The Kalman Filter

Benedikt Kolb*

April 7, 2015

[updated August 18, 2015]

Abstract

This manuscript gives an outline of the Kalman filter (KF) algorithm including derivations of the steps involved. It is based on Canova (2007), Jesús Fernández-Villaverde's slides "State Space Models and Filtering" (JFV in the following) and Hamilton (1994). Two examples illustrate the use of the KF: Maximising the likelihood of a DSGE model with a financial accelerator with respect to some parameters and estimating time-varying coefficients for a Taylor rule with US data from 1982Q1 to 2007Q2.

1 Purpose and properties of the Kalman filter

1.1 The Kalman filter in macroeconomics

For a given model structure, the Kalman filter (KF) generates one-period ahead forecasts of observables y_t, where we assume that these observables are driven by some unobservable states x_t. An example from economics could be a stochastic simple growth model with y_t being output and consumption and x_t being the capital stock and technology. The KF has two main advantages: It optimally estimates the unobservable states in a recursive procedure and furthermore generates exact finite-sample forecasts and the exact likelihood function for Gaussian ARMA processes.
The centerpiece of the KF is a model in state-space form, where observables y_t depend on states x_t. The state-space model is described by two equations: A transition (or state) equation, (1), and a measurement (or observation) equation, (2),

$$x_t = Fx_{t-1} + Gw_t \tag{1}$$
$$y_t = A + Bz_t + H'x_t + Cv_t \tag{2}$$

where $w_t \sim N(0, Q)$ and $v_t \sim N(0, R)$ are independent Gaussian martingale sequences and z_t is a vector of exogenous or predetermined variables. In the following, I will abstract from including

*This note is available on my homepage (http://www.bkolb.eu/codes/kalman.zip). Comments are welcome! Please send them to benedikt@bkolb.eu.

exogenous variables, so B is a zero matrix.[1] Moreover, I will assume $A = 0$ for simplicity and $C = I_n$, where I_n is the identity matrix of appropriate dimension. This leaves us with the simplified state-space model

$$x_t = Fx_{t-1} + Gw_t \tag{3}$$
$$y_t = H'x_t + v_t \tag{4}$$

Stability of the system requires that the eigenvalues of F are < 1.

1.2 Some caveats

1. Beware of the central assumptions:
   (a) Initial conditions of and innovations to ARMA processes are Gaussian.
   (b) The system is stable (all eigenvalues of F are < 1).
2. If initial conditions or innovations are not normal, the KF produces only the best linear forecast for the data y_t. In this case, non-linear filters might do a better job in predicting y_t, see Canova (2007), p. 219. For example, see Fernández-Villaverde and Rubio-Ramírez (2007) for an outline of the particle filter.
3. For unstable systems (some of the eigenvalues of F are ≥ 1), you can use the information filter by Anderson and Moore (1979) or the non-stationary KF by Koopman (1997), see Canova, p. 216.
4. The state space representation is not unique, see JFV, slide 4
5. You can write many different processes in state-space form by adjusting them appropriately, see JFV slides 5-12 and Canova (2007), Exercise 6.1

2 Algorithm

2.1 The KF algorithm

This subsection just presents the algorithm. For derivations, please refer to Section 3.
Start with initial values for states $x_{t|t-1}$ and their forecast error matrix $\Sigma_{t|t-1}$ – how to choose these values is outlined in the next subsection. We first get the forecast error v_t as

$$\begin{aligned} v_t &= y_t - y_{t|t-1} \\ &= y_t - H'x_{t|t-1} \end{aligned} \tag{5}$$

The mean squared – one-step-ahead – forecast error (MSFE) of the observables is given by

1. See Hamilton (1994) for derivations including a non-zero B matrix.

$$\Omega_{t|t-1} \equiv E\left((y - y_{t|t-1})(y_t - y_{t|t-1})'|y^{t-1}\right) \tag{6}$$
$$= H'\Sigma_{t|t-1}H + R$$

where y^{t-1} denotes the history of observables up to $t-1$. Remember that R is the variance-covariance matrix of the measurement equation (2), i.e. it represents the size of the measurement error.

At this step, you can also back out the (log) likelihood of the model, $\log l(y|y^{t-1}, F, G, H, Q, R)$, which will be summed up to the total log likelihood:

$$\begin{aligned} L_t &\equiv \log l(y^T|F, G, H, Q, R) \\ &= \sum_{t=1}^{T} \log l(y_t|y^{t-1}, F, G, H, Q, R) \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left[n \cdot \log(2\pi) + \log\left(\det(\Omega_{t|t-1})\right) + v_t'\Omega_{t|t-1}^{-1}v_t\right] \end{aligned} \tag{7}$$

where n is the number of states to be estimated.[2]

The Kalman gain is[3]

$$K_t = \Sigma_{t|t-1}H\Omega_{t|t-1}^{-1} \tag{8}$$

Why is this called a "gain function"? Intuitively, if we did a bad job forecasting $x_{t|t-1}$ ($\Sigma_{t|t-1}$ large), we will give a heavy weight to the new information (K_t large). If the new information is mostly noise (R large and hence $\Omega_{t|t-1}^{-1}$ small), we give heavy weight to the old prediction (K_t small).[4]

The MSFE of the *states* given history y^{t-1} is given by

$$\begin{aligned} \Sigma_{t+1|t} &\equiv E\left((x_{t+1} - x_{t+1|t})(x_{t+1} - x_{t+1|t})'|y^t\right) \\ &= F\Sigma_{t|t-1}F' + GQG' - FK_t\Omega_{t|t-1}K_t'F' \end{aligned} \tag{9}$$

The nowcast for the state x_t is formed by including the new information from y_t:

$$x_{t|t} = x_{t|t-1} + K_t v_t$$

where v_t is the forecast error. As the state forecast at this point, $x_{t+1|t}$, is simply $Fx_{t|t}$, we get

$$x_{t+1|t} = Fx_{t|t} = Fx_{t|t-1} + FK_t v_t \tag{10}$$

The next iteration then starts with the values $\Sigma_{t+1|t}$ and $x_{t+1|t}$.

A short algorithm sufficient to extract the log likelihood L_t of the model in state-space form is thus given by (5) to (10).

Note that the equations for $x_{t|t-1}$ and $\Omega_{t|t-1}$ (10 and 6) are referred to as "prediction equations" and the ones for $x_{t|t}$ and $\Sigma_{t|t}$ (see Section 3) are called "updating equations".

2. In the code, the log likelihood is added to the previous value in every iteration, starting at zero. So after the loop, we rescale by 0.5 and add the constant term $Tn\log(2\pi)$.

3. Canova (2007) calls $\tilde{K}_t \equiv FK_t$ the Kalman gain (see p. 216), while I follow JFV in calling K_t the Kalman gain.

4. See also JFV, slide 30.

2.2 Initial values

How to obtain initial values $x_{t|t-1}$ and $\Sigma_{t|t-1}$? For a stable system (eigenvalues of F are < 1), we usually use the unconditional mean as an initial value for the unknown states, $x_{1|0} = E(x)$, which follows from the distributional assumption. Similarly, the mean squared forecast error is initialised by the unconditional variance of the process: $\Omega_{1|0} = F\Omega_{1|0}F' + GQG'$, so that $\text{vec}(\Omega_{1|0}) = [I_{n^2} - (F \otimes F')]^{-1}\text{vec}(GQG')$, where I_z is the identity matrix of dimension z and n is the number of states (state variables and exogenous shock processes in the model) to be estimated.[5]

3 Derivations

In the following, I derive the equations of the KF (equations 6, 8 and 9). I make use of

$$
\begin{aligned}
y_t &= H'x_t + v_t \\
y_{t|t-1} &= H'x_{t|t-1} \\
E_t\left(v_t(x_t - x_{t|t-1})'|y^{t-1}\right) &= 0 \\
E_t\left(w_{t+1}(x_t - x_{t|t})'|y^{t}\right) &= 0
\end{aligned}
$$

As in Section 2, every iteration takes $x_{t|t-1}$ and $\Sigma_{t|t-1}$ as given.

$\boldsymbol{\Omega_{t|t-1}}$. First, the mean squared forecast error of observables and states given the history of observables y^{t-1}, $\Omega_{t|t-1}$, can be derived as

$$
\begin{aligned}
\Omega_{t|t-1} &= E\left((y_t - y_{t|t-1})(y_t - y_{t|t-1})'|y^{t-1}\right) \\
&= E\left((H'x_t + v_t - H'x_{t|t-1})(x_t'H + v_t' - x_{t|t-1}'H)|y^{t-1}\right) \\
&= E\left(H'(x_t - x_{t|t-1})(x_t - x_{t|t-1})'H|y^{t-1}\right) + E\left(v_t(x_t - x_{t|t-1})'H|y^{t-1}\right) \\
&\quad + E\left(H'(x_t - x_{t|t-1})v_t'|y^{t-1}\right) + E\left(v_tv_t'|y^{t-1}\right) \\
&= H'\Sigma_{t|t-1}H + 0 + 0 + R \\
&= H'\Sigma_{t|t-1}H + R
\end{aligned} \tag{6}
$$

$\mathbf{K_t}$. Second, the Kalman gain K_t is the coefficient matrix in

$$
K_t = \arg\min_k (y_t - y_{t|t-1}|y^{t-1})
$$

5. An alternative would be to use solve the Lyapunov equation, using e.g. Dynare's routine `lyapunov_symm`. I have added it to my code (commented out as default). The result is effectively the same, but `lyapunov_symm.m` is slightly slower.

$$K_t = E\left((y_t - y_{t|t-1})(x_t - x_{t|t-1})'|y^{t-1}\right)' \times \left[E\left((y_t - y_{t|t-1})(y_t - y_{t|t-1})'|y^{t-1}\right)\right]^{-1}$$
$$= \left[H'\Sigma_{t|t-1}\right]' \times \Omega^{-1}_{t|t-1}$$
$$= \Sigma_{t|t-1} H \Omega^{-1}_{t|t-1} \qquad (8)$$

Finally, the mean squared forecast error of states given history y^t, $\Sigma_{t+1|t}$, is a bit harder to derive analytically, as it requires the MSFE of states given history y^t, $\Sigma_{t|t}$.

$\mathbf{\Sigma_{t|t}}$. $\Sigma_{t|t}$ itself can be conceptually understood as the forecast error given (for now!) the MSFE given history y^{t-1}, $\Sigma_{t|t-1}$, minus the Kalman gain on this forecast error variance, $K_t H' \Sigma_{t|t-1}$:

$$\Sigma_{t|t} \equiv E\left((x_t - x_{t|t})(x_t - x_{t|t})'|y^t\right)$$
$$= \Sigma_{t|t-1} - K_t H' \Sigma_{t|t-1} \qquad (11)$$

A detailed derivation of this formula is given in Hamilton (1994), see equations [13.2.16] and [4.5.31], which can only be sketched here. First, write

$$\begin{aligned}\Sigma_{t|t} &\equiv E\left((x_t - x_{t|t})(x_t - x_{t|t})'|y^t\right)\\ &= E\left([(x_t - x_{t|t-1}) - K_t(y_t - H'x_{t|t-1})] \times [(x_t - x_{t|t-1})' - (y_t - H'x_{t|t-1})'K_t']|y^t\right)\\ &= E\left((x_t - x_{t|t-1})(x_t - x_{t|t-1})'|y^t\right) - E\left((x_t - x_{t|t-1})(y_t - H'x_{t|t-1})'K_t'|y^t\right)\\ &\quad - E\left(K_t(y_t - H'x_{t|t-1})(x_t - x_{t|t-1})'|y^t\right) + E\left(K_t(y_t - H'x_{t|t-1})(y_t - H'x_{t|t-1})'K_t'|y^t\right)\end{aligned}$$

Let us write the variance-covariance matrix of the system $([x_t', (y^{t-1})']', y_t, x_t)$ as

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_t & 0 & 0 \\ S_{21}S_{11}^{-1} & 1 & 0 \\ S_{31}S_{11}^{-1} & K_t H'(H'\Sigma_{t|t-1}H + R)^{-1} & 1 \end{bmatrix} \times$$
$$\begin{bmatrix} S_{11} & 0 & 0 \\ 0 & H'\Sigma_{t|t-1}H + R & 0 \\ 0 & 0 & \Sigma_{t|t-1} - K_t H'(H'\Sigma_{t|t-1}H + R)^{-1} H'\Sigma_{t|t-1} \end{bmatrix}$$
$$\begin{bmatrix} \mathbf{I}_t & S_{11}^{-1}S_{12} & S_{11}^{-1}S_{13} \\ 0 & 1 & (H'\Sigma_{t|t-1}H + R)^{-1}K_t'H \\ 0 & 0 & 1 \end{bmatrix}$$

As shown in Hamilton (1994), p. 379f. and p. 98f., this specifies the MSFE, $\Sigma_{t|t}$, as the last block of the block-diagonalised matrix:

$$\begin{aligned}\Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}H(H'\Sigma_{t|t-1}H + R)^{-1}H'\Sigma_{t|t-1}\\ &= \Sigma_{t|t-1} - K_t H'\Sigma_{t|t-1}\end{aligned}$$

This completes the sketched derivation of (11).

$\mathbf{\Sigma_{t+1|t}}$. The formula for $\Sigma_{t+1|t}$ can be derived as

$$\begin{aligned}
\Sigma_{t+1|t} &\equiv E\left((x_{t+1}-x_{t+1|t})(x_{t+1}-x_{t+1|t})'|y^t\right) \\
&= E\left((Fx_t+Gw_{t+1}-Fx_{t|t})(Fx_t+Gw_{t+1}-Fx_{t|t})'|y^t\right) \\
&= FE\left((x_t-x_{t|t})(x_{t-1}-x_{t|t})'|y^t\right)F' + GE\left(w_{t+1}(x_t-x_{t|t})'|y^t\right)F' \\
&\quad + FE\left((x_t-x_{t|t})w_{t+1}'|y^t\right)G' + GE\left(w_{t+1}w_{t+1}'|y^t\right)G' \\
&= F\Sigma_{t|t}F' + 0 + 0 + GQG' \\
&= F\Sigma_{t|t}F' + GQG' \qquad (12)
\end{aligned}$$

Using (11) and (8) in (12), we finally get

$$\Sigma_{t+1|t} = F\Sigma_{t|t-1}F' + GQG' - FK_t\Omega_{t|t-1}K_t'F' \qquad (9)$$

This completes the derivations for the KF algorithm.

4 Kalman filter for DSGE models

I only sketch how to obtain a log-linearised DSGE model in state-space form and how to adjust the matrices above in that case.
Using Dynare or other codes based on Blanchard and Kahn (1980), e.g. the one outlined in Uhlig (1995), we get a log-linearised DSGE model in the form

$$\begin{aligned}
\mathrm{NSV}_t &= P\cdot\mathrm{NSV}_{t-1} + Q\cdot\mathrm{XSV}_t \qquad (13)\\
\mathrm{CV}_t &= R\cdot\mathrm{NSV}_{t-1} + S\cdot\mathrm{XSV}_t \qquad (14)\\
\mathrm{XSV}_t &= N\cdot\mathrm{XSV}_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0,\Sigma)
\end{aligned}$$

where NSV denotes "endogenous state variables", XSV means "exogenous state variables" and CV "control variables" (see Uhlig 1995, p.19, who calls CV "endogenous other variables"). To write the likelihood function of this model given the observables y_t, we have to transform the model into state-space form as

$$\begin{aligned}
x_t &= Fx_{t-1} + G\varepsilon_t \\
y_t &= H'x_t + J\varepsilon_t
\end{aligned}$$

where $\varepsilon_t \sim N(0, I)$ and the states x_t and the matrices A and B are given by[6]

$$x_t = (\text{NSV}_t, \text{XSV}_t)$$
$$F = \begin{bmatrix} P & Q \cdot N \\ 0 & N \end{bmatrix}$$
$$G = \begin{bmatrix} Q \\ I \end{bmatrix} \Sigma^{1/2}$$

Try it, it works! Note, however, that this formulation in itself would not allow us to use any CVs (like e.g. output or investment in many models) as observables for our model. So we use the trick to include the those CVs for which we want to use observable counterparts into the NSVs via a simple identity. So the equation for NSV_t above, (13), contains variables that are not in fact states (so-called "pseudo-states").[7]

Now we only need to connect the individual observable series with their counterpart (controls or pseudo states) in the model. As now rows in y_t correspond to rows in x_t, we just need a "matching" matrix M to select the right columns from x_t: M has zero entries except for a unit entry for one element i per row, which connects the observable series in that row to row i in x_t:[8]

$$y_t = Mx_t + \nu_t$$

where $\nu_t \sim N(0, R)$ again represents measurement error.

Note that we do not in fact use (14): Intuitively, we get rid of all control variables CV_t on which we have no information from observables. There are more complex settings possible, but I will focus on this one below.[9]

6. If you have predetermined or lagged variables in your model, you will have to specify $x_t = (\text{NSV}_t, \text{NSV}_{t-1}, \text{XSV}_t)$, in which case F and G become

$$F = \begin{bmatrix} P & 0 & Q \cdot N \\ I & 0 & 0 \\ 0 & 0 & N \end{bmatrix}, \quad G = \begin{bmatrix} Q \\ 0 \\ I \end{bmatrix}$$

With more lags, matrices expand accordingly.

7. The equations are imputed into the Uhlig algorithm as the matrices for the following system of equations ($x = \text{ESV}$; $y = \text{CV}$; $z = \text{XSV}$):

$$\begin{aligned} 0 &= A\text{ESV}_t + B\text{ESV}_{t-1} + C\text{CV}_t + D\text{XSV}_t \\ 0 &= \mathbb{E}_t \left[F\text{ESV}_{t+1} + G\text{ESV}_t + H\text{ESV}_{t-1} + H\text{CV}_{t+1} + K\text{CV}_t + L\text{XSV}_{t+1} + M\text{XSV}_t\right] \\ \text{XSV}_{t+1} &= N\text{XSV}_t + \varepsilon_t, \quad \text{where } \varepsilon_t \sim N(0, I) \end{aligned}$$

Hence, we can easily include some pseudo-states into the vectors ESV_t and CV_t that give "harmless" identities (using $+1/-1$ entries in the matrices A and C as well as G and K). This leaves us with an enlarged vector $\tilde{\text{ESV}}_t$ to be used in the following.

8. So if $y_{1,t} = \text{GDP}_t$, and output is located as third variable in the state vector, $G_{13} = 1$, while $G_{i3} = 0$ for all $i \neq 3$.

9. For example, you can include variables with non-zero mean by including another column into H and F. You

5 Two examples

5.1 Maximum likelihood estimation of a DSGE model

This example shortly explains the use of the KF for maximising the likelihood of a log-linearised model.[10] Assume you used a routine like the one outlined in Uhlig (1995) to obtain a DSGE model in state-space form, as explained in Section 4. You start with the state-space model

$$x_t = Fx_{t-1} + G\varepsilon_t$$
$$y_t = Mx_t + \nu_t$$

where $\varepsilon_t \sim N(0, I)$, and the measurement error is $\nu_t \sim N(0, R)$. Make sure your data are chosen, detrended and scaled to fit their model counterparts in y_t as closely as possible (if in doubt, adjust the measurement error variance-covariance matrix R accordingly). Use my routine `example1.m` to calculate the (log) likelihood of the model – take the model, data and data treatment as given.[11]

To redo the analysis, run `example1.m`. I initialise a parameter vector Θ with several model parameters (again, details are not important for the KF here). After some initialisation steps, I use the Matlab minimisation routine `fmincon` on the function `model_solve.m`. This function gets the (negative) model (log) likelihood for a certain parameter vector by solving the model using `uhligsolve.m`[12] (calling Harald Uhlig's `solve2.m`) and passing the state-space matrices to `kalmanfilter.m` to obtain the likelihood. Then we can apply `kalmanf.m` to extract the model likelihood. The maximum likelihood estimates are indeed quite different from the initial values and the (negative log) likelihood increases somewhat for the new parameter values. While a more careful setting of bounds could allow even better results, one could in principle also add a full-blown Metropolis-Hastings algorithm for Bayesian estimation.[13]

also can include growth rates, if you have NSV_{t-1} in the model, below in the second row:

$$H' = \begin{bmatrix} \bar{y}_1 & P_2 & 0 & Q_2 \\ \bar{y}_2 & P_5 & -P_5 & Q_5 \cdot (N-1) \\ \bar{y}_3 & P_3 & 0 & Q_3 \\ \bar{y}_4 & P_1 & 0 & Q_1 \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & P & 0 & Q \cdot N \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & N \end{bmatrix}$$

10. This is based on a project for Fabio Canova, where the likelihood then entered a Metropolis-Hastings algorithm.
11. It is a standard New Keynesian DSGE model with a financial friction as in Bernanke, Gertler, and Gilchrist (1999). The observables are output, consumption, investment, CPI, policy rates and credit spreads as in Gilchrist and Zakrajsek (2012). Measurement error is assumed to take the intermediate level of $R = I \cdot 0.5$.
12. Note the inclusion of "pseudo states" like `p_y(t)` in this code.
13. Adding this to my homepage is a distant goal... Stay tuned :)

5.2 Fitting a Taylor rule with time-varying coefficients

Assume you want to estimate a Taylor Rule with time-varying coefficients with data for the US.[14] The process of interest is

$$R_t = \beta_{\pi,t}\Pi_t + \beta_{y,t}Y_t + \varepsilon_{r,t}, \text{ where } \varepsilon_{r,t} \sim N(0, \sigma_r^2) \tag{15}$$

Note that this is our measurement equation, while the transition of unobservable Taylor rule coefficients (our states) $\beta_{\pi,t}$ and $\beta_{y,t}$ is assumed to follow independent random walk processes:

$$\begin{bmatrix}\beta_{\pi,t}\\ \beta_{y,t}\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\beta_{\pi,t-1}\\ \beta_{y,t-1}\end{bmatrix} + \begin{bmatrix}\varepsilon_{\pi,t}\\ \varepsilon_{y,t}\end{bmatrix}, \text{ where } \begin{bmatrix}\varepsilon_{\pi,t}\\ \varepsilon_{y,t}\end{bmatrix} \sim N(\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\sigma_\pi^2 & 0\\ 0 & \sigma_y^2\end{bmatrix}\right))$$

I use three quarterly US time series for the time 1982Q1 to 2007Q2: The Federal funds rate for R_t as well as real GDP growth rate and the GDP deflator growth rate for Y_t and Π_t. As I have no strong prior on the most suitable standard deviations – σ_r, σ_π and σ_y – I use a grid search over some possible values, ranging from very small, $\sigma_i = e^{-6}$ to rather large $\sigma_i = 10$, for $i \in \{r, \pi, y\}$. I choose the variances depending on which gives the highest likelihood for the model.[15]

Run the code `example2.m` to redo the Kalman filtering. Note the twist here: As $\beta_{\pi,t}$ and $\beta_{y,t}$ are our unobservable states, the (time-varying, but deterministically given) vector $[\Pi_t, Y_t]$ replaces the time-constant matrix H in (4). Moreover, as we are interested in the variation of the states (i.e. $\beta_{\pi,t}$ and $\beta_{y,t}$) over time, I use a slightly altered variant of the KF routine, `kalmanf_xt.m` (the "`xt`" hints at the backing out of the states). Figure 1 plots the estimates for the Taylor rule coefficients as well as the used observable series.

14. This example is loosely based on a problem set by Massimiliano Marcellino, solved with Dominik Thaler.

15. Not surprisingly, the model favours high variances, allowing for a better fit, so σ_i is always close at the upper bound for all variables. The difference for the coefficient estimates, however, is negligible.

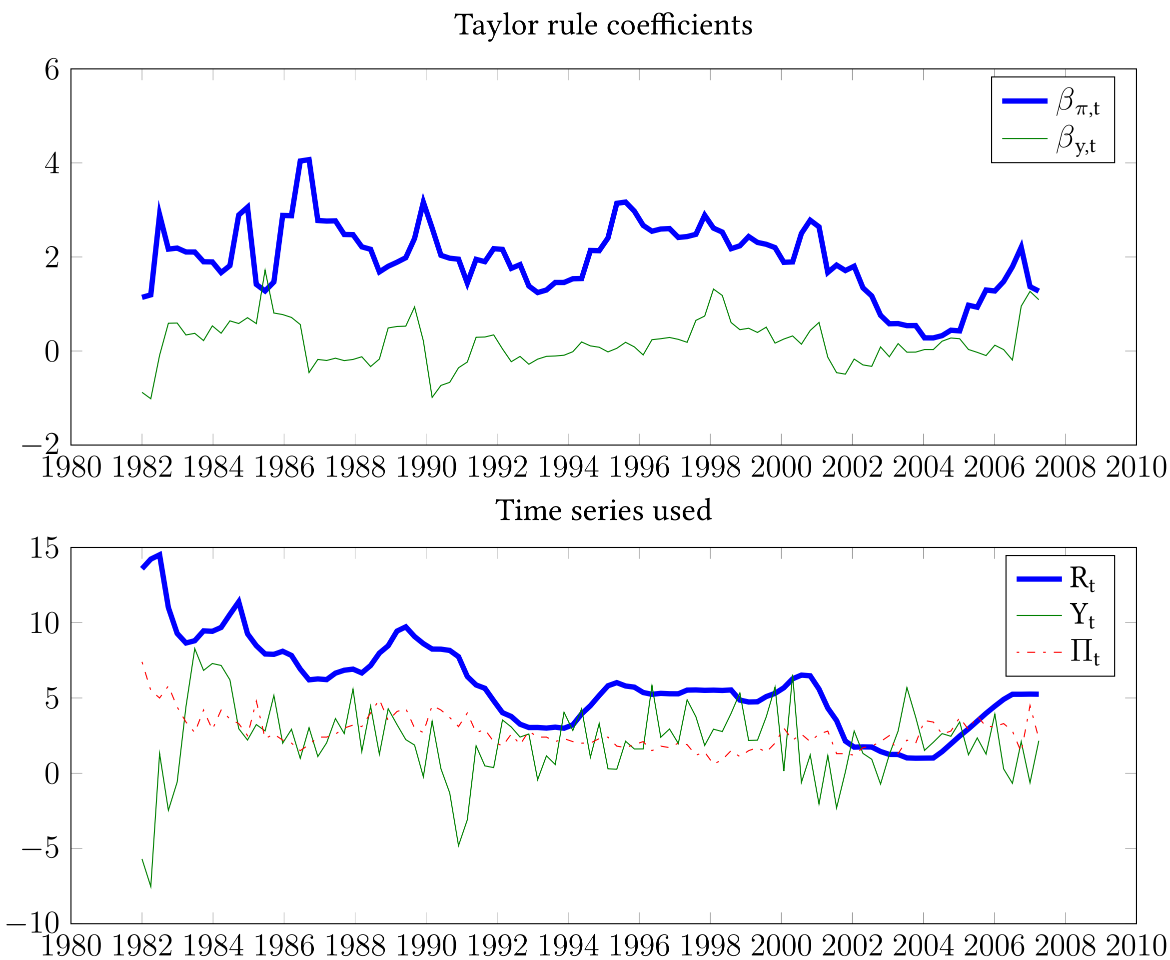

Figure 1: Results for Example 2 (time-varying TR coefficients for US)

There is indeed evidence for time-varying Taylor rule coefficients (e.g., one can recognise a period of an exceptionally low inflation coefficient around 2004). The averages for $\beta_{\pi,t}$ and $\beta_{y,t}$ over the considered period are 0.18 and 1.95, which is very close to standard calibration for those values in standard DSGE models.[16]

6 Note on notation

Notation means confusion. Compare my notation to the one in Hamilton (1994), Canova (2007) and the slides by JFV:

16. Note that I do not include a constant $\beta_{0,t}$ in eq. (15), which I found to absorb most of the movement in R_t (drastically reducing the means of $\beta_{\pi,t}$ and $\beta_{y,t}$).

here	Hamilton	Canova	JFV
y_t	$\mathbf{y}_t$	y_t	z_t
x_t	ξ_t	α_t	x_t
H	$\mathbf{H}$	x_{1t}	H
F	$\mathbf{F}$	$\mathbb{D}_{1t}$	F
$\Sigma_{t\|t}$	$\mathbf{P}_{t\|t}$	$\Omega_{t\|t}$	$\Sigma_{t\|t}$
$\Omega_{t\|t}$	–	$\Sigma_{t\|t}$	$\Omega_{t\|t}$
K_t	$\mathbf{K}_t$	$\tilde{K}_t$	K_t
B[17]	$\mathbf{A}'$	–	–

Table 1: Comparison of Notation

References

Anderson, Brian D.O., and John B. Moore. 1979. *Optimal Filtering.* Edited by Thomas Kailath. Prentice Hall.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist. 1999. "The financial accelerator in a quantitative business cycle framework." Chap. 21 in *Handbook of Macroeconomics,* edited by J. B. Taylor and M. Woodford, 1:1341–1393. Handbook of Macroeconomics. Elsevier.

Blanchard, Olivier Jean, and Charles M. Kahn. 1980. "The Solution of Linear Difference Models under Rational Expectations." *Econometrica* 48, no. 5 (July): 1305–11.

Canova, Fabio. 2007. *Methods for Applied Macroeconomic Research.* Princeton University Press.

Fernández-Villaverde, Jesús, and Juan F. Rubio-Ramírez. 2007. "Estimating Macroeconomic Models: A Likelihood Approach." *Review of Economic Studies* 74 (4): 1059–1087.

Gilchrist, Simon, and Egon Zakrajsek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102, no. 4 (June): 1692–1720.

Hamilton, James D. 1994. *Time Series Analysis.* Princeton University Press.

Koopman, Siem Jan. 1997. "Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models." *Journal of the American Statistical Association* 92, Issue 440:1630–1638.

Uhlig, H. 1995. *A toolkit for analyzing nonlinear dynamic stochastic models easily.* Discussion Paper 1995-97. Tilburg University, Center for Economic Research.