with size T x J, Ω is a diagonal matrix with size J x J and with non-negative numbers on the main diagonal and V' is the transpose of a orthogonal matrix V with size J x J. Then the factors are constructed by

$$\tilde{F} = U *' diag(\Omega), \tag{264}$$

where $*'$ is the element-wise multiplication operator, but as $diag(\Omega)$ has size 1 x J it also stacks $diag(\Omega)$ to construct a T x J matrix. $\tilde{F}$ is the estimated factors and has size T x J. When only L factors are wanted they are found by taking the L first columns of $\tilde{F}$, i.e

$$F = \tilde{F}(1:T, 1:L). \tag{265}$$

F is then the final estimate of the L factors. And the principal component coefficients are estimated by

[23] For some factor model χ_t may consist of more variables than a constant term, and a iterative algorithm must be used, see section 10.4.

[24] $*$ is the element-wise multiplication operator and Σ is constructed by stacking σ L times horizontally. I.e. Σ is scaling the factors back to the normal scale of Z_t.

$$G = V'. \tag{266}$$

The explained variance of each factor can then be calculated as

$$explained_i = 100 \frac{\frac{\Omega_{ii}^2}{T-k}}{\sum_{j=1}^{J} \frac{\Omega_{jj}^2}{T-k}} \text{ for } i \in [1, L], \tag{267}$$

where k is 1 if Z_t are demeaned or standardized before the principal components algorithm is used.

10.1.1 Test for the number of factors

10.1.1.1 Scree

10.1.1.2 Explained

10.1.1.3 Bai and Ng To select the number of factors can be difficult. Bai and Ng (2000) suggest a way to test for the number of factors to use. Let us define $\nu_{t,L}$ as the residual in equation 262 when L is the number of factors used, and then ν_L by stacking $\nu'_{t,L}$ vertically over time. The number of factors can be found by minimizing

$$IC_L = log(\frac{\nu'_L \nu_L}{T}) + CT_L \tag{268}$$

over L. Where CT_L is selected from:

1. $CT_L = Llog(\frac{JT}{J+T})\frac{J+T}{JT}$
2. $CT_L = L\frac{log(min(J,T))(J+T)}{JT}$
3. $CT_L = L\frac{log(min(J,T))}{min(J,T)}$
4. $CT_L = \frac{2L}{T}$
5. $CT_L = L\frac{log(T)}{T}$
6. $CT_L = \frac{2L}{J}$
7. $CT_L = L\frac{log(J)}{J}$
8. $CT_L = 2L\frac{J+T}{JT}$
9. $CT_L = L\frac{log(JT)(J+T)}{JT}$

T is the number of observations and J is the number of observed variables.

10.1.1.4 Trapani

10.1.1.5 Horn

10.2 Single equation factor model

A general single equation factor model can be written as

$$Y_t = B_1 X_t + B_2 F_t + u_t, \tag{269}$$

where Y_t is the dependent variables of the model with size M x 1, X_t is the exogenous variables of the model with size N x 1, F_t is the factors of the model with size J x 1, u_t is the residuals of the regression with size M x 1, B_1 has size M x N and B_2 has size M x J. The second equation of the model is the observation equation

$$O_t = D + \Lambda F_t + u_t, \tag{270}$$

where O_t has size G x 1, F_t has size J x 1, D has size G x 1 and is the constant term, Λ has size G x J and is the factor loadings, $u_t \sim N(0, R)$ (IID) and has size G x 1.

10.2.1 Estimation

First the J number of factors are estimated by PCA, i.e. equation 270 is estimated using PCA. See section 10.1 for more on this. Then equation 269 is estimated using OLS with the estimated factors from the first stage substituted in for F_t.

10.2.2 Bootstrapping single equation factor models

As the factor is estimated we have uncertainties in the regressors in the model, so we add additional steps to the bootstrapping algorithm presented in section 1.7.

1. Estimate the observation equation as discussed in the above section.
2. Re-sample the residual with replacement to construct Q simulated series of u, i.e. the residual of the observation equation. Abbreviate each simulated series by $\tilde{u}^i$.
3. For each simulated series of u construct an artificial series of O by $\tilde{O}_t^i = \hat{\Lambda} F_t + \hat{\rho} Y_t + \tilde{u}_t^i$, where i is the simulation index and $\hat{\Lambda}$ and $\hat{\rho}$ are the estimates of Λ and ρ respectively.
4. Re-estimate the observation equation where O_t is substituted with $\tilde{O}_t^i$ for all i. You will then end up with Q different estimates of F[25], Λ and ρ. I.e. $\hat{F}^i$, $\hat{\Lambda}^i$ and $\hat{\rho}^i$ for $i \in [1 : Q]$. This means that you have an estimate of the distribution of the factors F and all the parameters of the vector Λ and ρ.
5. Use the factors found in the 3 first steps to proceed with the steps of the algorithm in section 1.7. Using one draw of the factors to produce one parameter draws for the other equations of the model.

10.3 Step ahead factor model

The only difference with the single equation factor model is that equation 269 now reads

$$Y_{t+h} = B_1 X_t + B_2 F_t + u_{t+h}, \tag{271}$$

where $h > 0$. The reason we call this a step ahead model is that we can produce direct forecast using this model the number of selected periods ahead.

[25] F is the stacked series of F_t over time.

10.3.1 Estimation

The model is estimated in the same way as in section 10.2.1 except that the Y_t variable is leaded the number of wanted periods.

10.3.2 Bootstrapping the step ahead factor model

Uses the same algorithm as single equation factor models.

10.4 FA-VAR

FA-VAR stands for factor augmented vector auto-regression model. The model consists of two blocks, the observation equation

$$O_t = D + \Lambda F_t + \rho Y_t + u_t, \tag{272}$$

where O_t has size G x 1, F_t has size J x 1, D has size G x 1 and is the constant term, Λ has size G x J and is the factor loadings, Y_t has size M x 1, ρ has size G x M, $u_t \sim N(0, R)$ (IID) and has size G x 1. The VAR equation is given by

$$\begin{bmatrix} Y_t \\ F_t \end{bmatrix} = A \begin{bmatrix} Y_{t-1} \\ F_{t-1} \end{bmatrix} + BX_t + \varepsilon_t. \tag{273}$$

X_t has size N x 1, u_t has size U x 1, A has size M + J x M + J and B has size M + J x N.[26] ε_t is of size U x 1, and is assumed to to be $\Psi(\mathbf{0}, \Omega_t)$, for some multivariate distribution Ψ[27] and correlation matrix Ω_t. If Ω_t varies with t you need to correct the standard errors for heteroscedasticity, and use a bootstrapping procedure that is robust for heteroscedasticity if you want to draw from the distribution of the parameters of the model.

10.4.1 Estimation

To estimate the model above a two stage method is used. The first step is to estimate the observation equation (272), which can be done in two ways. The first algorithm uses the same steps as in Boivin *et al.* (2009). First the J number of factors are estimated by PCA, abbreviate these factors by $F_t^{(0)}$. Then 272 is estimated by OLS, and we get an estimate of $\rho^{(0)}$. We can then remove the variation of $O_t^{(0)} = O_t$ due to Y_t by $O_t^{(1)} = O_t^{(0)} - \rho^{(0)} Y_t$. To proceed we estimate J number of factors $F_t^{(1)}$ by PCN by using $O_t^{(1)}$ instead of $O_t^{(0)}$. The algorithm is the iterated until $|F_t^{(i)} - F_t^{(i-1)}| < \epsilon$, for some small number ϵ. Lastly equation 272 is estimated by OLS with the last estimated factors.

The second algorithm uses a classification of "slow-moving" and "fast-moving" variables. I.e. "slow-moving" variables are the observable variables not affected by Y_t contemporaneously. We then estimate the factors only using this class of observable variables. The last step is to estimate 272 without Y_t by OLS.

The second step of the estimation procedure is to estimate the VAR. This is estimated by OLS where the factors are substituted by their estimated equivalent.

10.4.2 Bootstrapping FA-VAR models

As the factors are estimated we have uncertainties in the regressors in the FA-VAR model, so we add additional steps to the bootstrapping algorithm presented in section 1.7:

[26] M + J need not be equal to U, as auxiliary variables (equation) can be added to make it into a dynamic system only dependent on one lag, and these equations do not have any error terms.

[27] The classical t-test assumes that Ψ is the multivariate normal distribution.

1. Estimate the FA-VAR equation as discussed in the above section.
2. Re-sample the residual with replacement to construct Q simulated series of ε, i.e. the residual of the FA-VAR equation. Abbreviate each simulated series by $\tilde{\varepsilon}^i$.
3. For each simulated series of ε construct an artificial series of Y and F by $\begin{bmatrix} \tilde{Y}_t^i \\ \tilde{F}_t^i \end{bmatrix} = \hat{A} \begin{bmatrix} Y_{t-1} \\ F_{t-1} \end{bmatrix} + \hat{B} X_t + \tilde{\varepsilon}_t{}^i$, where i is the simulation index and $\hat{A}$ and $\hat{B}$ are the estimates of A and B respectively.
4. Estimate the observation equation as discussed in the above section.
5. Re-sample the residual with replacement to construct Q simulated series of u, i.e. the residual of the observation equation. Abbreviate each simulated series by $\tilde{u}^i$.
6. For each simulated series of u construct an artificial series of O by $\tilde{O}_t^i = \hat{\Lambda}\tilde{F}_t^i + +\hat{\rho}\tilde{Y}_t^i + \tilde{u}_t^i$, where i is the simulation index and $\hat{\Lambda}$ and $\hat{\rho}$ are the estimates of Λ and ρ respectively.
7. Re-estimate the observation equation where O_t is substituted with $\tilde{O}_t^i$ for all i. You will then end up with Q different estimates of F[28], Λ and ρ. I.e. $\hat{F}^i$, $\hat{\Lambda}^i$ and $\hat{\rho}^i$ for $i \in [1:Q]$. This means that you have an estimate of the distribution of the factors F and all the parameters of the vector Λ and ρ.
8. Finally use the factors found in step 7, i.e. $\hat{F}^i$ instead of F, to proceed with the steps of the algorithm in section 1.7 to bootstrap equation 273. In other words, use $\hat{F}^i$ and $\tilde{Y}_t^i$ to bootstrap $\hat{A}$ and $\hat{B}$.

11 Dynamic factor model

We follow Banbura and Modugno (2010) and Banbura *et al.* (2010) for how we implements the dynamic factor model that also can handle mixed frequencies. The problem is to find a map from the N^F factors to the huge data set on the form (observation equation):

$$O_t = \begin{bmatrix} O_t^H \\ O_t^L \end{bmatrix} = \Lambda F_t + \Theta \begin{bmatrix} \epsilon_t^H \\ \epsilon_t^L \\ \vdots \\ \epsilon_{t-P}^L \end{bmatrix} = \Lambda F_t + \Theta \zeta_t, \tag{274}$$

where O_t^H is the observed variable with high frequency with size $G^H \times 1$, O_t^L is the observed variable with low frequency with size $G^L \times 1$, define $G = G^H \times G^L$, then Λ is the factor loadings with size $G \times J$ and Θ is the loadings on the idiosyncratic components with size $G \times (G^H + G^L \times P)$. ϵ_t^H and ϵ_t^L are the idiosyncratic component in the observed variables of high and low frequency respectively, and has size $G \times 1$. F_t has size $J \times 1$, where $J = N^F \times W$. Here $W = max(L, P)$, where L is the number of lags of the state equation and P is the needed number of lags in F_t to be able to handle mixed frequencies in the measurement equation, i.e. F_t is given by

$$F_t = \begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-W} \end{bmatrix}, \tag{275}$$

where we have that f_t contains the N^F unobserved factors of the model at time t. The observed variables are stored in Z_t, and we apply one of the standardizing options, it is applied for each observed variable i

[28] F is the stacked series of F_t over time.

None

$O_t^i = Z_t^i$.

Demean

$B^i = \frac{1}{T}\sum_{t=1}^{T} Z_t^i$, and $O_t^i = Z_t^i - B^i$.

Standardise

$B^i = \frac{1}{T}\sum_{t=1}^{T} Z_t^i$, $\sigma^i = \frac{1}{T-1}\sum_{t=1}^{T}(Z_t^i - B^i)^2$ and $O_t^i = \frac{Z_t^i - B^i}{\sigma^i}$.

Further, the factors are modelled as a VAR

$$F_t = AF_{t-1} + Cu_t. \tag{276}$$

$$A = \begin{bmatrix} A^1 & \cdots & \cdots & A^L \\ 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{277}$$

$$u_t \sim \Phi(0, Q). \tag{278}$$

u_t has size N^F x 1, A has size J x J, C has size J x N^F and where the upper N^F x N^F partition is I, Q is the covariance matrix of the residuals and has size N^F x N^F and Φ is some multivariate distribution, here assumed be multivariate normal.

Finally, we assume

$$\begin{bmatrix} \epsilon_t^H \\ \epsilon_t^L \end{bmatrix} = \begin{bmatrix} \beta^H & 0 \\ 0 & \beta^L \end{bmatrix} \begin{bmatrix} \epsilon_{t-1}^H \\ \epsilon_{t-1}^L \end{bmatrix} + \begin{bmatrix} e_t^H \\ e_t^L \end{bmatrix}. \tag{279}$$

$$\begin{bmatrix} e_t^H \\ e_t^L \end{bmatrix} \sim \Phi\left(0, \begin{bmatrix} \Sigma^H & 0 \\ 0 & \Sigma^L \end{bmatrix}\right) = \Phi\left(0, \Sigma\right). \tag{280}$$

β^H is a diagonal matrix size $G^H \times G^H$ and β^L is a diagonal matrix with size $G^L \times G^L$. e_t^H and e_t^L have sizes $G^H \times 1$ and $G^L \times 1$ respectively. Both Σ^H and Σ^L are diagonal covariance matrices of the residuals e_t^H and e_t^L, and has size $G^H \times G^H$ and $G^L \times G^L$ respectively. This means that each element in ϵ_t^H and ϵ_t^L follows an AR(1) process. Φ is some multivariate distribution, here assumed be multivariate normal. Let use also define $\beta = \begin{bmatrix} \beta^H & 0 \\ 0 & \beta^L \end{bmatrix}$ for later. We want to collapse the model into a state-space representation, which we can do by defining

$$\alpha_t = \begin{bmatrix} F_t \\ \epsilon_t^H \\ \epsilon_t^L \\ \vdots \\ \epsilon_{t-P}^L \end{bmatrix} = \begin{bmatrix} F_t \\ \epsilon_t^H \\ \zeta_t^L \end{bmatrix} = \begin{bmatrix} F_t \\ \zeta_t \end{bmatrix}, \tag{281}$$

$$\eta_t = \begin{bmatrix} u_t \\ e_t \end{bmatrix}, \tag{282}$$

$$T = \begin{bmatrix} A & 0 & 0 \\ 0 & \beta^H & 0 \\ 0 & 0 & T^L \end{bmatrix}, \tag{283}$$

$$T^L = \begin{bmatrix} \beta^L & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{284}$$

$$D = \begin{bmatrix} C & 0 \\ 0 & I \\ 0 & 0 \end{bmatrix}, \tag{285}$$

$$\Omega = \begin{bmatrix} Q & 0 \\ 0 & \Sigma \end{bmatrix}, \tag{286}$$

$$\Upsilon = \begin{bmatrix} \Lambda & \Theta \end{bmatrix}. \tag{287}$$

When dealing with mixed frequencies we have that

$$\Lambda = \beth \circ \begin{bmatrix} \Lambda^H & 0 & \cdots & 0 \\ \Lambda^L & \Lambda^L & \cdots & \Lambda^L \end{bmatrix}, \tag{288}$$

and

$$\Theta = \begin{bmatrix} J_{G,G^H} & \beth \end{bmatrix} \circ \begin{bmatrix} I & 0 & \cdots & 0 \\ 0 & I & \cdots & I \end{bmatrix}, \tag{289}$$

where J_{G,G^H} is a unit matrix of size $G \times G^H$, and $\beth$ is a matrix that will depend on the transformation of the data.[29] See section 11.3 for more on this matrix. In the case we are not dealing with mixed frequency we get

$$\Lambda = \begin{bmatrix} \Lambda^H & 0 & \cdots & 0 \end{bmatrix}, \tag{290}$$

and

$$\Theta = I. \tag{291}$$

We are now able to rewrite the system of 274 and 276 - 280 into

$$O_t = \Upsilon \alpha_t, \tag{292}$$

$$\alpha_t = T\alpha_{t-1} + D\eta_t, \tag{293}$$

where

$$\eta_t \sim \Phi(0, \Omega). \tag{294}$$

[29] $\circ$ is the Hadamard product, i.e. element-wise times operator.

11.1 Expected maximum likelihood

Here we follow Banbura and Modugno (2010) and Banbura *et al.* (2010). The expected maximum likelihood iterates over the following two steps until convergence

1. The expectation of the log-likelihood conditional on the data is calculated using the estimates from the previous iteration, i.e. conditional on the matrices in 283 - 287.
2. The updated estimate of the current iteration is found by maximizing the of the expected log-likelihood from the first step.

To start this iterative algorithm we need a initial values of the parameters of the model, this will be described in the next section, while the two steps of the algorithm will be described in sections 11.1.2 and 11.1.3.

11.1.1 Initialization

Let us define $O = \begin{bmatrix} O_1' & \cdots & O_S' \end{bmatrix}'$, where S is the sample length of the variables in O, i.e. O contains all the observations of the variables in the vector O_t stacked over time. We allow that some of the observation in the matrix O are missing. Let us define $\widehat{O}^j$ the times-series of each observed variable O^j where all missing observations in the middle of sample as been filled in by a cubic spline method. It will have length $\widehat{S} \leq S$. If $\widehat{S} < S$, then we estimate an AR(1) on $\widehat{O}^j$ which is used for backcasting and forecasting away leading and trailing missing observations of each observed time-series. The time series of the observed variable, where the backcasts and forecasts are appended, we abbreviate $\tilde{O}^j$, and the combined dataset constructed in this way for $\tilde{O}$.

The second step is to find an initial value of Λ and F, abbreviate these by Λ^0 and F^0. We use the N first principal components of $\tilde{O}^H$ and the associated loadings as initial estimates for these quantities. Here $\tilde{O}^H$ is a subset of $\tilde{O}$ only containing the high frequency observed variables, and it has size S x G^H. For more on principal components see section 10.1. Let $\tilde{O}^{L,i}$ be the i observed variable at the low frequency, then the unrestricted loadings for the low frequency observed variables can then found by OLS

$$\Lambda_{ur}^{0,L,i} = (F^{0\prime}F^0)^{-1}F^{0\prime}\tilde{O}^{L,i}. \tag{295}$$

The restricted version can be found

$$\Lambda^{0,L,i} = \Lambda_{ur}^{0,L,i} - (F^{0\prime}F^0)^{-1}K^{i\prime}\left(K^i(F^{0\prime}F^0)^{-1}K^{i\prime}\right)K^i\left(K^i\Lambda_{ur}^{0,L,i}\right), \tag{296}$$

where K^i will depend on the transformation of the low frequency observed variable i, see section 11.3 for more on this frequency mapping matrix.

The initial value of T is done by estimation of A_1,...,A_L and β by OLS on 293, where α^0, the initial value of the state variables, is constructed using F^0 and $\epsilon^0 = \tilde{O} - \Lambda^0 F^0$. Zero restrictions on the off diagonal terms are taken into account when the initial value β^0.

11.1.2 Step 1

The expectation step uses the Kalman smoother to filter out the α^j at iteration j given the parameters found at iteration $j-1$. During this step both the mean and the variance of the state variables are filtered out, and the likelihood is calculated. See section 2 for more on the Kalman smoother and how the likelihood is calculated.

11.1.3 Step 2

For simple exposition we here assume that $W = L = P$, but this is not to be taken as a requirement. We also define the selection matrices W_t of the high frequency variables, which is a diagonal matrix, with ones corresponding to the non-missing values in O^H. $W_t^{H,i}$ is on the other hand a scalar, which is 1 for all non-missing values of the high frequency variable i.

Banbura and Modugno (2010) shows that the maximization step leads to the following expressions

$$A_k^j = \left(\sum_{t=1}^{S} \mathbb{E}_j \left[f_t f'_{t-k}\right]\right) \left(\sum_{t=1}^{S} \mathbb{E}_j \left[f_{t-k} f'_{t-k}\right]\right)^{-1}, \tag{297}$$

$$Q^j = \frac{1}{S}\left(\sum_{t=1}^{S} \mathbb{E}_j \left[f_t f'_t\right] - \begin{bmatrix} A_1^j & \cdots & A_W^j \end{bmatrix} \sum_{t=1}^{S} \mathbb{E}_j \left[f_t F'_{t-1}\right]\right), \tag{298}$$

$$\beta_v^j = \left(\sum_{t=1}^{S} \mathbb{E}_j \left[\epsilon_t^v {\epsilon_{t-1}^v}'\right]\right) \left(\sum_{t=1}^{S} \mathbb{E}_j \left[(\epsilon_{t-1}^v)^2\right]\right)^{-1}, \tag{299}$$

$$\Sigma_v^j = \frac{1}{T}\left(\sum_{t=1}^{S} \mathbb{E}_j \left[(\epsilon_t^v)^2\right] - \beta_v \sum_{t=1}^{S} \mathbb{E}_j \left[\epsilon_t^v {\epsilon_{t-1}^v}'\right]\right), \tag{300}$$

$$vec\left(\Lambda^{j,H}\right) = \left(\sum_{t=1}^{S} \mathbb{E}_j \left[f_t f'_t\right] \otimes W_t\right)^{-1} vec\left(\sum_{t=1}^{S} W_t O_t^H \mathbb{E}_j \left[f'_t\right] + W_t \mathbb{E}_j \left[\epsilon_t^H f'_t\right]\right), \tag{301}$$

$$vec\left(\Lambda_{ur}^{j,L,i}\right) = \left(\sum_{t=1}^{T} \mathbb{E}_j \left[F_t F'_t\right] W_t^{L,i}\right)^{-1} \left(\sum_{t=1}^{S} W_t^{L,i} O_t^{L,i} \mathbb{E}_j \left[F'_t\right] + W_t^{L,i} \beth^{L,i} \mathbb{E}_j \left[\zeta_t^L F'_t\right]\right). \tag{302}$$

Finally, we need to map to the restricted factor loading of the low frequency variables in the same way as in 296, and see section 11.3 for the definition of $\beth^{L,i}$.

11.1.4 Convergence

The iterative procedure stops when the change in the log likelihood is less than some selected convergence criterion.

11.1.5 Block restrictions

The algorithm also allow for additional zero restriction on the factor loading matrix Λ, i.e. you can select a subset of observed variables to load on any given factor, f_t^n. In this case we partition the observed variables into separate group based on which columns of the factor loadings where the zero restrictions has been applied. E.g. if a set of observed variables load on all factors, we make that group 1, and if another group load only on the first and the second factors, we make that group 2, and so on. This leads to the following correction of equation 301

$$vec\left(\Lambda^{j,H,g}\right) = \left(\sum_{t=1}^{S} \mathbb{E}_j \left[f_t^g f_t^{g\prime}\right] \otimes W_t^g\right)^{-1} vec\left(\sum_{t=1}^{S} W_t^g O_t^{H,g} \mathbb{E}_j \left[f_t^{g\prime}\right] + W_t^g \mathbb{E}_j \left[\epsilon_t^{H,g} f_t^{g\prime}\right]\right), \tag{303}$$

where the g superscript implies restrictions on $\Lambda^{j,H}$, f_t, ϵ_t^H, W_t and O_t^H that is consistent with the block restrictions of group g. Equal corrections need to be applied to 302 and the initialization stage.

11.2 Time varying parameter bayesian estimation

Another approach is to follow Eraslan and Schröder (2022). In their setup we need to simplify 274, by assuming that $\beta = 0$, assume no lags of ϵ_t^L and $\Theta = I$, which leads to

$$O_t = \Lambda_t F_t + \zeta_t,\ \zeta_t \sim \Phi(0, \Sigma_t), \tag{304}$$

and with the state equation

$$F_t = A_t F_{t-1} + C u_t,\ u_t \sim \Phi(0, Q_t). \tag{305}$$

Let $\lambda_t = vec(\Lambda_t)$ and $a_t = vec(A_t)$. Both evolves as multivariate random walks

$$\lambda_t = \lambda_{t-1} + v_t,\ v_t \sim \Phi(0, V_t). \tag{306}$$

$$a_t = a_{t-1} + w_t,\ w_t \sim \Phi(0, W_t). \tag{307}$$

The estimation procedure is a two-step Kalman filter approach.

1. Set up the priors; $\Lambda_0 \sim N(0, \Lambda_{scale} \times I)$, $\Sigma_0 = \Sigma_{scale} \times I$, $Q_0 = Q_{scale} \times I$, and a Minnesota prior for A_0.
2. Pre-filter the raw data by removing the part due to exogenous variables, and standardize the data.
3. Get a initial estimate of F_t using an expected maximum likelihood approach, as in Stock and Watson (2002).
4. Run the Kalman filters for the parameter estimation. Q_t and Σ_t are updated using exponentially weighted moving average (EWMA) estimators. V_t and W_t are updated using forgetting-factor methods.
5. Run the Kalman filters for the factor estimation to get the final estimate of F_t.
6. Use equation 304 to get an estimate of the missing observations of O_t.

For more details see Eraslan and Schröder (2022).

11.3 Mixed frequency mapping

Here we illustrate the mapping where the high frequency is monthly, while the high frequency is quarterly. The mapping will depend on the transformation of the low frequency variable. The supported cases are:

Level summed
$O_t^{L,i} = X_t^i + X_{t-1}^i + X_{t-2}^i$

Diff summed
$O_t^{L,i} = X_t^i + 2X_{t-1}^i + 3X_{t-2}^i + 2X_{t-3}^i + X_{t-4}^i$

Level average
$O_t^{L,i} = \frac{1}{3}(X_t^i + X_{t-1}^i + X_{t-2}^i)$

Diff average
$O_t^{L,i} = \frac{1}{3}X_t^i + \frac{1}{2}X_{t-1}^i + X_{t-2}^i + \frac{1}{2}X_{t-3}^i + \frac{1}{3}X_{t-4}^i$

End
$O_t^{L,i} = X_t^i$

where Y_t^i is the low frequency variable, while X_t^i represents the low frequency measure of the same variable. This means that the different approximations leads to the mapping matrices:

Level summed

$$\beth^{L,i} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Diff summed

$$\beth^{L,i} = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \end{bmatrix}$$

Level average

$$\beth^{L,i} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \end{bmatrix}$$

Diff average

$$\beth^{L,i} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

End

$$\beth^{L,i} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Let $\beth_j^{L,i}$ mean the j'th column of $\beth^{L,i}$. This means that $\beth$ in equations 288 and 289 is given by

$$\beth = \begin{bmatrix} I & 0 & \cdots & \cdots & 0 \\ \beth_1^{L,1} & \cdots & \beth_5^{L,1} & \cdots & \beth_5^{L,G^H} \end{bmatrix}. \tag{308}$$

We also use the following definition in section**??**

$$\beth^L = \begin{bmatrix} \beth_1^{L,1} & \cdots & \beth_5^{L,1} & \cdots & \beth_5^{L,G^H} \end{bmatrix}. \tag{309}$$

Finally we have that K^i used in equation 296 is given by

$$K^i = \begin{bmatrix} \beth_{-1}^{L,i\prime} & -I \end{bmatrix}, \tag{310}$$

where $\beth_{-1}^{L,i}$ is equal to $\beth^{L,i}$, but with the first element removed.

12 Missing observations

There are different ways of handling missing observations. As of now, this option is available when estimating VAR-models that are estimated with OLS. Be aware that the method here is only needed for estimators that does not handle missing observations in itself. There are two methods for dealing with missing observations in the discussed cases. The first approach uses automatically fitted AR models to fill in for the missing observations, while the second uses a conditional forecast approach. Both have its advantages and disadvantages. The methods will be described in the next subsections. Please be aware that the forecasting routine only handles maximum one missing observation at the end of the sample for any variable of the model.

12.1 AR

For each series that has missing observations an AR(p) model is estimated on the data up to the time of the first missing observation, and then this model is used to forecast the missing observation. Finally the model is estimated using the estimated values of the missing observations.

When to use?

1. If the missing observations are only at the end of the sample.

2. If some of the exogenous variables of the model have missing observations.

When not to use?

1. If the missing observations are in the middle of the sample.
2. If none of the exogenous variables of the model have missing observation. The conditional forecast method is then more appropriate.

12.1.1 Nowcasts

Nowcasts are produced by using the forecasts from the AR(p) models. Be it point or density nowcasts.

12.2 Conditional forecast

This approach uses conditional forecast routines inspired by work by Maih (2010). In this method we utilize the observations on the non-missing variables to fill in for the missing observations. In this way this approach utilizes more information than the AR(p) approach. We will document the approach using a VAR example. We start out with a model on the form 156, where X_t has size 0 x 1, i.e. there are no exogenous variables included in the model. This is just to simplify the example. First the model is estimated on the balanced data set. We then must find the theoretical autocorrelation matrices, Σ_k, for $k = 1, ..., L$, where L is the number of lags of the VAR. See section 14.4 for how these are calculated. Stack the autocorrelation matrices Σ_k for the different k's as follows

$$\Sigma = \begin{bmatrix} \Sigma_0 & \Sigma_1 & \cdots & \cdots & \cdots & \Sigma_L \\ \Sigma_1 & \ddots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \Sigma_0 & \Sigma_1 & \cdots & \vdots \\ \vdots & \vdots & \Sigma_1 & \Sigma_0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_L & \cdots & \cdots & \cdots & \cdots & \Sigma_0 \end{bmatrix}. \tag{311}$$

And stack the observations of the variables as follows

$$X = \begin{bmatrix} X_{T-L} \\ X_{T-L+1} \\ \vdots \\ X_T \end{bmatrix}, \tag{312}$$

where T is the end of the sample. As some of the observations in X are missing, we must partition the variables into the sets $\mathbf{X}_1$ and $\mathbf{X}_2$. Where $\mathbf{X}_2$ is the set of variables with a known value with dimension I, and $\mathbf{X}_1$ is the set of variable with a missing value with dimension $N - I$. If we re-order them according to $\tilde{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, i.e. with all $X_k \in \mathbf{X}_2$ stacked in a vector X_2 and with all $X_i \in \mathbf{X}_1$ stacked in a vector X_1, we can partition the correlation matrix as $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}$. Σ_{11} has size $N - I \times N - I$, Σ_{21} has size $I \times N - I$, Σ_{12} has size $N - I \times I$ and Σ_{22} has size $I \times I$. This means that we can find the most probable value of the missing observation given the observation we have and the model from

$$X_1 = \Sigma_{12}\Sigma_{22}^{-1}X_2. \tag{313}$$

Finally the model is estimated using the estimated values of the missing observations.

When to use?

1. If there is missing observation at the end of the sample for the dependent variables of the model.
2. If none of the exogenous variables of the model have missing observations.

When not to use?

1. If the missing observations are in the middle of the sample.
2. If some of the exogenous variables of the model has missing observation. This method does not apply, see the AR method instead.

12.2.1 Nowcasts

When producing the nowcasts we use the estimated parameters where the missing observations are filled in, i.e. we do an update on the most probable values of the missing observations given the full sample estimates. Lets abbreviate the update of Σ as $\tilde{\Sigma}$. Then the final mean nowcasts are given by

$$\tilde{X}_1 = \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}X_2. \tag{314}$$

If we are interested in density forecast we can draw them by using that they are distributed as $N(\tilde{X}_1, \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21})$. If you want to take into account parameter uncertainty the steps here are looped for each parameter draw. See section 12.3 for how the parameters are drawn in this case.

12.3 Bootstrapping

The following algorithm is used to bootstrap the parameters of the models with missing observations:

1. Estimate the model using one of the methods put forward in sections 12.1 or 12.2.
2. Re-sample the residual with replacement to construct Q simulated series of the residuals.
3. For each of the Q simulated series of the residual construct a simulated series of the dependent variables of the model.
4. Re-estimate the model where the actual observations of the dependent variables are substituted with one simulation of the dependent variables, and where the originally missing observations are stripped. Repeat for all Q simulations. You will then end up with Q different estimates of the parameters of the model. This means that you have an estimate of the distributions of all the parameters of the model.
5. Use the distributions of the parameters to construct simulations of the object of interest, and use the $\alpha/2$ and $1-\alpha/2$ percentiles to calculate two-sided α confidence intervals.

12.4 Combination of forecasts and nowcasts

We are able to account for possible combinations of forecasts and nowcasts. For instance, let us presume that we have a bivariate VAR(2) model with inflation and GDP. As often can be the case, we have one additional observation of inflation compared to GDP. In such a case, we are able to combine the nowcast of inflation to the forecast of GDP. This will further be accounted for if the model is used to produce combined forecasts.

13 DSGE

Dynamic stochastic general equilibrium (DSGE) models can be put on the form

$$E_t[F(X_{t+1}^+, X_t, X_{t-1}^-, U_t, \theta_t)] = 0, \tag{315}$$

where F is a set of possibly non-linear equations of the model, X_t are the endogenous variables of the model, X_{t+1}^+ and X_{t-1}^- are a subset of X_t that appear with a lead and lag respectively in the model. U_t are the exogenous variables of the model, which cannot appear with any order of lead or lag, with size n^i. Let the number of equations be given by M, which must also be the number of endogenous variables ($= n$).[30] θ_t is the parameters of the model that may change unexpectedly over time. In what follows we assume that $\theta_t \in \mathbb{R}^P$, and that there is countable number of different regimes, i.e. $\theta_t \in \{\theta^1, ..., \theta^s\}$ for some finite $s \in \mathbb{N}$. $P \in \mathbb{N}$ will be the number of parameters of the model.

The timing used here is the same as Dynare[31], which means that variables decided in period t have a time subscript t. This means that stock variable that are decided at the end-of-period must have time subscript $t-1$.

We follow Villemot (2011) and use the following classification of the variables

Static endogenous variables:

Those that appear only at the current period. Numbered as n_s, and their location in X_t are l_j^s, for $j = \{1, ..., n_s\}$

Purely forward endogenous variables

Those that appear at the future period, possible at the current period, but not at the previous period. Numbered as n_{++}, and their location in X_t are l_j^{++}, for $j = \{1, ..., n_{++}\}$

Purely backward endogenous variables

Those that appear at the previous period, possible at the current period, but not at the future period. Numbered as n_{--}, and their location in X_t are l_j^{--}, for $j = \{1, ..., n_{--}\}$

Mixed endogenous variables

Those that appear at the previous period, possible at the current period, and at the future period. Numbered as n_m, and their location in X_t are l_j^m, for $j = \{1, ..., n_m\}$

We must then have that $n_s + n_{++} + n_{--} + n_m = n$. In addition we define

Forward endogenous variables

Those that appear at the future period. Numbered as n_+, and their location in X_t are l_j^+, for $j = \{1, ..., n_+\}$. $n_+ = n_m + n_{++}$.

Backward endogenous variables

Those that appear at the previous period. Numbered as n_-, and their location in X_t are l_j^-, for $j = \{1, ..., n_-\}$. $n_- = n_m + n_{--}$.

Dynamic endogenous variables

All non-static endogenous variables. Numbered as n_d, and their location in X_t are l_j^d, for $j = \{1, ..., n_d\}$. $n_d = n - n_s$.

[30] Except if the model is solved with optimal monetary policy.

[31] For more on dynare go to http://www.dynare.org/.

We denote ζ_j^+, for $j = \{1, ..., n_m\}$, the location of the mixed endogenous variable inside the locations of the forward endogenous variables (l_j^+), i.e. such that $l^+_{\zeta_j^+}$ is a mixed endogenous variable. In similar fashion we define ζ_j^- as the location of the mixed endogenous variable inside the locations of the backward endogenous variables. Similarly we define π_j^+ (resp. π_j^-) for the purely forward (resp. purely backward) endogenous variables inside l_j^+ (resp. l_j^-).

Finally we define the state variables by the union of backward endogenous variables at the previous period (X_{t-1}^-) and the exogenous variable at the current period (U_t).

If the model include more than one lag we append additional auxiliary variables to the model. Thus for every endogenous variable with more than one lag, abbreviate these with x_t, we do the following

1. Find the max number of lags of that variable, and abbreviate it by k.
2. Introduce $k-1$ new endogenous variables, z_t^j, for $j = 1, ..., k-1$. These new variables get the name of the original series with the prefix AUX_LAG_ and postfix _j is added, for $j = \{1, ..., k-1\}$.
3. Add $k-1$ new equations to the model

$$\begin{cases} z_t^1 = x_{t-1}, \\ z_t^j = z_{t-1}^{j-1}, \text{ for } j = \{1, ..., k-1\}. \end{cases} \tag{316}$$

4. Replace all occurrences of x_{t-1-j} in the original model by z_{t-1}^j in the transformed model for all $j = \{1, ..., k-1\}$.

We do not support variables with more than one lead. Use Dynare or RISE toolbox in this case.

In this way we have transformed the model into the form required by equation 315.

In this documentation of DSGE models we will use the notation that the first order derivative of a function $Q(Y_t)$ is given by $H_{Y_t} = \frac{\partial Q}{\partial Y_t}$.

13.1 Standard 1st order solution

To be able to solve the non-linear system in 315 we need to do an approximation. In this section we document the 1st order approximation solution. First we need a well defined steady-state of the system

$$U_{ss} = 0, \tag{317}$$

$$F(X_{ss}^{+,r}, X_{ss}^r, X_{ss}^{-,r}, 0, \theta^r) = 0 \ \forall r \in [1, s]. \tag{318}$$

r is then the regime we want to solve the model in. Finding the deterministic steady-state involves solving a multivariate non-linear system. This solution can be provided by its own MATLAB .m file[32], or using a numerical solver.

We want to find a solution on the form

$$X_t = G(X_{t-1}^-, U_t, \theta_t). \tag{319}$$

From here on we drop the reference to the regime of the parameters (θ_t) in the functions F and G, as the steps from here on will be the same for each regime, since the switches between regimes are unexpected. The function G must satisfy

[32] See section 13.6.1 for more on this.

$$E_t[F(G^+(G^-(X_{t-1}^-,U_t),0),G(X_{t-1}^-,U_t),X_{t-1}^-,U_t,\theta)]=0, \tag{320}$$

where $X_t^+ = G^+(X_{t-1}^-,U_t)$ and $X_t^- = G^-(X_{t-1}^-,U_t)$. A first order solution around the steady-state is then given by

$$F(X_{ss}^+,X_{ss},X_{ss}^-,U_{ss})+F_{X_{t+1}^+}\left(G^+_{X_{t-1}^-}(G^-_{X_{t-1}^-}X_{t-1}^- + G^-_{U_t}U_t)\right)+F_{X_t}\left(G_{X_{t-1}^-}X_{t-1}^- + G_{U_t}U_t\right)+F_{X_{t-1}^-}X_{t-1}^- \\ +F_{U_t}U_t=0. \tag{321}$$

where we have used that H^+ and H^- are the restrictions of the general function H to the subspaces X^+ and X^- of X respectively. By using the definition of the steady-state and reordering we get

$$\left(F_{X_{t+1}^+}G^+_{X_{t-1}^-}G^-_{X_{t-1}^-}+F_{X_t}G_{X_{t-1}^-}+F_{X_{t-1}^-}\right)X_{t-1}^- + \left(F_{X_{t+1}^+}G^+_{X_{t-1}^-}G^-_{U_t}+F_{X_t}G_{U_t}+F_{U_t}\right)U_t=0. \tag{322}$$

13.1.1 Recovering $G_{X_{t-1}^-}$

Equation 322 implies that

$$F_{X_{t+1}^+}G^+_{X_{t-1}^-}G^-_{X_{t-1}^-}+F_{X_t}G_{X_{t-1}^-}+F_{X_{t-1}^-}=0, \tag{323}$$

which is the same as

$$F_{X_{t+1}^+}X_{t+1}^+ + F_{X_t}X_t + F_{X_{t-1}^-}X_{t-1}^- = 0. \tag{324}$$

As Villemot (2011) we take care of the static variables by constructing a n x n^s matrix S, by selecting the columns of F_{X_t} for the static endogenous variables, i.e. $S_{i,j} = F_{X_t,i,l_j^s}$. A QR decomposition gives $S = QR$, where Q is an n x n orthogonal matrix and R an n x n^s upper triangular matrix. We assume here that

$$rank(R) = n^s. \tag{325}$$

We can then rewrite equation 324 as

$$A^+X_{t+1}^+ + A^0X_t + A^-X_{t-1}^- = 0, \tag{326}$$

where $A^+ = Q'F_{X_{t+1}^+}$, $A^0 = Q'F_{X_t}$ and $A^- = Q'F_{X_{t-1}^-}$. By using the QR decomposition we have made the columns of the static endogenous variables in A^0 are zero in their lower part.

13.1.1.1 Non-static endogenous variable

We define $\tilde{A}^+$ to be the last n^d rows of A^+, and $\tilde{A}^-$ in the same way. We also define $\tilde{A}^{0+}$ to be a sub matrix of A^0 where only the last n^d rows and the columns of the forward endogenous variables are kept ($\tilde{A}^{0+}_{i,j} = A^0_{n^s+i,l_j^+}$), and $\tilde{A}^{0-}$ to be a sub matrix of A^0 where only the last n^d rows and such that the columns of the purely backward endogenous variable are taken from A^0 ($\tilde{A}^{0-}_{i,\pi_j^-} = A^0_{n^s+i,l_j^{--}}$), while the rest of the columns are 0. Then we can take the n^d lower rows of 324, and rewrite it to get

$$\tilde{A}^+X_{t+1}^+ + \tilde{A}^{0+}X_t^+ + \tilde{A}^{0-}X_t^- + \tilde{A}^-X_{t-1}^- = 0. \tag{327}$$

By doing this the static endogenous variable no longer appear in the problem to be solved. We can collapse 327 into

$$\underbrace{\begin{bmatrix}\tilde{A}^{0-} & \tilde{A}^{+}\\ I^{-} & 0\end{bmatrix}}_{D}\begin{bmatrix}X_t^{-}\\ X_{t+1}^{+}\end{bmatrix}=\underbrace{\begin{bmatrix}-\tilde{A}^{-} & \tilde{A}^{0+}\\ 0 & I^{+}\end{bmatrix}}_{E}\begin{bmatrix}X_{t-1}^{-}\\ X_t^{+}\end{bmatrix}, \tag{328}$$

where I^- is an n^m x n^- selection matrix for the mixed endogenous variables, i.e. $I^-_{i,\zeta_i^-}=1$, and 0 otherwise. Similarly we have that I^+ is an n^m x n^+ selection matrix for the mixed endogenous variables, i.e. $I^+_{i,\zeta_i^+}=1$, and 0 otherwise. By this D and E are square matrices with size $n^{++}+n^{--}+2n^m$. To proceed we follow Klein (2000), and we do a generalized Schur decomposition of the pencil (D,E)

$$D=QTZ, \tag{329}$$

$$D=QSZ. \tag{330}$$

where T is upper triangular, S quasi upper triangular, and Q and Z are orthogonal matrices. The decomposition is ordered so that the stable generalized eigenvalues (modulus less than 1) are located in the upper left corner of T and S. We can then write equation 328 as

$$\underbrace{\begin{bmatrix}T_{11} & T_{12}\\ 0 & T_{22}\end{bmatrix}}_{T}\underbrace{\begin{bmatrix}Z_{11} & Z_{12}\\ Z_{21} & Z_{22}\end{bmatrix}}_{Z}\begin{bmatrix}X_t^{-}\\ X_{t+1}^{+}\end{bmatrix}=\underbrace{\begin{bmatrix}S_{11} & S_{12}\\ 0 & S_{22}\end{bmatrix}}_{S}\underbrace{\begin{bmatrix}Z_{11} & Z_{12}\\ Z_{21} & Z_{22}\end{bmatrix}}_{Z}\begin{bmatrix}X_{t-1}^{-}\\ X_t^{+}\end{bmatrix}, \tag{331}$$

where we have partitioned the matrices T and S such that T_{11} and S_{11} are square and contain the stable generalized eigenvalues, while T_{22} and S_{22} are square and contain the explosive generalized eigenvalues.

By introducing the auxiliary variables

$$\begin{bmatrix}Y_t^{s}\\ Y_t^{e}\end{bmatrix}=\underbrace{\begin{bmatrix}Z_{11} & Z_{12}\\ Z_{21} & Z_{22}\end{bmatrix}}_{Z}\begin{bmatrix}X_t^{-}\\ X_{t+1}^{+}\end{bmatrix} \tag{332}$$

we can write 332 as

$$\underbrace{\begin{bmatrix}T_{11} & T_{12}\\ 0 & T_{22}\end{bmatrix}}_{T}\begin{bmatrix}Y_t^{s}\\ Y_t^{e}\end{bmatrix}=\underbrace{\begin{bmatrix}S_{11} & S_{12}\\ 0 & S_{22}\end{bmatrix}}_{S}\begin{bmatrix}Y_{t-1}^{s}\\ Y_{t-1}^{e}\end{bmatrix}. \tag{333}$$

Here the partition $\begin{bmatrix}Y_t^{s}\\ Y_t^{e}\end{bmatrix}$ is done so to make it conforms to the partition of T and S. To exclude explosive solution we impose

$$Y_t^{e}=0, \tag{334}$$

which means that we can find the solution of Y_t^s from 333

$$Y_t^{s}=T_{11}^{-1}S_{11}Y_{t-1}^{s}. \tag{335}$$

Then by using equation 332 we can find the final solution to be

$$X_t^+ = Z_{21} Z_{11}^{-1} X_{t-1}^-, \tag{336}$$

$$X_t^- = Z_{11} T_{11}^{-1} S_{11} Z_{11}^{-1} X_{t-1}^-. \tag{337}$$

By using that $X_t^- = G_{X_{t-1}^-}^- X_{t-1}^-$ and $X_t^+ = G_{X_{t-1}^-}^+ X_{t-1}^-$ we must have

$$G_{X_{t-1}^-}^+ = Z_{21} Z_{11}^{-1}, \tag{338}$$

$$G_{X_{t-1}^-}^- = Z_{11} T_{11}^{-1} S_{11} Z_{11}^{-1}. \tag{339}$$

If $\left|S_{n^-,n^-}\right| > \left|T_{n^-,n^-}\right|$ or $\left|S_{n^-+1,n^-+1}\right| < \left|T_{n^-+1,n^-+1}\right|$ the Blanchard and Kahn (1980) order condition will fail, i.e. the requirement to have as many explosive eigenvalues as forward or mixed endogenous variables fail. If $rank(Z_{11}) < n^-$ the Blanchard and Kahn (1980) rank condition will fail.

13.1.1.2 Static endogenous variable

We define $\hat{A}^+$ to be the first n^s rows of A^+, and $\hat{A}^-$ in the same way. We also define $\tilde{A}^{0s}$ to be a sub matrix of A^0 where only the first n^s rows and the columns of the static endogenous variables are kept ($\tilde{A}^{0s}_{i,j} = A^0_{i,l_j^s}$), and $\tilde{A}^{0d}$ to be a sub matrix of A^0 where only the first n^s rows and such that the columns of the dynamic endogenous variable are taken from A^0 ($\tilde{A}^{0-}_{i,j} = A^0_{i,l_j^d}$). Then we can take the n^s lower rows of 324, and rewrite it to get

$$\hat{A}^+ X_{t+1}^+ + \hat{A}^{0s} X_t^s + \hat{A}^{0d} X_t^d + \hat{A}^- X_{t-1}^- = 0. \tag{340}$$

Note that A^{0s} is a square upper triangular matrix, and is invertible because of assumption 325. We can rewrite 340 as

$$\hat{A}^+ G_{X_{t-1}^-}^+ G_{X_{t-1}^-}^- X_{t-1}^- + \hat{A}^{0s} X_t^s + \hat{A}^{0d} G_{X_{t-1}^-}^d X_{t-1}^- + \hat{A}^- X_{t-1}^- = 0, \tag{341}$$

where $G_{X_{t-1}^-}^d$ is the restrictions of $G_{X_{t-1}^-}$ to the dynamic endogenous variables which is obtained by combining $G_{X_{t-1}^-}^+$ and $G_{X_{t-1}^-}^-$. By solving for X_t^s we get

$$X_t^s = (\hat{A}^{0s})^{-1} \left[\hat{A}^+ G_{X_{t-1}^-}^+ G_{X_{t-1}^-}^- + \hat{A}^{0d} G_{X_{t-1}^-}^d + \hat{A}^- \right] X_{t-1}^-, \tag{342}$$

which means that

$$G_{X_{t-1}^-}^d = (\hat{A}^{0s})^{-1} \left[\hat{A}^+ G_{X_{t-1}^-}^+ G_{X_{t-1}^-}^- + \hat{A}^{0d} G_{X_{t-1}^-}^d + \hat{A}^- \right]. \tag{343}$$

13.1.2 Recovering G_{U_t}

Equation 322 implies that

$$F_{X_{t+1}^+} G_{X_{t-1}^-}^+ G_{U_t}^- + F_{X_t} G_{U_t} + F_{U_t} = 0 \tag{344}$$

which can be reordered to

$$\left[F_{X_{t+1}^+} G_{X_{t-1}^-}^+ J^- + F_{X_t} \right] G_{U_t} + F_{U_t} = 0, \tag{345}$$

where J^- is an n^- x n matrix selecting only the backward endogenous variables. We can therefore solve for G_{U_t} to get

$$G_{U_t} = -\left[F_{X^+_{t+1}} G^+_{X^-_{t-1}} J^- + F_{X_t}\right]^{-1} F_{U_t}. \tag{346}$$

13.1.3 Companion form

We now have the solution on the form

$$X_t = G_{X^-_{t-1}} X^-_{t-1} + G_{U_t} U_t, \tag{347}$$

but we want it on the form

$$X_t = AX_{t-1} + CU_t. \tag{348}$$

This means that

$$A_{i,l^-_j} = G_{X^-_{t-1},i,j}, \tag{349}$$

where the rest of the elements of A are 0. Finally we get that

$$C = G_{U_t}. \tag{350}$$

13.2 Optimal monetary policy: Commitment

Here we uses the same algorithm as presented in the section 13.1, i.e. Klein (2000). In the case of optimal monetary policy a loss function must be specified, see section 13.4 for how to implement it in the model file. The loss function must be a quadratic loss function on the form

$$\sum_{t=0}^{\infty} \beta^t X_t' W X_t, \tag{351}$$

where β is the central banks discount rate and W is the weights on the different target variables.

First we take the first order derivative of the problem 315 with respect to the inputs except the parameters to get

$$A_1 E_t\left[X_{t+1}\right] + A_0 X_t + A_{-1} X_{t-1} + BU_t = 0, \tag{352}$$

where have that $A_1 = F_{X_{t+1}}$, $A_0 = F_{X_t}$, $A_{-1} = F_{X_{t-1}}$ and $B = F_{U_t}$. Then we need to optimize 351 s.t. 352. To do this we set up the lagrangian

$$\mathcal{L} \equiv E_{-1} \sum_{t=0}^{\infty} \beta^t \left\{ X_t' W X_t + \lambda_{t-1}' \beta^{-1} A_1 X_t + \lambda_t' \left[A_1 X_{t+1} + A_0 X_t + A_{-1} X_{t-1} + BU_t\right] \right\}, \tag{353}$$

with λ_{-1} and X_{-1} given.

The first order conditions can then be given by

$$\frac{\partial \mathcal{L}}{\partial \lambda_t} = A_1 E_t\left[X_{t+1}\right] + A_0 X_t + A_{-1} X_{t-1} + BU_t = 0, \tag{354}$$

$$\frac{\partial \mathcal{L}}{\partial X_t} = 2WX_t + \beta A'_{-1} E_t \left[\lambda_{t+1}\right] + A'_0 \lambda_t + \beta^{-1} A'_1 \lambda_{t-1} = 0. \tag{355}$$

We can stack these first order conditions to get

$$H_1 E[Y_{t+1}] + H_0 Y_t + H_{-1} Y_{t-1} + DU_t = 0, \tag{356}$$

where

$$Y_t = \begin{bmatrix} X_t \\ \lambda_t \end{bmatrix}, \tag{357}$$

$$H_1 = \begin{bmatrix} A_1 & 0 \\ 0 & \beta A'_{-1} \end{bmatrix}, \tag{358}$$

$$H_0 = \begin{bmatrix} A_0 & 0 \\ 2W & A'_0 \end{bmatrix}, \tag{359}$$

$$H_{-1} = \begin{bmatrix} A_{-1} & 0 \\ 0 & \beta^{-1} A'_1 \end{bmatrix}, \tag{360}$$

$$D = \begin{bmatrix} B \\ 0 \end{bmatrix}. \tag{361}$$

From this representation it is easy to apply the algorithm by Klein (2000) to solve the model under rational expectations.

13.3 Optimal monetary policy: Loose commitment

We use the algorithm presented in Debortoli *et al.* (2010) to solve a DSGE model with optimal monetary policy under loose commitment. In this case a loss function must be specified, see section 13.4 for how to implement it in the model file. The loss function must be a quadratic loss function on the form

$$\sum_{t=0}^{\infty} \beta^t X'_t W X_t, \tag{362}$$

where β is the central banks discount rate and W is the weights on the different target variables.

Debortoli *et al.* (2010) defines the concept of loose commitment. In this it is meant that the central bank do have the possibility to commit, but may from time to time revise their plan (be discretionary). The occurence of re-optimization is in their setup an exogenous two-state Markov stochastic process

$$\eta_t = \begin{cases} 1 & \text{with probability } \gamma \\ 0 & \text{with probability } 1-\gamma, \end{cases} \tag{363}$$

where $0 < \gamma < 1$. If $\eta_t = 1$ the previous commitments are honored, otherwise the central bank re-optimize their objective. In the special cases where $\gamma = 1$ and $\gamma = 0$ we have that the central bank is fully committed or re-optimizes at all periods (fully discretionary).

First we take the first order derivative of the problem 315 with respect to the inputs except the parameters to get

$$A_1 E_t\left[X_{t+1}\right] + A_0 X_t + A_{-1} X_{t-1} + B U_t = 0, \tag{364}$$

where have that $A_1 = F_{X_{t+1}}$, $A_0 = F_{X_t}$, $A_{-1} = F_{X_{t-1}}$ and $B = F_{U_t}$.

Following Tambalotti (2004) and Debortoli *et al.* (2010), the policymaker's problem can be written as

$$X'_{-1} P X_{-1} + d = min_{\{X_t\}_{t=0}^{\infty}} E_{-1} \sum_{t=0}^{\infty} (\beta\gamma)^t \left[X'_t W X_t + \beta(1-\gamma)(X'_t P X_t + d)\right] \tag{365}$$

$$\text{s.t. } \gamma A_1 E_t\left[X_{t+1}\right] + (1-\gamma) A_1 E_t\left[X^r_{t+1}\right] + A_0 X_t + A_{-1} X_{t-1} + B U_t = 0 \ \forall t \geq 0. \tag{366}$$

At each period the policymaker sum over the period loss function and a part indicating the value the policymaker obtains if a re-optimization occurs in the next period. The constraints that the policymaker face depends on the expectation of future variables that consists of two terms. The first term, X_{t+1}, relates to the allocation prevailing when current plans are honored, while the second term, X^r_{t+1}, refers to the choices made in period $t+1$ if re-optimization occurs. As in Debortoli *et al.* (2010), $E_t\left[X^r_{t+1}\right]$ only depend on X_t

$$E_t\left[X^r_{t+1}\right] = \tilde{H} X_t. \tag{367}$$

which leads to the lagrangian

$$\begin{aligned} \mathcal{L} \equiv & E_{-1} \sum_{t=0}^{\infty} (\beta\gamma)^t \Big\{ X'_t \left[W + (1-\gamma)\beta P\right] X_t + \lambda'_{t-1} \beta^{-1} A_1 X_t + \\ & \lambda'_t \left[A_{-1} X_{t-1} + \left(A_0 + (1-\gamma) A_1 \tilde{H}\right) X_t + B U_t\right] \Big\}, \end{aligned} \tag{368}$$

with λ_{-1}, $\tilde{H}$ and X_{-1} given.

The first order conditions can then be given by

$$\frac{\partial \mathcal{L}}{\partial \lambda_t} = \left[A_0 + (1-\gamma) A_1 \tilde{H}\right] X_t + \gamma A_1 E_t\left[X_{t+1}\right] + A_{-1} X_{t-1} + B U_t = 0, \tag{369}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial X_t} = & 2 W X_t + \beta(1-\gamma) A'_{-1} E_t\left[\lambda^r_{t+1}\right] + (A_0 + (1-\gamma) A_1 \tilde{H})' \lambda_t + \\ & \mathcal{I} \beta^{-1} A'_1 \lambda_{t-1} + \beta\gamma A'_{-1} E_t\left[\lambda_{t+1}\right] = 0. \end{aligned} \tag{370}$$

The λ^r_{t+1} term comes from the derivative of the value function w.r.t. Y_t, i.e. from the envelop theorem we have

$$\frac{\partial X'_t P X_t}{\partial X_t} = 2 P X_t = A'_{-1} E_t\left[\lambda^r_{t+1}\right]. \tag{371}$$

Finally, the term $\mathcal{I}$ is an indicator function

$$\mathcal{I} = \begin{cases} 0 & \text{if } \gamma = 0 \\ 1 & \text{otherwise.} \end{cases} \tag{372}$$

Following Debortoli *et al.* (2010) a undetermined coefficients approach is then used to solve the problem.

First we guess a solution to the problem in the following form

$$\begin{bmatrix} X_t \\ \lambda_t \end{bmatrix} = \underbrace{\begin{bmatrix} H_{yy} & H_{y\lambda} \\ H_{\lambda y} & H_{\lambda\lambda} \end{bmatrix}}_{H} \begin{bmatrix} X_{t-1} \\ \lambda_{t-1}+ \end{bmatrix} \underbrace{\begin{bmatrix} G_y \\ G_\lambda \end{bmatrix}}_{G} U_t, \tag{373}$$

where the matrices H and G depend on the unknown matrix $\tilde{H}$.

When re-optimization occurs $\lambda_{t-1} = 0$, which has been proved by Debortoli and Nunes (2010). By using this and equation 373 it follows that

$$X_t^r = H_{yy} X_{t-1} + G_y U_t. \tag{374}$$

Moving this equation one period forward and taking expectations one obtains

$$E[X_{t+1}^r] = H_{yy} X_t, \tag{375}$$

which means that in an rational expectation equilibrium must have

$$\tilde{H} = H_{yy}. \tag{376}$$

By substitution of this plus the expressions we get when leading equation 373, and take expectation of those equations again, we get

$$\Gamma_0 \begin{bmatrix} X_t \\ \lambda_t \end{bmatrix} = \Gamma_1 \begin{bmatrix} X_{t-1} \\ \lambda_{t-1} \end{bmatrix} + \Gamma_v U_t, \tag{377}$$

with

$$\Gamma_0 \equiv \begin{bmatrix} A_0 + A_1 H_{yy} & \gamma A_1 H_{y\lambda} \\ 2W + \beta A'_{-1} H_{\lambda y} & A'_0 + (1-\gamma) H'_{yy} A'_1 + \beta\gamma A_{-1} H_{yy} \end{bmatrix}, \tag{378}$$

$$\Gamma_1 \equiv \begin{bmatrix} A_{-1} & 0 \\ 0 & \mathcal{I}\beta^{-1} A'_1 \end{bmatrix}, \tag{379}$$

$$\Gamma_v \equiv \begin{bmatrix} B \\ 0 \end{bmatrix}. \tag{380}$$

The resulting law of motion is

$$\begin{bmatrix} X_t \\ \lambda_t \end{bmatrix} = -\Gamma_0^{-1}\Gamma_1 \begin{bmatrix} X_{t-1} \\ \lambda_{t-1} \end{bmatrix} - \Gamma_0^{-1}\Gamma_v U_t. \tag{381}$$

The final step is to verify that $H = \Gamma_0^{-1}\Gamma_1$. The steps of the algorithm proposed by Debortoli *et al.* (2010) can then be summarized as

1. Using a guess H_{guess}, form Γ_0 and Γ_1.
2. Compute $H = \Gamma_0^{-1}\Gamma_1$.
3. Check if $|H - H_{guess}| < \epsilon$, where $|.|$ is a distance measure and $\epsilon > 0$. If the guess and the solution have converged, proceed to step 4. Otherwise, update the guess as $H_{guess} = H$ and repeat steps 1-3 until convergence.
4. Finally, form Γ_v and compute $G = \Gamma_0^{-1}\Gamma_v$.

13.3.1 Initialization of H_{guess}

You can initialize H_{guess} with one of the following

Zeros

$H_{guess} = 0$

Backward

$H_{guess} = A_0^{-1} A_{-1}$

Commitment

Sets $H_{guess} = H_c$, where H_c is equal to the solution found when the model is solved under commitment using the algorithm of section 13.2.

Last

Default option. Uses H_c if not already solved, but uses the last found solution if re-solved, which is done for example during estimation or when solving the model with more regimes (see section 13.10).

13.3.2 Companion form

We are interested in the companion form of the solution on the form

$$X_t = AX_{t-1} + CU_t. \tag{382}$$

This can be found by redefine X_t by $\begin{bmatrix} X_t \\ \lambda_t \end{bmatrix}$ and set $A = H$ and $C = G$.

13.4 Model file

The model file must be a .nb (or .mod) file, and it must consists of the following blocks

endogenous

List the endogenous variables of the models. Can have lags and leads, but for now maximum one lead. Finish this block with a semi-colon.

exogenous

List the exogenous variables of the models. Can not have lags or leads. Finish this block with a semi-colon.

parameters

List of the parameter names of the model. Finish this block with a semi-colon. This is an optional block.

model

List the equations of the model. Each equation must end with a semi-colon. One equation can span more lines. The equation is typed in using the endogenous variable names, exogenous variable names, parameter names or numbers (hard coded parameters). To indicate lead or lags use the syntax varName(+1) and varName(-1). Finish this block with a end command (on its own line).

observables

List the observables variables of the models. These are the variables you declare to have data on when doing filtering or estimation. Finish this block with a semi-colon. This is an optional block.

planner_objective

Syntax: planner_objective{options} lossFunction;. The options can be discount and/or commitment. Here follows an example; planner_objective{discount=0.99,commitment=0} 0.5*(INF^2+LAM_Y*Y^2); As you see from the example the command must end with a semi-colon. In this example the central banks discount factor is set to 0.99 ($\beta = 0.99$) and probability of commitment is set to 0 ($\gamma = 0$). The loss function is quadratic and has weight 1 on the variable INF and weight LAM_Y on Y. LAM_Y must here be declared as a parameter in the parameters block, and INF and Y must be endogenous variables of the model.

breakPoint

Syntax: breakPoint{date} param1,param2;. The date input must be a date on the same frequency as the data of the model, see NB toolbox manual for more on the supported date format. Example: breakPoint{2018Q1}LAM_Y,LAM_X;. To assign the values after of this break-point use the original parameter names and append it with _date, e.g. in the example we want to assign LAM_Y_2018Q1 a value. The value will change unexpected at the date you specify, i.e. in our example at 2018Q1.

In the model block you can use # to define temporary parameters that are substituted in for before solving the model. The syntax is as follows;

```
#delta_1 = (1-delta);
k = delta_1*k(-1) + i;
```

Be aware that it will be substituted just as you define it! So the above give you

```
k = (1-delta)*k(-1) + i;
```

Had you defined it as follows

```
#delta_1 = 1-delta;
k = delta_1*k(-1) + i;
```

It will wrongly result in

```
k = 1-delta*k(-1) + i;
```

13.4.1 Example

```
endogenous
PAI, Y, R, RR

exogenous
EPS_R

parameters
betta, eta, kappa, rhor, sigr, mu, psi

model
    1 = betta*(1-.5*kappa*(PAI-1)^2)*Y*R(+1)/((1-.5*kappa*(PAI(+1)-1)^2)*Y(+1)*exp(mu)*PAI(+1));

    1-eta+eta*(1-.5*kappa*(PAI-1)^2)*Y+betta*kappa*(1-.5*kappa*(PAI-1)^2)*(PAI(+1)-1)*
    PAI(+1)/(1-.5*kappa*(PAI(+1)-1)^2)-kappa*(PAI-1)*PAI;
```

```
(R(-1)/steady_state(R))^rhor*(PAI/steady_state(PAI))^((1-rhor)*psi)*exp(sigr*EPS_R) -
R/steady_state(R);

RR = R/PAI(+1);
```

13.4.2 Macro processor

With the use of the macro processor you can use conditional statements and loop statements in the model file. The main statements that can be used in the macro processing language are:

@#echo
: Displays some text on the command line during parsing.

@#error
: Throws an error during parsing.

@#include
: File inclusion. This command can be used without turning on the macro processor.

@#define
: Define a macro variable that can be used in the conditional and loop statements.

@#if
: Opening of a conditional block.

@#ifdef
: Test if a macro variable is defined.

@#ifndef
: Test if a macro variable is not defined.

@#else
: Else statement in a @#if, @#ifdef or @#ifndef block.

@#endif
: End of conditional block.

@#for
: Opening of a loop block.

@#endfor
: End of loop block.

\\
: Continue macro processing statement on the next line.

/*NB-IGNORE-BEGIN*/
: Start a multi-line comment ignored by the model parser. This command can be used without turning on the macro processor.

/*NB-IGNORE-END*/
: Start a multi-line comment ignored by the model parser. This command can be used without turning on the macro processor.

In the following the syntax and some examples will be provided.

13.4.2.1 @#echo

Syntax

```
@#echo string
```

Example

```
@#echo "Some information message"
```

13.4.2.2 @#error

Syntax

```
@#error string
```

Example

```
@#error "Some error message"
```

13.4.2.3 @#include

Syntax

```
@#include "filename.ext"
```

Example

```
@#include "modelPart.nb"
```

13.4.2.4 @#define

Syntax

```
@#define statement
```

Example

```
@#define macroVar = 1
```

13.4.2.5 @#if

Syntax

```
@#if statement
    included if statement is not equal to 0
@#endif

@#if statement
    included if statement is not equal to 0
@#else
    included if statement is equal to 0
@#endif
```

Example

```
@#if macroVar
   i = lambda_i*i(-1);
@#else
   i = lambda_i*i(-1) + e_i;
@endif
```

13.4.2.6 @#ifdef

Syntax

```
@#ifdef macroVar
    included if macroVar is defined
@#endif
```

```
@#ifdef macroVar
    included if macroVar is defined
@#else
    included if macroVar is not defined
@#endif
```

13.4.2.7 @#ifndef

Syntax

```
@#ifndef macroVar
    included if macroVar is not defined
@#endif
```

```
@#ifndef macroVar
    included if macroVar is not defined
@#else
    included if macroVar is defined
@#endif
```

13.4.2.8 @#for

Syntax

```
@#for str in statement
    loop body
@#endfor
```

Example

```
@#for c in ["h","f"]
    i_@{c} = lambda_@{c}_i*i_@{c}(-1);
@#endfor
```

13.4.2.9 Multi-line comment

Examples

```
/*NB-IGNORE-BEGIN*/

Here is the comment ignored by the parser

/*NB-IGNORE-END*/

/*NB-IGNORE-BEGIN*/ This is also ignored. /*NB-IGNORE-END*/
```

13.4.2.10 Macro expressions

The following operators are supported by the macro processor

Operators on integers

- Arithmetic operators; +, -, *, /
- Comparison operators; <, >, ≤, ≥, ==, ~= (or !=)
- Logical operators; &&, ||, ~(or !)
- Integer ranges: 1:4 is equivalent to integer array [1,2,3,4].

Operators on characters

- Concatenation; +
- Comparison operators; ==, ~= (or !=)
- Extraction of substrings; Let str be a macro variable storing a string (character array), then str[2] get the second character of that character array (word). You can also do str[2:4] to get the characters located at the elements from 2 to 4.

Operators on arrays

- Concatenation; +
- Extraction of an element of an array; Let a be an array, then a[2] returns the second element of that array. For array of string you can also use a2.
- Extraction of sub-arrays; Use a[2:4].
- Testing if an element is part of the array using the in operator; "b" in ["a","b","c"] or 1 in [1,2,3]. Return a logical true (1) or false (0).
- Remove the elements of one array that is found in first array. ["a","b","c"] - ["a"] = ["b","c"].

Macro expressions can be used after the @#if command, after the in command of the @#for loop, after the @#define statement and when enclosed like @{expression} anywhere in the model file. Be aware that the macro processor language cannot be used in the steady-state provided.

13.5 Calibration

Calibration of DSGE model is easily done, by providing value to the parameters of the model.

13.6 Solve for the steady-state

In this step of the solution we solve for the point we want to approximate our DSGE model around. It can be done analytically using a steady-state file or numerically. See the following sections.

13.6.1 Steady-state file

The steady-state file must be a MATLAB .m file. It must take two inputs and have two outputs. The first input (model) is a struct containing the fields endogenous, while the second input (pp) is a struct storing the parameters of the model. The field names are the parameter names of the model, while the fields stores the values. These two inputs are passed into the steady-state file automatically by the solver.

To use the parameters in the steady-state file either reference the pp input, or add the following code segment in the steady-state file to be able to use it as normal variable in MATLAB

```
pnames = fieldnames(pp);
parnr  = length(pnames);
for j = 1:parnr
    pname = deblank(pnames{j});
    eval([pname,' = pp.(pname);'])
end
```

i.e. after this code segment you can reference a parameter just by its name.

The first output (ss) of the function must be a struct. This struct must have the fieldnames as the names of the endogenous variables, and its fields set to the steady-state value of the corresponding endogenous variable. This can be done by

```
ss.varName = 0.2;
```

where varName must be substituted with the name of the endogenous variable. You do not need to assign the steady-state value to an endogenous variable with steady-state of 0, as this is the default. You can also assign the steady-state values by using

```
varName = 0.2;
```

when combined with the following code segment at the end of the .m file

```
mylist      = model.endogenous(~model.isAuxiliary);
my_endo_nbr = length(mylist);
for j = 1:my_endo_nbr
    ss.(mylist{j}) = eval(mylist{j});
end
```

Be aware that this last way forces you to declare the steady-state of all endogenous variables, even if they have steady-state value of 0.

Finally the second output (param) must be a struct storing the parameters that the value is solved for in the steady-state. Use the syntax

```
param.paramName = 0.2;
```

where paramName must be substituted with the name of the parameter you want to assign the value of.

```
function [ss,param] = model_steadystate(model,pp)

    param  = struct();
    ss.PAI = 1;
    ss.Y   = (pp.eta - 1)/pp.eta;
    ss.R   = exp(pp.mu)/pp.betta*ss.PAI;
    ss.RR  = ss.R/ss.PAI;

end
```

13.6.2 Solve it numerically

To solve the steady-state of a DSGE model is supported in NB toolbox (Paulsen, 2021). See the provided references for how to do this.

13.7 Solve model

See NB toolbox (Paulsen, 2021) for how to solve a DSGE model. The solution algorithm available where discussed in sections 13.1, 13.2 and 13.3.

13.8 Estimation

Bayesian estimation of DSGE models is supported. The approach used is to separate the problem of maximizing the posterior distribution and sampling from it.

13.8.1 Mode estimation

The posterior is given by

$$p(\theta|Y,\mathbb{M}) = \frac{\mathcal{L}(Y|\theta,\mathbb{M})\pi(\theta,\mathbb{M})}{m(Y,\mathbb{M})}, \tag{383}$$

where Y is the data on the observed variables, $\mathbb{M}$ indicate knowing the structure of the model, $\mathcal{L}(Y|\theta)$ is the likelihood, $\pi(\theta,\mathbb{M})$ is the prior and $m(Y,\mathbb{M})$ is the marginal density of the data, and is given by

$$m(Y,\mathbb{M}) = \int \mathcal{L}(Y|\theta,\mathbb{M})\pi(\theta,\mathbb{M})d\theta. \tag{384}$$

It is normal in the literature that the prior is specified a set of marginal independent priors, even though this is strictly not a theoretical requirement

$$p(\theta|\mathbb{M}) = p(\theta^1) \times \cdots \times p(\theta^P). \tag{385}$$

As $m(Y\mathbb{M})$ does not depend on the θ, maximizing $p(\theta|Y)$ amount to maximize the objective $\mathcal{L}(Y|\theta,\mathbb{M})\pi(\theta,\mathbb{M})$. The likelihood can be evaluated using the appropriate Kalman filter (see section 2.3) and the prior is just a

function of θ so it is easy to evaluate. To do the maximization of the objective a optimizer in MATLAB is used.

13.8.1.1 System priors We follow Andrle and Benes (2013) which makes it possible to estimate a DSGE model with system priors. These priors is used to apply priors to some properties of the model. These properties will be a function of the model and the parameters of the model, i.e. $\mathbb{Z} = H(\mathbb{M}, \theta)$. These properties will itself be a probabilistic model $\mathbb{Z} \sim h(\mathbb{S})$, where h is a distribution function with parameters $\mathbb{S}$. This will result in the system prior on the form $p(\mathbb{S}|\theta, \mathbb{M})$, and the joint prior is then given by

$$p(\theta|\mathbb{M}, \mathbb{S}) = p(\mathbb{S}|\theta, \mathbb{M}) \times p(\theta|\mathbb{M}). \tag{386}$$

A Bayesian approach constitute of estimating the parameters θ by maximization of the posterior distribution given by

$$p(\theta|Y, \mathbb{M}, \mathbb{S}) = \frac{\mathcal{L}(Y|\theta, \mathbb{M}) \times p(\theta|\mathbb{M}, \mathbb{S})}{m(Y|\mathbb{M})} \propto \mathcal{L}(Y|\theta, \mathbb{M}) \times p(\theta|\mathbb{M}, \mathbb{S}). \tag{387}$$

13.8.2 Sampling

To sample from the posterior an altered version of the random walk Metropolis-Hastings algorithm is used. It is most like the approach found in Haario et al. (2001), but in contrast to that paper we also add a scale parameter that is updated to target the wanted acceptance rate.

13.9 Filtering

See section 2 for a theoretical discussion on the filtering steps.

13.10 DSGE model with break-point

See Appendix A of Kravik and Paulsen (2020) for how break-points are handled in DSGE models.

14 Analysis

When a model is estimated and solved[33] it is possible to do some further analysis on the model. For some models it is possible to produce impulse response function, shock decomposition and forecast error variance decomposition. Each method will be described in turn in the next sections.

14.1 Impulse response functions

Let's start with the companion form of the model

$$Y_t = AY_{t-1} + BX_t + C\varepsilon_t \tag{388}$$

$$E\varepsilon_t'\varepsilon_t = I. \tag{389}$$

[33] Some models must also be identified.

Y_t has size M x 1, X_t has size N x 1, ε_t has size Q x 1, A has size M x M, B has size M x N and C has size M x Q.[34]

In a linear model an impulse response function is a response in an endogenous variable to a one-off disturbance to one of the structural shocks, or in mathematical terms

$$E\left[\frac{\partial Y_{i,t+s}}{\partial \varepsilon_{j,t}}\right] \text{, for all } i,j \text{ and } s \in [0, S-1], \tag{390}$$

where S is the total number of steps ahead, $i \in [1, M]$ is the dependent variable index and $j \in [1, Q]$ is the index of the structural shock. With the use of matrix notation and equation 388 we can express it as

$$E\left[\frac{\partial Y_{t+s}}{\partial \varepsilon_{j,t}}\right] = A^s C e_j. \tag{391}$$

Again for all j and $s \in [0, S-1]$, and where e_j is the a vector of zeros except at the element j where it is 1.

Density IRFs are produced by drawing the parameters from a known distribution. In classical estimation this can be done by using a bootstrap method, see section 1.7, while for bayesian estimation it amounts to draw from the posterior distribution. For each parameter draw the model is solved, i.e. written on the companion form[35], and the formula from equation 391 is used to produce a draw from the distribution of the IRFs. The error band is then produced by the wanted percentiles of this distribution for all s and j.

14.2 Shock decomposition

Let's start with the companion form of the model

$$Y_t = AY_{t-1} + BX_t + C\varepsilon_t. \tag{392}$$

Y_t has size M x 1, X_t has size N x 1, u_t has size Q x 1, A has size M x M, B has size M x N and C has size M x Q.[36]

As we know the values of ε_t and X_t at each t we can decompose the difference between the unconditional forecast given by AY_{t-1} and the actual observation over the history by

$$Y_t - AY_{t-1} = BX_t + C\varepsilon_t. \tag{393}$$

At period $t = 1$ we don't know Y_{t-1}, but we can find the contribution of all the shocks hitting before this date by

$$IC_1 = Y_1 - BX_1 - C\varepsilon_1. \tag{394}$$

IC_1 is the initial conditions, which can at the start of the decomposition be quite big, but should die out after a while. By using 393 the contribution of the shock and the exogenous variables at time s can be found by iterating on

$$K_t = AK_{t-1} \tag{395}$$

$$K_t = K_t + \begin{bmatrix} B & C \end{bmatrix} * \begin{bmatrix} \tilde{X}_t & \tilde{\varepsilon}_t \end{bmatrix}. \tag{396}$$

[34] M need not be equal to Q, as auxiliary variables (equation) can be added to make it into a dynamic system only dependent on one lag, and these equations do not have any error terms.

[35] For the models that need to be identified this is also done at this step.

[36] M need not be equal to Q, as auxiliary variables (equation) can be added to make it into a dynamic system only dependent on one lag, and these equations does not have any error terms.

From $t = 1$ until $t = s$. Where Q_t has size M x N + Q, $K_0 = 0$, $*$ is the element-wise multiplication operator, $\tilde{X}_t = \begin{bmatrix} X_t' \\ \vdots \\ X_t' \end{bmatrix}$ and $\tilde{\varepsilon}_t = \begin{bmatrix} \varepsilon_t' \\ \vdots \\ \varepsilon_t' \end{bmatrix}$. I.e. $\tilde{X}_t$ and $\tilde{\varepsilon}_t$ is constructed by stacking X_t' and ε_t' M times vertically. Then the contribution from the initial conditions are given by

$$IC_t = Y_t - sum(K_t, 2). \tag{397}$$

14.3 Forecast error variance decomposition (FEVD)

Let's start with the companion form of the model

$$Y_t = AY_{t-1} + BX_t + C\varepsilon_t. \tag{398}$$

Y_t has size M x 1, X_t has size N x 1, ε_t has size Q x 1, A has size M x M, B has size M x N and C has size M x Q.[37]

Then the total mean squared forecast error at time $t + 1$ is given by

$$MSE(Y_{t+1}) = \Omega. \tag{399}$$

While the total mean squared forecast error at time $t + h$, for $h > 2$, is given by

$$MSE(Y_{t+h}) = MSE(Y_{t+h-1}) + \rho_h \Omega \rho_h, \tag{400}$$

where

$$\Omega = \tilde{C}\tilde{C}' \tag{401}$$

$$\tilde{C} = C(1 : Q, 1 : Q) \tag{402}$$

$$\rho_h = A^{h-1}(1 : Q, 1 : Q) \text{ for } h > 2. \tag{403}$$

While the total mean squared forecast error at the infinite horizon is given by

$$MSE(Y_{inf}) = \rho_{inf} \Omega \rho_{inf}, \tag{404}$$

where

$$\rho_{inf} = inv(I_{kk} - \beta) \tag{405}$$

$$\beta = \sum_{p=1}^{nLags} \beta_p, \tag{406}$$

and β_p is the estimated coefficients in front of each lagged value of Y_t in the estimated equation.

The mean squared forecast error contributed by shock j at time $t + 1$ is then

$$MSE(Y_{t+1})_j = \Omega_j. \tag{407}$$

[37] M need not be equal to Q, as auxiliary variables (equation) can be added to make it into a dynamic system only dependent on one lag, and these equations does not have any error terms.

While the total mean squared forecast error at time $t+h$, for $h > 2$, is given by

$$MSE(Y_{t+h})_j = MSE(Y_{t+h-1})_j + \rho_h \Omega_j \rho_h, \tag{408}$$

where

$$\Omega_j = \tilde{C}_j \tilde{C}_j' \tag{409}$$

$$\tilde{C}_j = C(1:Q, j). \tag{410}$$

While the mean squared forecast error contributed by shock j at the infinite horizon is

$$MSE(Y_{inf})_j = \rho_{inf} \Omega_j \rho_{inf}. \tag{411}$$

The variance of the forecast error contributed by shock j at period h is then given by

$$FEVD_{j,h} = diag(MSE(Y_{t+h})_j ./ MSE(Y_{t+h})), \tag{412}$$

where ./ is the element-wise division operator, and the function *diag* picks out the diagonal element of a matrix. $FEVD_{j,h}$ will then have size M x 1, i.e. the shock j's contribution to the forecast error variance of all the endogenous variables.

14.4 Theoretical moments

We start from equation 55, where we assume that the process is stationary. We can find the mean of this process as

$$\mathbb{E}[X] = c + A\mathbb{E}[X] + B\mathbb{E}[u]. \tag{413}$$

By solving this for $\mathbb{E}[X]$ we get

$$\begin{aligned} \mathbb{E}[X] &= (I - A)^{-1}(c + B\mathbb{E}[u]) \\ &= (I - A)^{-1} c. \end{aligned} \tag{414}$$

The contemporaneous covariance can equally be found from solving the equation

$$\Omega_0 = A\Omega_0 A' + BQB', \tag{415}$$

where we have used that $u_t \sim N(0, Q)$. This is an equation on the form of a lyapunov equation, and it has no close form solution. In NB toolbox this equation is solved with and fixed point iteration algorithm. Remember here that X is a vector of size $Nx1$, so Ω is the covariance matrix with size NxN. To find the contemporaneous correlation matrix (Σ) we must use the following normalization

$$\Sigma_0 = \Omega_0 \oplus \omega_0, \tag{416}$$

where $\oplus$ is the element wise division operator and

$$\omega_0 = \sqrt{diag(\Omega_0)} \sqrt{diag(\Omega_0)}'. \tag{417}$$

To get the autocovariance matrix at lag k we can use the formula

$$\Omega_k = A^k \Omega_0. \tag{418}$$

And equivalently for the autocorrelation matrix at lag k

$$\Sigma_k = A^k \Sigma_0. \tag{419}$$

15 Forecasting

15.1 Unconditional forecast

Let us first start with some theory. A model on the companion for can be forecasted as

$$Y_{t+h} = AY_{t-1+h} + BX_{t+h} + u_{t+h} \tag{420}$$

$$u_{t+h} \sim N(0, \Omega). \tag{421}$$

Y_t has size M x 1, X_t has size N x 1, u_t has size Q x 1, A has size M x M, B has size M x N, t is the end date of history and h is the number of steps ahead forecast.

Point forecast is then produced by setting $u_{t+h} = 0$, for all h, and conditional on lagged values of the endogenous variables (Y_{t-1+h}) and some values of the exogenous variables (X_{t+h}). H step forecast is then produce by iterating on the equation 420 for $h \in [1 : H]$. If the variables in X_{t+h} only contain deterministic variables, as constant, time-trend and seasonal dummies, no further condition information is needed, and we call the forecast produced in this way unconditional forecast.

Density forecast is produce by drawing the parameters from a know distribution. In frequentist estimation this can be done by using a bootstrap method, see section 1.7, while for Bayesian estimation it amounts to draw from the posterior distribution. For each parameter draw the model is solved, i.e. written on the companion form[38], and then the residuals, u_{t+h}, is drawn from the multivariate distribution given by equation 421 as many times as wanted. A H step ahead density forecast is then produce by iterating this process and using the equation 420 for $h \in [1 : H]$.

15.2 Conditional forecast

When doing conditional forecast the model must be written in a slightly different way, as it is possible to translate the conditional information into restrictions on the structural shocks

$$Y_t = AY_{t-1} + BX_t + C\varepsilon_t \tag{422}$$

$$E\varepsilon_t'\varepsilon_t = I. \tag{423}$$

Y_t has size M x 1, X_t has size N x 1, ε_t has size Q x 1, A has size M x M, B has size M x N and C has size M x Q.[39]

If the model contains non-deterministic exogenous variables no forecast can be produced unless some conditional information is added. This information can come from forecast of other models or be subjective forecast. This is to condition on exogenous variables.

[38] There is no reason to identify a model when doing unconditional forecast, so this step is discarded in the algorithm.

[39] M need not be equal to Q, as auxiliary variables (equation) can be added to make it into a dynamic system only dependent on one lag, and these equations does not have any error terms.

To produce forecast conditional on information about some endogenous variables you need to identify the shocks to match those restrictions. Lets make a simple example, which we also used in section 8.8

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \end{bmatrix} = A \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \end{bmatrix} + \begin{bmatrix} C_{1,1} & 0 & 0 \\ C_{2,1} & C_{2,2} & 0 \\ C_{3,1} & C_{3,2} & C_{3,3} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}. \tag{424}$$

And say we want to condition on the variable $Y_{3,t}$. Then you can back out the difference between the unconditional forecast and the conditional information, which has to be matched by some shocks by

$$\begin{bmatrix} C_{3,1} & C_{3,2} & C_{3,3} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix} = Y_{3,t} - \begin{bmatrix} A_{3,1} & A_{3,2} & A_{3,3} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \end{bmatrix}, \tag{425}$$

so in this case all three shocks can be used to explain the conditional information. You need to select which shocks to use to match the added conditional information, and as you see from equation 424, to select different shocks will give different forecast for all the other endogenous variables of the model. If you condition on less endogenous variables than the number of selected shocks, the total variance of the shocks that should match those restriction is minimized. Had we instead conditioned on $Y_{1,t}$, we see from equation 424 that the only shock that could have match this information is the $\varepsilon_{1,t}$ shock. I.e. the identification is unique. If you had tried another shock in this example, an error would have occurred.

For conditional forecasting with forward looking models we have made the algorithms of Maih (2010) and Paulsen (2017) available in NB toolbox. (Paulsen (2021)).

15.3 Conditional forecast using a Kalman filter

For models that has a measurement equation, the algorithm of the last section does not work, and we must instead use the Kalman-filter introduced in section 2. You need to set an option in the toolbox, if you want to use the Kalman-filter instead of the algorithm of the last section.

Caution: DSGE models must use the algorithm of the last section!

15.4 Forecast evaluation

Sometimes you may be interested in the forecasting properties of your model. In general you have two options. Do in-sample forecast, which means to estimate your model over the full sample, and then produce forecast for each period of the sample, or a sub-sample. From there you evaluate the forecast by comparing it to actual observations by a given metric. The second option is to do out-of-sample forecast, which means to estimate your model recursively, and then use the estimated parameters for the sub-sample $1:s$ to produce forecast at time s. Again you compare the forecast against the actual data using a given metric.

15.4.1 Point forecast

Let us define the forecast at time s at horizon h by $f_{s,h}$, and the actual data at time s as Y_s. Both $f_{s,h}$ and Y_s has size 1 x M, where M is the number of forecasted variables. The forecast error at time s at horizon h is then given by $\eta_{s,h} = f_{s,h} - Y_{s+h}$. To evaluate the forecast a metric must be chosen:

Mean squared error (MSE)

$$score_h = \frac{1}{T-t+1} \sum_{s=t}^{T} \eta_{s,h}^2$$

Root mean squared error (RMSE)

$$score_h = \sqrt{\frac{1}{T-t+1} \sum_{s=t}^{T} \eta_{s,h}^2}$$

Mean absolute error (MAE)

$score_h = \frac{1}{T-t+1} \sum_{s=t}^{T} |\eta_{s,h}|$

Standard dev. of error (STD)

$score_h = \sqrt{\frac{1}{T-t} \sum_{s=t}^{T} \eta_{s,h}^2}$

Mean error (ME)

$score_h = \frac{1}{T-t+1} \sum_{s=t}^{T} \eta_{s,h}$

Where t is the start date and T is the end date of the recursive forecast.

15.4.2 Density forecast

Let us define the density forecast at time s at horizon h for variable v by $\varphi_{s,h,v}$, and the actual data at time s for variable v as $Y_{s,v}$. $\varphi_{s,h,v}$ is now a distribution, while $Y_{s,v}$ has size 1 x 1. $v \in [1, M]$. The forecast error at time s at horizon h for variable v is then given by $\eta_{s,h,v} = mean(\varphi_{s,h,v}) - Y_{s+h,v}$. To evaluate the forecast a metric must be chosen:

Mean squared error (MSE)

$score_{h,v} = \frac{1}{T-t+1} \sum_{s=t}^{T} \eta_{s,h}^2$

Root mean squared error (RMSE)

$score_{h,v} = \sqrt{\frac{1}{T-t+1} \sum_{s=t}^{T} \eta_{s,h}^2}$

Mean absolute error (MAE)

$score_{h,v} = \frac{1}{T-t+1} \sum_{s=t}^{T} |\eta_{s,h}|$

Standard dev. of error (STD)

$score_{h,v} = \sqrt{\frac{1}{T-t} \sum_{s=t}^{T} \eta_{s,h}^2}$

Mean error (ME)

$score_{h,v} = \frac{1}{T-t+1} \sum_{s=t}^{T} \eta_{s,h}$

Exponential of the sum of the log scores (ESLS)

$score_{h,v} = exp(\sum_{s=t}^{T} log(\varphi_{s,h,v}(Y_{s+h,v})))$

Exponential of the mean of the log scores (EELS)

$score_{h,v} = exp(\frac{\sum_{s=t}^{T} log(\varphi_{s,h,v}(Y_{s+h,v}))}{T-t+1})$

Mean of the log scores (MLS)

$score_{h,v} = \frac{\sum_{s=t}^{T} log(\varphi_{s,h,v}(Y_{s+h,v}))}{T-t+1}$ (Cannot be used for combinations)

Where t is the start date and T is the end date of the recursive forecast.

15.4.3 Mincer-Zarnowitz test

The Mincer-Zarnowitz approach can be used to test for bias in the forecast from one model. If the forecast is unbiased the forecast error is not predictable with information up until time t

$$\varepsilon_{t+h|t} = \beta_0 + \theta y_{t+h|t} + u_t. \tag{426}$$

$y_{t+h|t}$ is the forecast from the model at time $t+h$ with information up until time t, while $\varepsilon_{t+h|t}$ is the forecast error defined by

$$\varepsilon_{t+h|t} = y_{t+h} - y_{t+h|t}. \tag{427}$$

y_{t+h} is the actual observation at time $t+h$. By substituted this into equation 426 we get the regression the test is based on

$$y_{t+h} = \beta_0 + \beta_1 y_{t+h|t} + u_t. \tag{428}$$

Then a F-test is used to test if $\beta_0 = 0$ and $\beta_1 = 1$. The F-test is distributed as F(2,T-2), where T is the number of forecasting periods. The null hypothesis is that the forecast is unbiased.

15.4.4 Diebold-Mariano test

The Diebold-Mariano test can be used to test for equal predictive accuracy of two models. Let us define the forecast error from the two models as

$$\varepsilon^1_{t+h|t} = y_{t+h} - y^1_{t+h|t} \tag{429}$$

$$\varepsilon^2_{t+h|t} = y_{t+h} - y^2_{t+h|t}. \tag{430}$$

y_{t+h} is the actual observation at time $t+h$ and $y^i_{t+h|t}$ is the forecast from model i at time $t+h$ with information up until time t. Because the forecast errors use overlapping data (information) the forecast errors are most likely autocorrelated. We are interested in testing the null hypothesis

$$H_0 : E[L(\varepsilon^1_{t+h|t})] = E[L(\varepsilon^2_{t+h|t})] \tag{431}$$

against the alternative hypothesis

$$H_1 : E[L(\varepsilon^1_{t+h|t})] \neq E[L(\varepsilon^2_{t+h|t})], \tag{432}$$

where $L(x) = x^2$ in NB toolbox Paulsen (2021). I.e. the square error. Let us define

$$d_{t,h} = L(\varepsilon^1_{t+h|t}) - L(\varepsilon^2_{t+h|t}). \tag{433}$$

Then the test statistic is

$$dmTest_h = \left(\frac{\sum_{t=1}^{T-h} d_{t,h}}{T-h} \left(\frac{T-h}{\lambda_d} \right)^{\frac{1}{2}} \right)^2, \tag{434}$$

where

$$\lambda_d = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \text{ and } \gamma_j = cov(d_t, d_{t-j}). \tag{435}$$

This means that λ_d is the long run variance. As the sample is finite we need a way to estimate it. See section 1.5 for how this is done. It uses the Newey-West bandwidth selection algorithm. The test is distributed as χ^2 with 1 degree of freedom.

15.4.5 PIT

PIT (probability integral transform) uses the fact that when a random variable X has a CDF given by F_X then the random variable given by $Z = F_X(X)$ has a uniform distribution.

We can use this property for our density forecast. Let us assume that we have recursive density forecast given by: φ_s, for periods $s \in [1:S]$. Then we can for each period s use the realized value Y_s to evaluate the density , i.e. $Z_s = \varphi_s(Y_s)$. If the distributions φ_s is properly estimated the distribution of Z_s should be uniform. We can illustrate this by a histogram or we can test it by using a Pearson χ^2 test.

If we group each observation of Z_s in intervals with length 0.1, so we have 10 subintervals. Let abbreviate the number of observation falling into each interval as O_i and E_i the expected number of observation that should fall into each interval, where $i \in [1, 10]$. In our case we have $E_i = S * 0.1$. The test statistic is then given by

$$T = \sum_{=1}^{10} \frac{(O_i - E_i)^2}{E_i}, \tag{436}$$

which follows asymptotically a χ^2 distribution. The null hypothesis of the test is that the Z_s follows a uniform distribution.

15.5 Forecast combination

You can combine both point forecast and density forecast, using the different scores discussed in section 15.4, or assigning equal weights to all models. For point forecast the combined forecast is just a weighted mean of the combined models point (or mean) forecast, i.e.

$$CF_{h,v,t} = \sum_{m=1}^{M} w_{h,v,t,m} Y_{h,v,t,m}, \tag{437}$$

where $w_{h,v,m,t}$ is the weights, which are calculated using recursive forecast scores up until the given date, and is constructed as follows

$$w_{h,v,s,m} = \frac{\frac{1}{score_{h,v,s,m}}}{\sum_{m=1}^{M} \frac{1}{score_{h,v,s,m}}}. \tag{438}$$

$CF_{h,v,t}$ is the combined point forecast and $Y_{h,v,t,m}$ is the individual model point (mean) forecast. The indexing is as follows; h is the horizon of the forecast, i.e. we see that the weights may differ over the forecast horizon, v is the selected variable, t is the date of the forecast, and m is the model index.

For density forecast it is only possible to weight forecast by using a linear opinion pool

$$\varphi^{CF}_{h,v,t} = \sum_{m=1}^{M} w_{h,v,t,m} \varphi_{h,v,t,m}, \tag{439}$$

where the weights are constructed as

$$w_{h,v,s,m} = \frac{score_{h,v,s,m}}{\sum_{m=1}^{M} score_{h,v,s,m}}. \tag{440}$$

For EELS and ESLS, otherwise as in 438. $\varphi^{CF}_{h,v,t}$ is the combined density forecast, and $\varphi_{h,v,t,m}$ is the individual model density forecast.

15.6 Aggregate forecast

See Paulsen *et al.* (2021).

16 Model selection

16.1 VAR

In NB toolbox (Paulsen, 2021) it is possible to select the M best VAR models out of a big suit of VAR models. To create the suit of VAR models to select from you need to select a variable of interest and a set of other variables you think are good at predicting the variable of interest. These other variables should be selected by economic reasoning. Let the number of other variables be abbreviated by V. Let $minV$ be the minimum number of other variables and $maxV$ be the maximum number of other variables to be used in the suits of VAR models. I.e. the suits of VAR models are created by adding $minV$ to $maxV$ other variables (without replication) to the variable of interest. For each VAR with given variables we test for lags between 1 to $nLags$, where $nLags$ is the maximum number of lags to be tested for. I.e. we create a suit of VAR models with size $R = nLags \sum_{K=minV}^{maxV} \frac{V!}{K!(V-K)!}$.

16.2 ARIMA

In NB toolbox (Paulsen, 2021) it is also possible to select the M best ARIMA models out of a big suit of ARIMA models. To create the suit of ARIMA models to select from you need to select a variable of interest. Then you can choose what exogenous variables you would like to add by economic reasoning. These will be forecasted using separate $AR(p)$ processes. See below for more on the different options when dealing with ARIMA model selection.[40]

16.3 General

The models are then estimated and forecasted. To select the best models an evaluation criterion is used, for more on the available criteria see section 15.4.

[40] The level of integration is fixed to be 0, so the series must be stationary.

References

Andrle, M. and Benes, J. (2013) System Priors; Formulating Priors about DSGE Models' Properties, IMF Working Papers 13/257, International Monetary Fund.

Bai, J. and Ng, S. (2000) Determining the Number of Factors in Approximate Factor Models, Boston College Working Papers in Economics 440, Boston College Department of Economics.

Banbura, M., Giannone, D. and Reichlin, L. (2010) Nowcasting, CEPR Discussion Papers 7883, C.E.P.R. Discussion Papers.

Banbura, M. and Modugno, M. (2010) Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data, Working Paper Series 1189, European Central Bank.

Binning, A. (2013) Underidentified SVAR models: A framework for combining short and long-run restrictions with sign-restrictions, Working Paper 2013/14, Norges Bank.

Blanchard, O. J. and Kahn, C. M. (1980) The Solution of Linear Difference Models under Rational Expectations, *Econometrica*, **48**, 1305–1311.

Boivin, J., Giannoni, M. P. and Mihov, I. (2009) Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data, *American Economic Review*, **99**, 350–384.

Davidson, R. and Monticini, A. (2014) Heteroskedasticity-and-Autocorrelation-Consistent Bootstrapping, DISCE - Working Papers del Dipartimento di Economia e Finanza def012, UniversitÃ Cattolica del Sacro Cuore, Dipartimenti e Istituti di Scienze Economiche (DISCE).

Debortoli, D., Maih, J. and Nunes, R. (2010) Loose commitment in medium-scale macroeconomic models: Theory and an application, Working Paper 2010/25, Norges Bank.

Eraslan, S. and Schröder, M. (2022) Nowcasting gdp with a pool of factor models and a fast estimation algorithm, *International Journal of Forecasting*.

Giannone, D., Lenza, M. and Primiceri, G. E. (2012) Prior Selection for Vector Autoregressions, Working Papers ECARES ECARES 2012-002, ULB – Universite Libre de Bruxelles.

Giannone, D., Lenza, M. and Primiceri, G. E. (2016) Priors for the Long Run, CEPR Discussion Papers 11261, C.E.P.R. Discussion Papers.

Hamilton, J. and Press, P. U. (1994) *Time Series Analysis*, Princeton University Press.

Hounyo, U. (2014) The wild tapered block bootstrap, CREATES Research Papers 2014-32, Department of Economics and Business Economics, Aarhus University.

Klein, P. (2000) Using the generalized Schur form to solve a multivariate linear rational expectations model, *Journal of Economic Dynamics and Control*, **24**, 1405–1423.

Koop, G. and Korobilis, D. (2009) Bayesian Multivariate Time Series Methods for Empirical Macroeconomics, Working Paper series 47, Rimini Centre for Economic Analysis.

Kravik, E. M. and Paulsen, K. S. (2020) Handling structural break points in nemo, Working Paper 2020/01, Norges Bank.

Lenza, M. and Primiceri, G. E. (2020) How to estimate a var after march 2020, Working Paper 27771, National Bureau of Economic Research.

Litterman, R. B., Sims, C. A. and Doan, T. (1986) Forecasting and conditional projection using realistic prior distribution, Tech. rep.

MacKinnon, J. G. (1996) Numerical Distribution Functions for Unit Root and Cointegration Tests, *Journal of Applied Econometrics*, **11**, 601–618.

MacKinnon, J. G., Haug, A. A. and Michelis, L. (1999) Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration, *Journal of Applied Econometrics*, **14**, 563–577.

Maih, J. (2010) Conditional forecasts in DSGE models, Working Paper 2010/07, Norges Bank.

Negro, M. D. and Schorfheide, F. (2004) Priors from general equilibrium models for vars, *International Economic Review*, **45**, 643–673.

Osborne, M., Presnell, B. and Turlach, B. (2000) On the lasso and its dual, *Journal of Computational and Graphical Statistics*, **9**, 319–337.

Park, T. and Casella, G. (2008) The bayesian lasso, *Journal of the American Statistical Association*, **103**, 681–686.

Paulsen, K. S. (2017) Conditional forecasting with DSGE models - A conditional copula approach, Working Paper 2017/4, Norges Bank.

Paulsen, K. S. (2021) *NB toolbox: Reference Manual*.

Paulsen, K. S., Fastbø, T. and Ingebrigsten, T. (2021) Aggregate density forecast of models using disaggregated data - a copula approach, Working paper, Norges Bank.

Sims, C. A. and Zha, T. (1996) Bayesian methods for dynamic multivariate models, Tech. rep.

Stock, J. and Watson, M. (2002) Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics*, **20**, 147–62.

Tambalotti, A. (2004) Optimal monetary policy and productivity growth, Money Macro and Finance (MMF) Research Group Conference 2003 99, Money Macro and Finance Research Group.

Villemot, S. (2011) Solving rational expectations models at first order: what Dynare does, Dynare Working Papers 2, CEPREMAP.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, **48**, 817–838.

Wu, D.-M. (1973) Alternative tests of independence between stochastic regressors and disturbances, *Econometrica*, **41**, 733–50.