HG ECON 5101

Sixth lecture – 26 Feb. 2014

1. Introduction to the Kalman filter

- Rudolf Kálmán, an electrical engineer, was born in Budapest in 1930, and emigrated to the US in 1943.
- Noted for his co-invention of the Kalman filter (or Kalman-Bucy Filter) developed by Kalman (and others before him) (1958 – 1961).
- Applied by Kalman under the Apollo program (1960) for navigation of space crafts.

General overview.

- The Kalman filter (KF) is an efficient way to organize many complex econometric models for estimation and prediction purposes.
- The Kalman filter is basically a VAR(1) model [*or VARX(1) with exogenous covariate series*], where some of the variables in the random VAR–vector are *latent* (i.e., non- observable).
- The variables in the random VAR-vector are divided into two classes called *state variables* (denoted by $\{\xi_t\}$ by Hamilton) and *observable variables*, (denoted by $\{y_t\}$ by Hamilton).
- The latent variables in the model (except some error terms) are always included in the *state vector,* ξ_t. The state vector, however, may also include some observable variables.
- The total random vector to be modelled, $\begin{pmatrix} \xi_t \\ y_t \end{pmatrix}$ is modelled with a VAR(1) (or VARX(1)) structure, where *separate* equations are formulated for the state vector ξ_t and the observation vector y_t.

The Kalman filter à la Hamilton

The random vector time series $\begin{pmatrix} \xi_t \\ y_t \end{pmatrix}$: r states in $\xi_t \sim r \times 1$,

n observable variables in $y_t \sim n \times 1$

State equation (SE): $\xi_{t+1} = \underset{r\times r}{F}\,\xi_t + v_{t+1}, \quad t = 0,1,2,\ldots$

Observation equation (OE): $y_t = \underset{n\times r}{H'}\,\xi_t + \underset{n\times k}{A'}\,x_t + w_t, \quad t = 1,2,\ldots$

where the errors, $\begin{pmatrix} v_t \\ w_t \end{pmatrix}$ are white noise, uncorrelated with the history, $D_{t-1} = \{y_1, y_2, \ldots, y_{t-1}, x_1, \ldots, x_t\}$ (i.e., independent under normality). (See Hamilton section 13.1 for details.)

In addition the two vectors v_{t+1}, w_t are uncorrelated with each other.

As white noise v_{t+1}, w_t have expectation 0 and covariance matrices Q and R respectively. Written in short:

$$v_{t+1} \sim WN(0,Q), \quad w_t \sim WN(0,R)$$

where, in general, Q and R are assumed to be positive semi-definite (*which includes the possibility that some of the error terms may be zero*).

Example 1: Suppose $\{y_t\}$ is an AR(2) –process,

$$y_{t+1} - \mu = \varphi_1(y_t - \mu) + \varphi_2(y_{t-1} - \mu) + \varepsilon_{t+1}, \quad t = 2,3,\ldots$$

(or, if you wish, $y_{t+1} = \eta + \varphi_1 y_t + \varphi_2 y_{t-1} + \varepsilon_{t+1}$, where $\eta = \mu(1 - \varphi_1 - \varphi_2)$)

where $\varepsilon_t \sim WN(0, \sigma^2)$.

State space form: Set $\xi_{t+1} = \begin{pmatrix} y_{t+1} - \mu \\ y_t - \mu \end{pmatrix}$

SE: $\xi_{t+1} = \begin{pmatrix} y_{t+1} - \mu \\ y_t - \mu \end{pmatrix} = F\xi_t + v_{t+1} = \begin{pmatrix} \varphi_1 & \varphi_2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} y_t - \mu \\ y_{t-1} - \mu \end{pmatrix} + \begin{pmatrix} \varepsilon_{t+1} \\ 0 \end{pmatrix}, \quad t = 2,3,\ldots$

OE: $y_t = H'\xi_t + A'x_t + w_t = [1,0]\xi_t + \mu + 0 = [1,0]\begin{pmatrix} y_t - \mu \\ y_{t-1} - \mu \end{pmatrix} + \mu + 0$, (i.e., $w_t = 0$)

Notice:

$$Q = \text{var}(v_{t+1}) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix} \sim \text{ positive semi-definite}, \quad R = \text{var}(w_t) = 0 \ \sim \ \text{ positive semi-definite}$$

Further notes.

- In econometric applications it is often assumed that $(v'_{t+1}, w'_t)'$ are multinormally distributed – which among other things – implies that *uncorrelatedness* can be replaced by *independence*.

- Hamilton's formulation is a simplification (probably for pedagogical reasons). It is more common (e.g., in the Stata manual) to include the possibility of exogenous variables in the **SE** as well:

  New SE: $\xi_{t+1} = \underset{r\times r}{F}\,\xi_t + Bx_{t+1} + v_{t+1}, \quad t = 0,1,2,\ldots$

- Note that many authors formulate the **SE** with *t* instead of *t*+1. Note that Hamilton's **SE** is equivalent to saying

  SE: $\xi_t = \underset{r\times r}{F}\,\xi_{t-1} + v_t, \quad t = 1,2,3,\ldots$

- The KF provides formulas for (linearly) predicting - at each given time point *t* – $\left(\hat{\xi}_{t+1|t}, \hat{y}_{t+1|t}\right)$ based on information $(\hat{\xi}_t, y_t)$ where $\hat{\xi}_t$ is calculated from the history at time *t*, $D_t = \{y_t, y_{t-1}, y_{t-2}, \ldots \text{ and } x_t, x_{t-1}, \ldots\}$. $\hat{\xi}_t$ represents an update of $\hat{\xi}_{t|t-1}$ when y_t becomes available. This procedure is called forward prediction (called "*filtering*").

- The KF also provides formulas for backward prediction (called "*smoothing*"- i.e., updating all earlier predictions $\hat{\xi}_t$ based on the information from the total observed series, $y_0, y_1, \ldots, y_t, \ldots, y_T$.

- The KF provides an efficient way to formulate the likelihood (usually Gaussian) for many complicated models. For example Stata uses Kalman filters for estimating ARMA-models.

- The maximum likelihood estimators (ML) based on the Gaussian likelihood are true ML-estimators when the error terms in the model are Gaussian white noise vectors. Otherwise they are quasi maximum likelihood estimators with an asymptotic normal limit distribution under general stability conditions of the KF.

- There are many fruitful extensions of the KF in the literature. The present KF is *time homogeneous*. A common extension is to let some of the matrices in SE and OE depend on *t*. The present KF is *linear* – so there are non-linear KF's in the literature. Of course, there are KF extensions to continuous time as well etc.

- In general, the KF idea is particular suited for Bayesian analysis where, at each time *t*, there are prior distributions for state variables ξ_t and parameters produced at time $t-1$, from which posterior distributions at time *t* are calculated as observation y_t becomes available.

Example 2: Ex ante real interest rate

(Briefly sketched in Hamilton page 376)
Let

$\pi_t^e =$ expected (in the market) inflation in quarter *t*, i.e., expected ΔCPI_t (latent)

$i_t =$ interest rate quarter t (observable)

$\pi_t =$ observed inflation in quarter t $= \Delta CPI_t$ (observable)

$i_t - \pi_t^e =$ ex ante real interest rate (latent)

$i_t - \pi_t =$ observed (ex post) real interest rate (observable)

We imagine that the ex ante r.i.r. varies stationarily (as in Hamilton) around an unknown mean, μ. The latent state, $\xi_t = i_t - \pi_t^e - \mu$, then varies stationarily around 0, and we assume an AR(1) process for ξ_t :

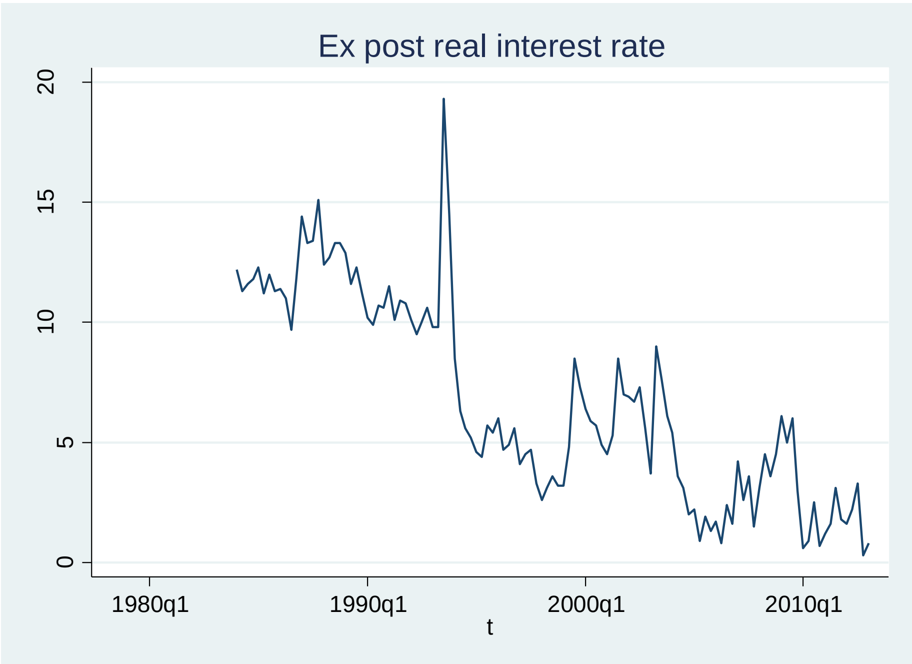
SE: $\xi_{t+1} = \varphi \xi_t + v_{t+1}$ where $v_t \sim WN(0, \sigma_s^2)$ ($\Rightarrow$ $r = 1$ og $F = \varphi$)

For the observation equation we have
$y_t = i_t - \pi_t = i_t - \pi_t^e + \pi_t^e - \pi_t = \mu + \xi_t + w_t$, where the error $w_t = \pi_t^e - \pi_t$ is supposed to be independent of the history $D_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$, and behaves like white noise. Hence

OE: $y_t = i_t - \pi_t = \mu + \xi_t + w_t$, where $w_t \sim WN(0, \sigma_o^2)$
($\Rightarrow A'x_t = \mu \cdot 1$ og $H' = 1$)

In addition we assume that the error terms, v_t, w_t, are independent and normally distributed.

We look at some Norwegian data



Not very stationary (!), but can possibly be estimated by an AR(1) stationary near the unit circle… (There is a continuous transition from stationarity to a non-stationary random walk (in finite time periods!) as some root of the AR lag-polynomial approaches the unit circle.)

Estimation of the model (example 2).

In Stata we use the command `sspace` to handle KF's.

Model:

SE: $\xi_{t+1} = \varphi\xi_t + v_{t+1}$ where $v_t \sim WN(0, \sigma_s^2)$

OE: $y_t = i_t - \pi_t = \mu + \xi_t + w_t$, where $w_t \sim WN(0, \sigma_o^2)$

Stata estimation.

```
Ex ante/post realrente                        (In data I call y = expostrr)

constraint 1 [ expostrr ]ksi=1                (I define the constraint that
                                              the coef. in front of ksi is 1 in OE)

sspace (ksi L.ksi ,state noconstant)( expostrr ksi), covstate(diagonal) constraints(1)

                                              (I define the latent variable in SE
                                              as ksi. All the equations are listed
                                              in sspace in parentheses. An SE-
                                              equation has the option state. L.ksi
                                              means ksi lagged one time unit. There
                                              are two methods for specifying error
                                              terms. I have chosen covstate in the
                                              options(see manual).
searching for initial values ..........
(setting technique to bhhh)
Iteration 0:   log likelihood = -266.84772
Iteration 1:   log likelihood = -232.19078
Iteration 2:   log likelihood = -228.70186
Iteration 3:   log likelihood = -228.45484
Iteration 4:   log likelihood = -228.44293
(switching technique to nr)
Iteration 5:   log likelihood = -228.44167
Iteration 6:   log likelihood = -228.44114
Iteration 7:   log likelihood = -228.44114
Refining estimates:
Iteration 0:   log likelihood = -228.44114
Iteration 1:   log likelihood = -228.44114

State-space model

Sample: 1984q1 - 2013q1                           Number of obs    =        117
                                                  Wald chi2(1)     =     781.88
Log likelihood = -228.44114                       Prob > chi2      =     0.0000
 ( 1)  [expostrr]u = 1
--------------------------------------------------------------------------------
               |                 OIM
      expostrr |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
ksi            |
           ksi |
           L1. |   .9476893   .0338918    27.96   0.000     .8812626    1.014116
---------------+----------------------------------------------------------------
expostrr       |
           ksi |          1  (constrained)
(my-hat)_cons  |   6.661696   2.144945     3.11   0.002     2.457682    10.86571
---------------+----------------------------------------------------------------
(sigma(s)sq.:)
      var(ksi)|   1.921453   .6540334     2.94   0.002     .6395714    3.203335
(sigma(o) sq.:)
 var(expostrr)|   .5427117   .3846031     1.41   0.079            0     1.29652
--------------------------------------------------------------------------------
```

```
Note: Tests of variances against zero are one sided, and the two-sided confidence
intervals are truncated at zero.
```

Some interpretation of output:

- It turns out that the Kalman-filter estimes φ to 0.95 which indicates almost random walk (RW).
- The variances, σ_s^2, σ_o^2 were estimated as 1.9 and 0.5 respectively, indicating relatively high volatility in the *ex ante r.i.r.* (too high for the approximate RW to be interpreted reasonably as a *locally constant trend RW*.

- It is tempting to interpret the fact that $\hat{\sigma}_o^2$ is relatively small as a tendency that *ex ante r.i.r.* follows (the observable) *ex post r.i.r.* rather tightly – i.e., that much of the variation in y_t has been "taken over" by ξ_t.
- In other words: The results seem to indicate a co-integration relationship between *ex post* and *ex ante r.i.r.*

Using the `sspace` post-estimation command `predict`, we can get the forward one-step-ahead predictions (the filterings) of `ksi`, and the backwards smoothing predictions of `ksi` from the total series:

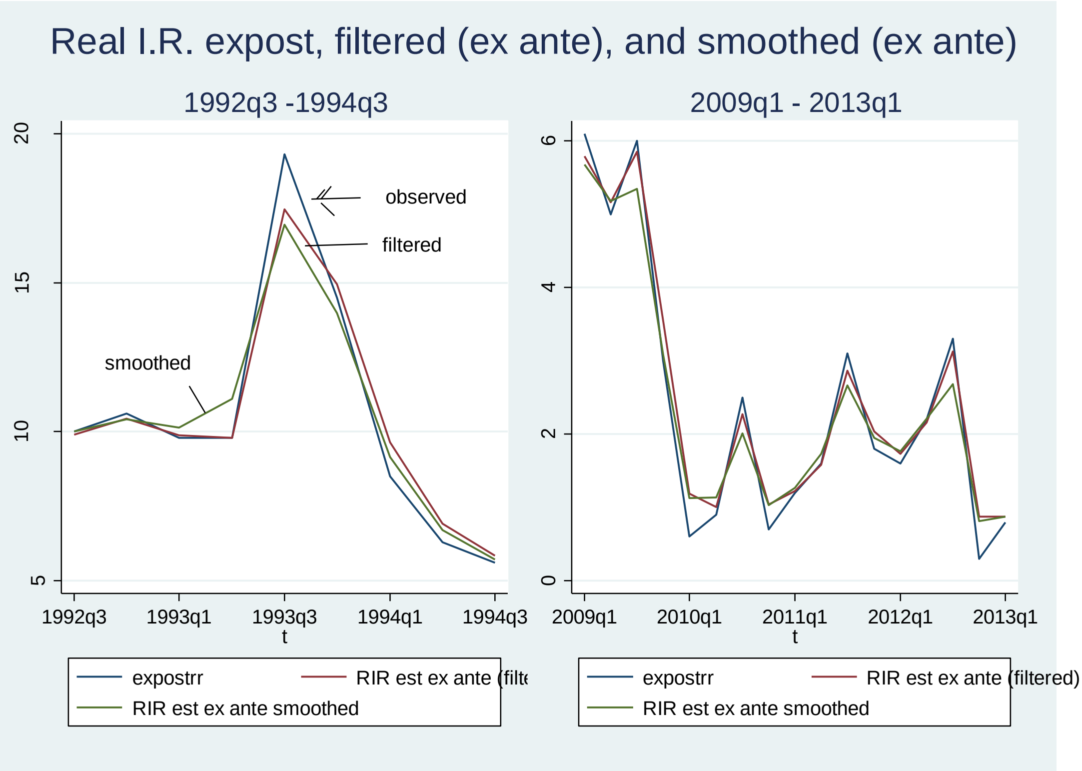
```
predict exfilter1,states smethod(filter) equation(ksi)

predict exsmooth1,states smethod(smooth) equation(ksi)
```

To list some values to compare with $y_t =$ expostrr, I need to add $\hat{\mu} = 6.6617$ to the predicted ksi's:

```
. gen rrfilter1= exfilter1+6.6617
(25 missing values generated)

. gen rrsmooth1= exsmooth1+6.6617
(25 missing values generated)


list t expostrr rrfilter1 rrsmooth1 if tin(2009q1,2013q1), compress noobs

+--------------------------------+
  |      t   exp~r   rrf~1   rrs~1 |
  |--------------------------------|
  | 2009q1     6.1     5.8    5.68 |
  | 2009q2       5    5.16    5.18 |
  | 2009q3       6    5.86    5.35 |
```

2009q4	3	3.55	3.09
2010q1	.6	1.19	1.13
2010q2	.9	1.01	1.14
2010q3	2.5	2.27	2.01
2010q4	.7	1.04	1.03
2011q1	1.2	1.23	1.27
2011q2	1.6	1.58	1.73
2011q3	3.1	2.86	2.66
2011q4	1.8	2.04	1.95
2012q1	1.6	1.73	1.77
2012q2	2.2	2.16	2.21
2012q3	3.3	3.13	2.68
2012q4	.3	.872	.817
2013q1	.8	.871	.871

Using the `twoway` command `graph combine`, I can produce the following graphs,



One can filter forwards and backwards

- Forward filtering (called *filtering*) filters, updates, and predicts next period ($t+1$) every time a new observation y_t is available.

- Backwards filtering (called *smoothing*)) is used for the complete data set, $y_1, y_2, \ldots, y_T$, and updates earlier predictions by recursive calculations backwards from time *T*.

Forwards filtering

SE: $\xi_{t+1} = \underset{r\times r}{F}\xi_t + v_{t+1}, \quad t = 0,1,2,\ldots \qquad \dim(\xi_t) = r$

OE: $y_t = \underset{n\times k}{A'}x_t + \underset{n\times r}{H'}\xi_t + w_t, \quad t = 1,2,\ldots \quad \dim(y_t) = n$

Example 2 r.i.r:

$$n = r = 1$$
$$y_t = i_t - \pi_t$$
$$\xi_t = i_t - \pi_t^e - \mu \qquad (\mu = E(\text{ex ante r.i.r.}) = E(i_t - \pi_t^e))$$

SE: $\xi_{t+1} = \varphi\xi_t + v_{t+1}$ where $v_t \sim WN(0,\sigma_s^2)$ $\quad (\Rightarrow \; r = 1$ og $F = \varphi)$

OE: $y_t = i_t - \pi_t = \mu + \xi_t + w_t$, where $w_t \sim WN(0,\sigma_o^2)$

$$(\Rightarrow A'x_t = \mu\cdot 1 \text{ og } H' = 1)$$

$$Q = \sigma_s^2, \quad R = \sigma_o^2$$

Time period *t*:

Step 1 – before y_t is observed:

Info from period *t* – 1:

$$\hat{\xi}_{t|t-1} = E(\xi_t | D_{t-1}) \overset{Ex.2}{\sim} 1\times 1 \qquad (D_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\})$$

$$P_{t|t-1} = MSE(\hat{\xi}_{t|t-1}) = E\left[\left(\xi_t - \hat{\xi}_{t|t-1}\right)\left(\xi_t - \hat{\xi}_{t|t-1}\right)' \middle| D_{t-1}\right] \overset{Ex.2}{=} p_{t|t-1} \sim 1\times 1$$

$\Rightarrow$

$$\hat{y}_{t|t-1} = A'x_t + H'\hat{\xi}_{t|t-1} \overset{Ex.2}{=} \mu + \hat{\xi}_{t|t-1}$$

$$MSE_{t-1}\left(\hat{y}_{t|t-1}\right) = H'P_{t|t-1}H + R \overset{Ex.2}{=} p_{t|t-1} + \sigma_o^2$$

Step 2 – updating after y_t is observed:

$$\hat{\xi}_{t|t} = E(\xi_t \big| D_t) = \hat{\xi}_{t|t-1} + P_{t|t-1}H\left(H'P_{t|t-1}H + R\right)^{-1}\left(y_t - \hat{y}_{t|t-1}\right)$$

$$\overset{Ex.2}{=} \hat{\xi}_{t|t-1} + \frac{p_{t|t-1}}{p_{t|t-1} + \sigma_o^2}(y_t - \mu - \hat{\xi}_{t|t-1})$$

$$P_{t|t} = MSE_t(\hat{\xi}_{t|t}) = P_{t|t-1} - P_{t|t-1}H\left(H'P_{t|t-1}H + R\right)^{-1}H'P_{t|t-1} =$$

$$\overset{Ex.2}{=} p_{t|t} = p_{t|t-1} + \frac{p_{t|t-1}^2}{p_{t|t-1} + \sigma_o^2}$$

Step 3 – predict ξ_{t+1} after y_t is observed:

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t} = F\hat{\xi}_{t|t-1} + K_t(y_t - \hat{y}_{t|t-1}) \quad \text{where}$$

$K_t = FP_{t|t-1}H\left(H'P_{t|t-1}H + R\right)^{-1}$ is called the "*gain matrix*".

In the example:

$$\hat{\xi}_{t+1|t} = \varphi\hat{\xi}_{t|t} = \varphi\hat{\xi}_{t|t-1} + \varphi\frac{p_{t|t-1}}{p_{t|t-1} + \sigma_o^2}(y_t - \hat{y}_{t|t-1})$$

with MSE

$$P_{t+1|t} = MSE_t(\hat{\xi}_{t+1|t}) = FP_{t|t}F' + Q =$$

$$\overset{Ex.2}{=} p_{t+1|t} = \varphi^2\left(p_{t|t-1} + \frac{p_{t|t-1}^2}{p_{t|t-1} + \sigma_o^2}\right) + \sigma_s^2$$

Steady state

Proposition 13.2 in Hamilton says that under stationarity of the state vector, $\{\xi_t\}$ series, the gain, K_t, and the MSE, $P_{t|t-1}$ will often stabilize when *t* increases. More precisely:

Prop. 13.2. If (i) the eigenvalues of F in **SE** all are less than 1 in absolute value (i.e., that $\Rightarrow$ stationarity of ξ_t), and (ii) if at least one of the two error covariance matrices, Q and R, are positive definite (not only semi pos. def.), then

$P_{t|t-1} \underset{t\to\infty}{\rightarrow} P$ (for an arbitrary pos. semi-definite start, $P_{1|0}$), and

$K_t \underset{t\to\infty}{\rightarrow} K = FPH\left(H'PH+R\right)^{-1}$ where P is the solution of

(*) $P = F\left[P - PH\left(H'PH+R\right)^{-1}H'P\right]F' + Q$

and where P and K are uniquely determined no matter the start values at $t=0$.[1]

In example 2 (which satisfies the conditions if $|\varphi|<1$), $p_{t|t-1} \underset{t\to\infty}{\rightarrow} p$, where p is the solution of the second degree equation

$$p = \varphi^2\left[p - \frac{p^2}{p+\sigma_o^2}\right] + \sigma_s^2,$$

which has only one positive solution, i.e.,

$$p_{t|t-1} \underset{t\to\infty}{\rightarrow} p = \frac{1}{2}\left(-B + \sqrt{B^2 + 4\sigma_o^2\sigma_s^2}\right) \text{ where } B = (1-\varphi^2)\sigma_o^2 - \sigma_s^2$$

(estimate p: 3.9383)

$$K_t \underset{t\to\infty}{\rightarrow} K = \varphi\frac{p}{p+\sigma_o^2}$$ (estimate K: 0.88176)

[1] If both Q and R are only pos. semi definite, the stabilization/convergence is still valid (prop 13.1) but the limits P and K are no longer necessarily unique. They may depend on the start values at $t = 0$.

Stable updating relations in example 2.

$$\hat{y}_{t|t-1} = 6.6617 + \hat{\xi}_{t|t-1}$$

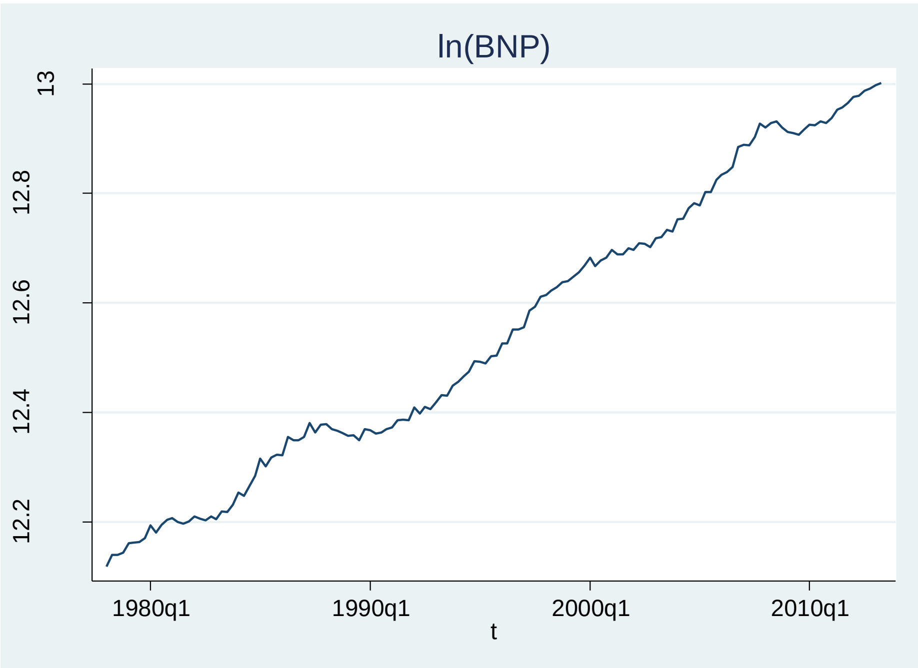
$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + 0.9304(y_t - \hat{y}_{t|t-1})$$

$$\hat{\xi}_{t+1|t} = \varphi\hat{\xi}_{t|t} = \varphi\hat{\xi}_{t|t-1} + K(y_t - \hat{y}_{t|t-1}) =$$
$$(0.9477)\hat{\xi}_{t|t-1} + (0.88176)(y_t - \hat{y}_{t|t-1})$$

$$MSE_t(\hat{\xi}_{t+1|t}) = p = 3.9383$$

Example 3 Local linear trend (structural time series)

Consider the ln BNP data.



We want to model the trend. There are many ways to model trends. One way is Harvey's local linear trend model as an example of his structural time series approach..

Let $y_t = \ln BNP_t$. We imagine that that y_t can be decomposed

(1) $y_t = \mu_t + w_t$

where μ_t is the trend component and w_t an error. Both μ_t and w_t are latent random variables.

One possibility is the "locally constant" trend model where

(2) $\mu_t = \mu_{t-1} + v_t$ where $v_t \sim WN(0, \sigma_v^2)$ and independent of $\{\mu_{t-1}, \mu_{t-2}, \ldots, y_t, y_{t-1}, \ldots\}$

This implies that $\mu_t = \mu_0 + v_1 + \cdots + v_t$ is a random walk. However, μ_t behaves as a trend only if the variance of the error v_t is substantially smaller than $\sigma_w^2 = \text{var}(w_t)$ in (1).

It is in this sense we call $\{\mu_t\}$ "locally constant".
This model is well suited for estimation by the KF.

The "locally linear" trend model is an extension of this: Adding a constant, *b*, to (2) we get

(3) $\mu_t = b + \mu_{t-1} + v_{1t} \quad \Rightarrow \quad \mu_t = \mu_0 + bt + v_{11} + \cdots + v_{1t}$

i.e., a random walk with a deterministic trend component (called a "*random walk with drift*"). However, the rate of increase of the deterministic part is fixed in this model. It is natural to think that the rate of increase may vary randomly although slowly. This we can achieve by replacing *b* by a random rate of increase, β_t, where $\beta_t = \beta_{t-1} + v_{2t}$ with $v_{2t} \sim WN(0, \sigma_{v_2}^2)$ and where $\sigma_{v_2}^2 = \text{var}(v_{2t})$ is small relatively to $\text{var}(w_t)$, i.e., a "locally constant" rate of increase. So the locally linear trend model is specified as the following state space model.

OE: $y_t = \mu_t + w_t$

SE:
$$\mu_t = \mu_{t-1} + \beta_{t-1} + v_{1t}$$
$$\beta_t = \beta_{t-1} + v_{2t}$$

or in matrix form

SE: $\xi_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$

OE: $y_t = [1,0]\xi_t + w_t$

The KF will provide one-step-ahead predictions (filtering), or full-sample predictions (smoothing) for the ξ_t's, and estimates for the variance parameters, $\sigma_w^2, \sigma_{v_1}^2, \sigma_{v_2}^2$, where the two last ones should be substantially smaller than σ_w^2 for μ_t to behave like a trend.

Note, in passing, that this model implies that $y_t \sim I(2)$ (unless $\text{var}(v_{2t}) = 0$), since $\Delta y_t = \Delta\mu_t + \Delta w_t = \beta_{t-1} + v_{1t} + \Delta w_t$, i.e., a random walk, and only $\Delta^2 y_t$ is stationary. If, however, $\text{var}(v_{2t}) = 0$, the β_t becomes a constant, and we are back to the case where y_t is a random walk with drift, i.e., $I(1)$.

In the following Stata `sspace` command, I call the two state equation for “s1” and “s2”, and the observation equation for “y”. The restrictions are

```
. constraint 1 [s1]L.s1=1

. constraint 2 [s1]L.s2=1

. constraint 3 [s2]L.s2=1

. constraint 4 [y]s1=1
```

The KF command

```
. sspace (s1 L.s1 L.s2,state noconstant) (s2 L.s2,state noconstant) (y s1,noconst),
constraints(1/4)

-----  Iteration output omitted  (22 iterations)----


State-space model

Sample: 1978q1 - 2013q2                          Number of obs   =        142
Log likelihood =   442.9024
 ( 1)  [s1]L.s1 = 1
 ( 2)  [s1]L.s2 = 1
 ( 3)  [s2]L.s2 = 1
 ( 4)  [y]s1 = 1
------------------------------------------------------------------------------
             |                 OIM
```

```
             y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
s1           |
          s1 |
         L1. |          1   (constrained)
             |
          s2 |
         L1. |          1   (constrained)
-------------+----------------------------------------------------------------
s2           |
          s2 |
         L1. |          1   (constrained)
-------------+----------------------------------------------------------------
y            |
          s1 |          1   (constrained)
-------------+----------------------------------------------------------------
      var(s1)|   2.35e-17          .        .       .            .           .
      var(s2)|   5.20e-06   1.55e-06     3.36   0.000     2.17e-06    8.23e-06
       var(y)|   .0000433   6.07e-06     7.13   0.000     .0000314    .0000552
------------------------------------------------------------------------------
Note: Model is not stationary.
Note: Tests of variances against zero are one sided, and the two-sided confidence
    intervals are truncated at zero.
```

We notice that state variances are substantially smaller than the observation error variance. The $\text{var}(v_{1t}) \approx 0$, but the rate of increase appears to be random $\text{var}(v_{2t}) > 0$.

We obtain the residuals $\hat{w}_t$ for y_t by by the `predict` post-estimation command:

```
. predict resy,resid

. corrgram resy

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1       -0.0220  -0.0221   .06932  0.7923          |                  |
2       -0.0123  -0.0133   .09119  0.9554          |                  |
3        0.0427   0.0422   .35532  0.9493          |                  |
4        0.0564   0.0586   .82032  0.9357          |                  |
5        0.0021   0.0054   .82097  0.9757          |                  |
6       -0.0944  -0.0961   2.1427  0.9061          |                  |
7        0.0108   0.0021   2.1602  0.9504          |                  |
8       -0.1904  -0.1976   7.6215  0.4713         -|                 -|
9        0.0701   0.0714    8.368  0.4975          |                  |
10      -0.1088  -0.1094   10.177  0.4251          |                  |
11      -0.0154   0.0025   10.214  0.5113          |                  |
12      -0.0417  -0.0497   10.484  0.5735          |                  |
13      -0.0415  -0.0394   10.754  0.6314          |                  |
14       0.0760   0.0505   11.666  0.6331          |                  |
15      -0.1159  -0.1082     13.8  0.5407          |                  |
16       0.0067  -0.0564   13.808  0.6130          |                  |
17      -0.0850  -0.0851   14.976  0.5972          |                  |
18      -0.0465  -0.1298   15.327  0.6394          |                 -|
19      -0.0403  -0.0564   15.594  0.6841          |                  |
20      -0.0514  -0.0897   16.031  0.7147          |                  |
21      -0.0304  -0.0828   16.185  0.7591          |                  |
22      -0.2384  -0.3108   25.763  0.2620         -|                --|
23       0.1235  -0.0032   28.357  0.2026          |                  |
```

```
24        0.0157  -0.0352   28.399  0.2436          |                  |
25        0.0450  -0.0277   28.749  0.2745          |                  |
26       -0.0254  -0.0981   28.862  0.3174          |                  |
27        0.1325   0.1345   31.951  0.2339          |-                 |-
28        0.0795  -0.0179   33.074  0.2330          |                  |
29        0.0394   0.0670   33.351  0.2638          |                  |
30        0.0917  -0.0450    34.87  0.2474          |                  |
31       -0.1131  -0.0599   37.203  0.2050          |                  |
32        0.1133   0.0703   39.567  0.1680          |                  |
33       -0.0503  -0.1019   40.037  0.1862          |                  |
34       -0.0453  -0.1792   40.422  0.2077          |                 -|
35       -0.0146  -0.0474   40.462  0.2418          |                  |
36       -0.0687  -0.1421   41.365  0.2479          |                 -|
37        0.0970   0.0063   43.181  0.2241          |                  |
38       -0.0745  -0.2018   44.264  0.2242          |                 -|
39        0.1188   0.0795    47.04  0.1765          |                  |
40        0.0945   0.1534   48.816  0.1599          |                  |-
```

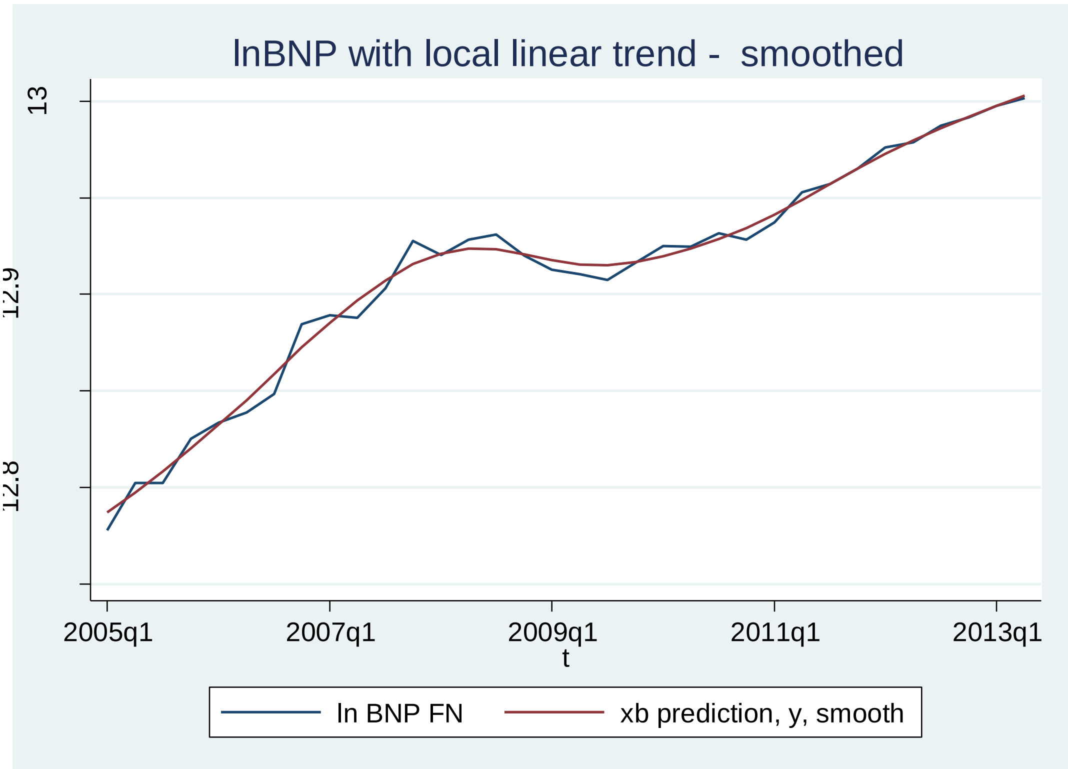
No strong evidence against w_t being WN.

We get the smoothed prediction of the trend $\hat{\mu}_t$ by the `predict` as well:

```
. predict ysmooth,xb smethod(smooth)
```

and the plot

```
. tsline y  ysmooth if t>=tq(2005q1)
```

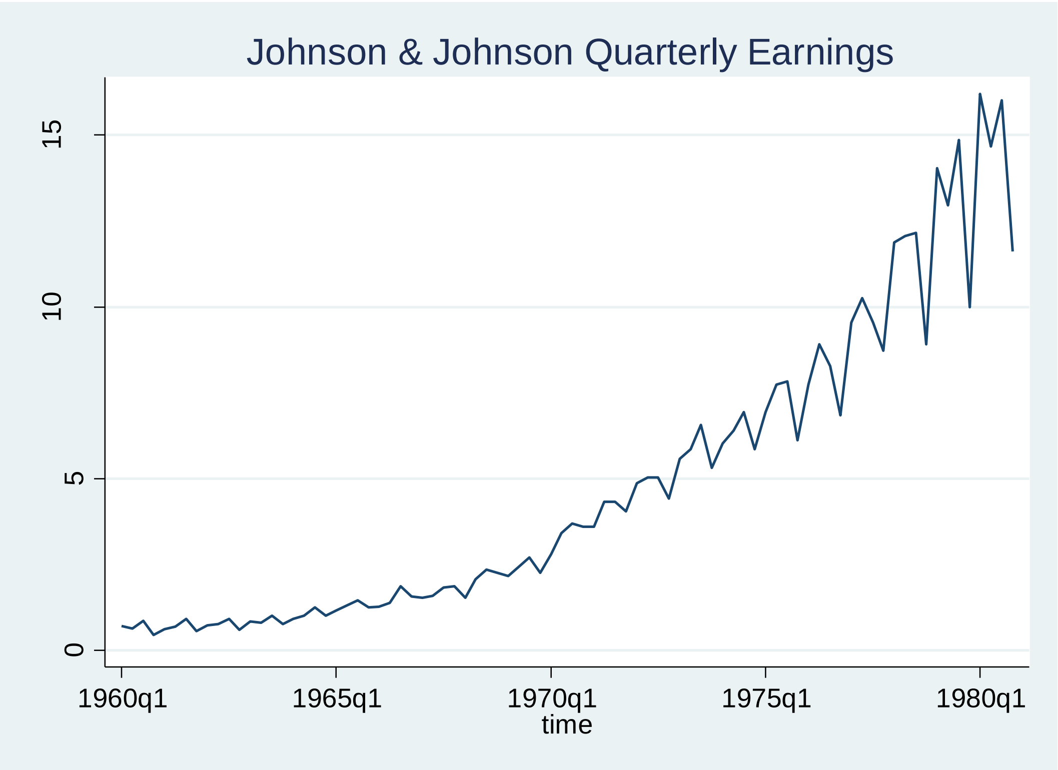


Note. This is an example that the KF handles a non-stationary dynamic model with standard asymptotic normal inference theory. There are general results in the literature that says that this is the case when the parameters determining the non-stationary part of the model are all specified as *known* (i.e., the case here).

As the cointegration theory will show, this is not always the case when some parameters of the non-stationary part are unknown and must be estimated – especially when some of the eigenvalues of F in the **SE** lie exactly on the unit circle. Then the inference theory must be modified (i.e., the Søren Johansen theory). In such cases the Stata `sspace` command will refuse to work.

On the other hand, `sspace` sometimes works well in other non-stationary cases with unknown parameters for the non-stationary part of the model as long as there are no eigenvalues on the unit circle. The following example taken from the recommendable, R.H. Shumway and D.S. Stoffer, "*Time Series Analysis and Its Applications*", Springer 2000, is an illustration.

Example 4 Johnson & Johnson Quarterly Sales Data



These data are highly non-stationary and have strong seasonal effects. They also appear to show an increasing volatility.

It turns out, as Shumway and Stoffer (S&S) point out, that these data are hard to bring to stationarity by differencing and transformations (log or n-th root transformations). So they decide to analyze the data using a *wrong* model (with the error terms failing to be white noise), but where the KF give meaningful results. They use a different estimation approach than provided by `sspace`.

But even `sspace` can handle wrong models to a certain degree by the quasi likelihood method (QML),, which turns out to provide quite similar results as in S&S.

Again we use the structural approach as in example 3, decomposing the quarterly observed time series, y_t, as

(4) $$y_t = T_t + S_t + w_t$$

where T_t represents the trend, S_t, the seasonal effect (i.e., the seasonal deviation from the trend), and w_t the error – preferably WN. In the deterministic case T_t would be some fixed function of *t* and S_t fixed seasonal deviations, i.e., $S_1 + S_2 + S_3 + S_4 = 0$ (since there are 4 quarters in a year), or, i.e., $S_t + S_{t-1} + S_{t-2} + S_{t-3} = 0$ for all t.

To increase the flexibility we make both components random:
Looking at the plot, the trend appears to be close to exponential growth in the observation period. So we specify

(5) $$T_t = \varphi T_{t-1} + v_{1t} \quad \text{where} \quad \varphi > 1$$

and where the error term v_{1t} makes the trend random. Similarly for the seasonal effects we specify

(6) $S_t + S_{t-1} + S_{t-2} + S_{t-3} = v_{2t}$ where v_{2t} is a error term that should be WN according to the KF assumptions. Looking at the plot, this does not appear very realistic since the seasonal effects appear to grow steadily with time.

Let the state vector be $\xi_t' = (T_t, S_t, S_{t-1}, S_{t-2})$. Then

$$\textbf{SE:} \quad \xi_t = \begin{pmatrix} T_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{pmatrix} = \begin{pmatrix} \varphi & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} T_{t-1} \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \\ 0 \\ 0 \end{pmatrix}$$

$$\textbf{OE:} \quad y_t = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} T_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{pmatrix} + w_t$$

The errors of the model satisfy $w_t \sim WN(0, R)$, where $R = \sigma_w^2$, and

$$v_t = \begin{pmatrix} v_{1t} \\ v_{2t} \\ 0 \\ 0 \end{pmatrix} \sim WN(0, Q), \quad \text{where} \quad Q = \begin{pmatrix} \sigma_{v_1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{v_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

For sspace I call the 4 state equations u1,…,u4, and the observation equation y. We need the following 7 restrictions,

```
. constraint 1 [u2]L.u2=-1

. constraint 2 [u2]L.u3=-1

. constraint 3 [u2]L.u4=-1

. constraint 4 [u3]L.u2=1

. constraint 5 [u4]L.u3=1

. constraint 6 [y]u1=1

. constraint 7 [y]u2=1
```

To compare with the QML I do the regular ML estimation first.

```
. sspace (u1 L.u1,state noconst) (u2 L.u2 L.u3 L.u4,state noconstant) (u3
L.u2,state noerror noconst) (u4 L.u3, state noerror noconst) (y u1 u2,
noconst),covstate(diagonal) constraints(1/7)

------ Iteration output omitted  -----

State-space model

Sample: 1960q1 - 1980q4                          Number of obs   =         84
                                                 Wald chi2(1)    =  165386.98
Log likelihood = -48.239979                      Prob > chi2     =     0.0000
 ( 1)  [u2]L.u2 = -1
 ( 2)  [u2]L.u3 = -1
 ( 3)  [u2]L.u4 = -1
 ( 4)  [u3]L.u2 = 1
 ( 5)  [u4]L.u3 = 1
 ( 6)  [y]u1 = 1
 ( 7)  [y]u2 = 1
------------------------------------------------------------------------------
             |                 OIM
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
u1           |
          u1 |
         L1. |   1.035097   .0025452   406.68   0.000     1.030108    1.040085
-------------+----------------------------------------------------------------
u2           |
          u2 |
         L1. |         -1  (constrained)
             |
          u3 |
         L1. |         -1  (constrained)
             |
          u4 |
         L1. |         -1  (constrained)
-------------+----------------------------------------------------------------
u3           |
          u2 |
         L1. |          1  (constrained)
-------------+----------------------------------------------------------------
u4           |
          u3 |
         L1. |          1  (constrained)
-------------+----------------------------------------------------------------
y            |
          u1 |          1  (constrained)
          u2 |          1  (constrained)
-------------+----------------------------------------------------------------
      var(u1)|   .0196384   .0061475     3.19   0.001     .0075896    .0316873
      var(u2)|   .0503249   .0110313     4.56   0.000      .028704    .0719458
       var(y)|   2.84e-15          .        .       .            .           .
------------------------------------------------------------------------------
Note: Model is not stationary.
Note: Tests of variances against zero are one sided, and the two-sided confidence
    intervals are truncated at zero.
```

I also want out-of-sample predictions for four years (the sample period is 1960q1 – 1980q4). So

```
.  tsappend, add(16)
```

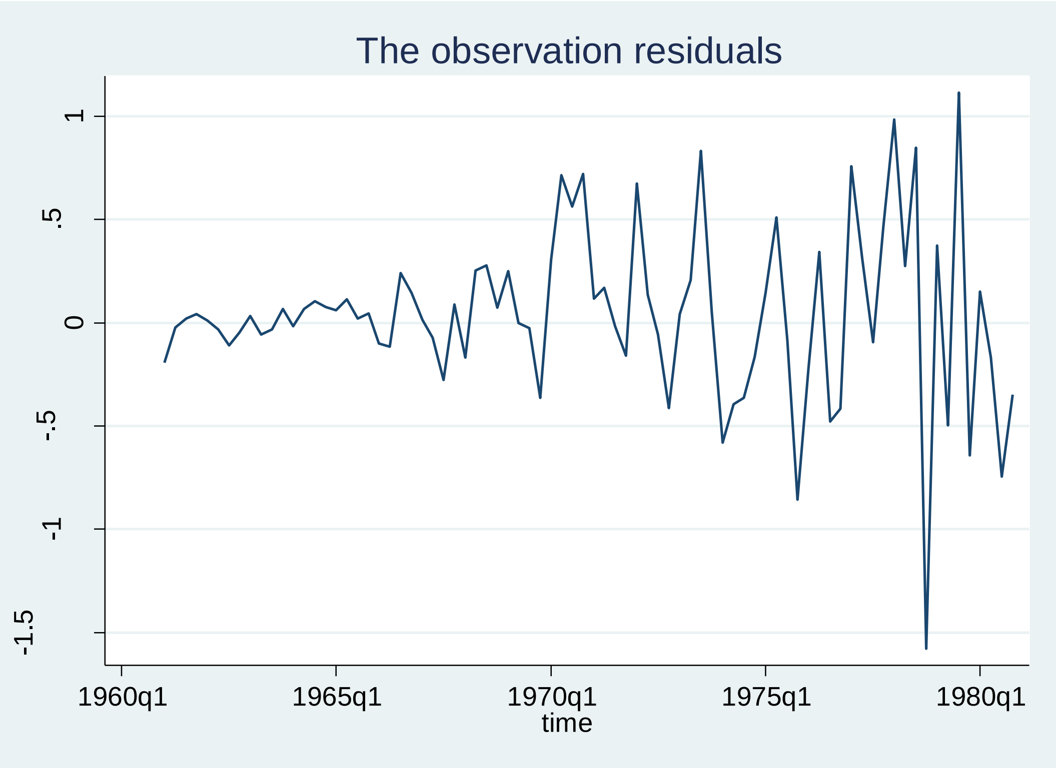
Robust standard deviations (QML – quasi max. likelihood)

```
. sspace (u1 L.u1,state noconst) (u2 L.u2 L.u3 L.u4,state noconstant) (u3
L.u2,state noerror noconst) (u4 L.u3,state noerror noconst) (y u1
u2,noconst),covstate(diagonal) constraints(1/7) vce(robust)


--   iteration output omitted  --

State-space model

Sample: 1960q1 - 1980q4                          Number of obs   =         84
                                                 Wald chi2(1)    =   74318.85
Log likelihood = -48.239979                      Prob > chi2     =     0.0000
 ( 1)  [u2]L.u2 = -1
 ( 2)  [u2]L.u3 = -1
 ( 3)  [u2]L.u4 = -1
 ( 4)  [u3]L.u2 = 1
 ( 5)  [u4]L.u3 = 1
 ( 6)  [y]u1 = 1
 ( 7)  [y]u2 = 1
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
u1           |
          u1 |
         L1. |   1.035097   .0037969   272.61   0.000     1.027655    1.042538
-------------+----------------------------------------------------------------
u2           |
          u2 |
         L1. |         -1  (constrained)
             |
          u3 |
         L1. |         -1  (constrained)
             |
          u4 |
         L1. |         -1  (constrained)
-------------+----------------------------------------------------------------
u3           |
          u2 |
         L1. |          1  (constrained)
-------------+----------------------------------------------------------------
u4           |
          u3 |
         L1. |          1  (constrained)
-------------+----------------------------------------------------------------
y            |
          u1 |          1  (constrained)
          u2 |          1  (constrained)
-------------+----------------------------------------------------------------
     var(u1)|   .0196384   .0107404     1.83   0.034            0    .0406892
     var(u2)|   .0503249   .0212708     2.37   0.009      .008635    .0920148
      var(y)|   2.84e-15          .        .       .            .           .
------------------------------------------------------------------------------
Note: Model is not stationary.
Note: Tests of variances against zero are one sided, and the two-sided confidence
    intervals are truncated at zero.
```

It is interesting to note that the variance of the observation error, w_t, is almost zero, so the trend and seasonals explain almost all of the variation in y_t.

The residuals of OE:

```
. predict resy,resid

. tsline resy if e(sample)
```

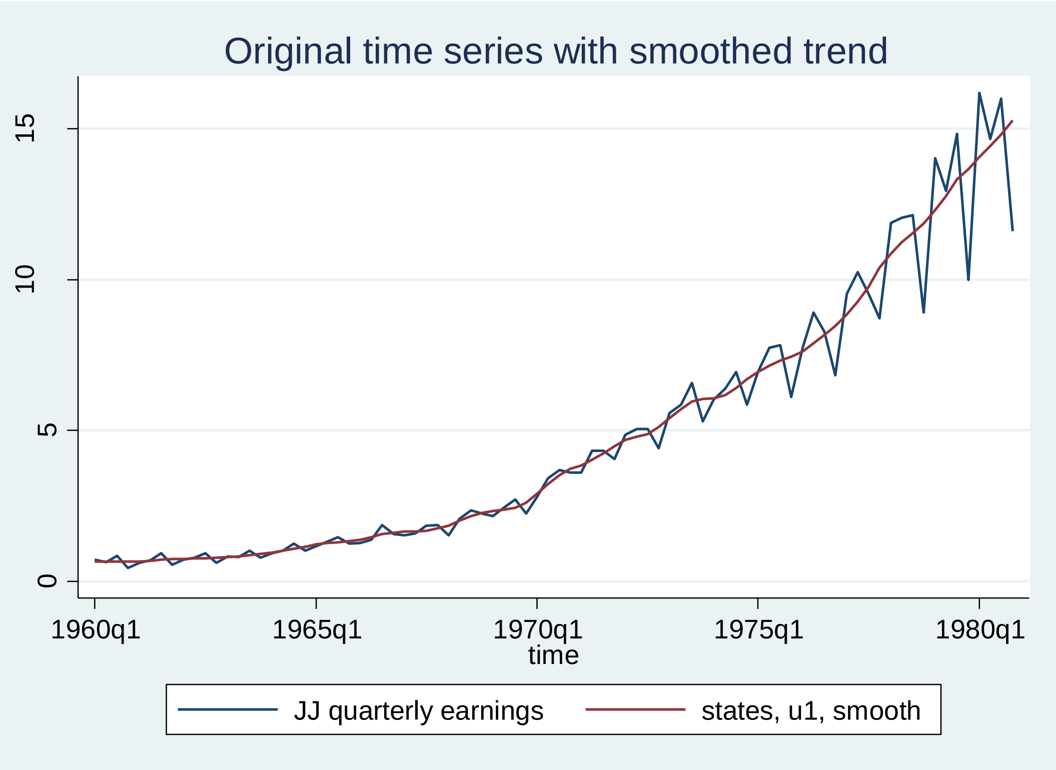


Clearly far from WN.

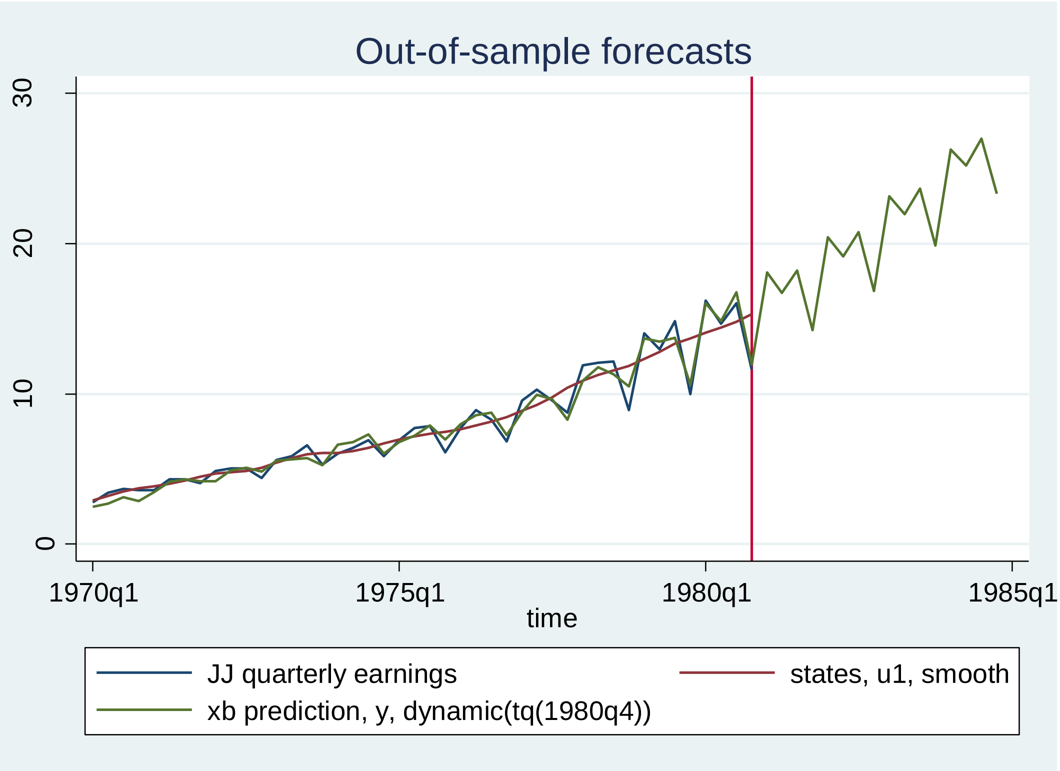
Now forecasts and trend (including "root mean square errors" for the forecasts

```
. predict f_y,dynamic(tq(1980q4)) rmse(rmse_f_y)

. predict trend if e(sample), states smethod(smooth) equation(u1)
```

```
. tsline y trend if e(sample)
```

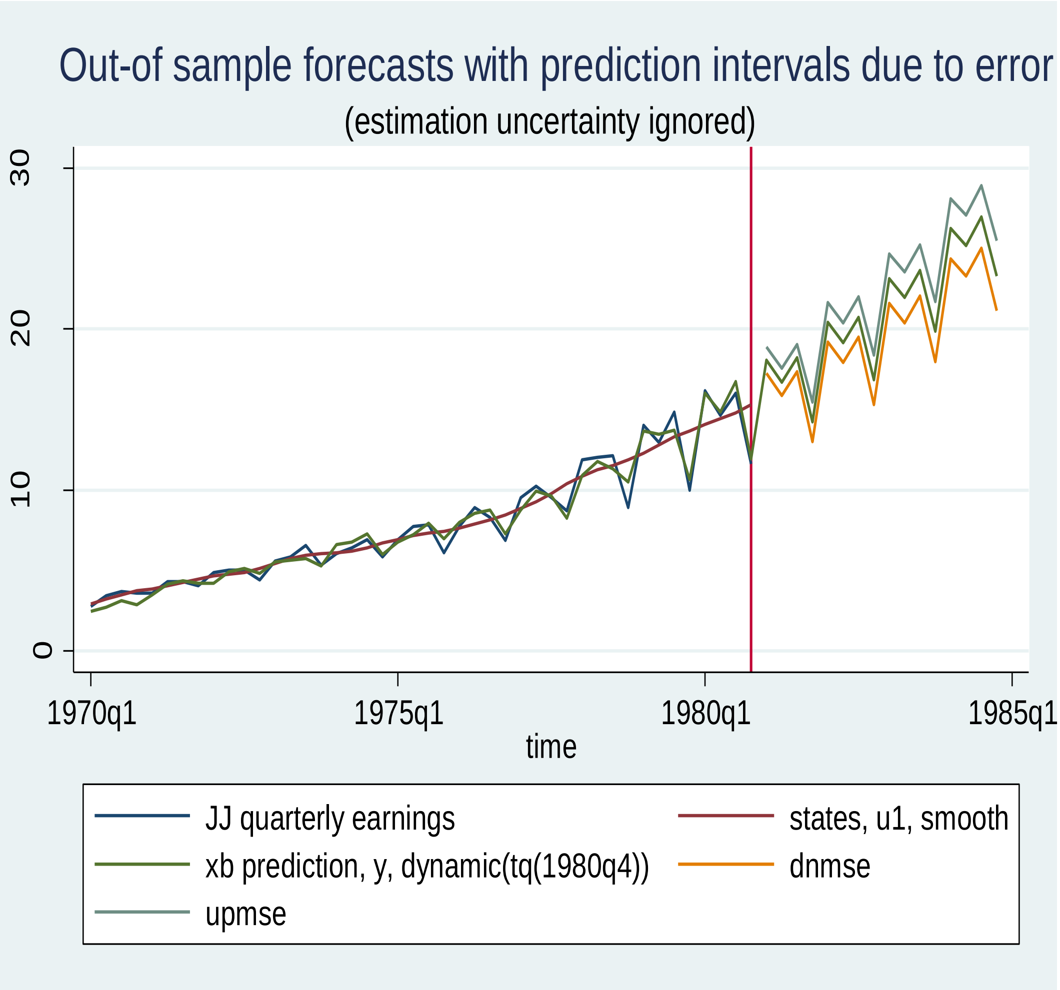



```
. tsline y trend f_y if time >=tq(1970q1), tline(1980q4
```



Partial prediction uncertainty from root MSE based on error variation only

```
. gen upmse =f_y+invnorm(0.975)* rmse_f_y if time >=tq(1981q1)
(84 missing values generated)

. gen dnmse =f_y+invnorm(0.025)* rmse_f_y if time >=tq(1981q1)
(84 missing values generated)

. tsline y trend f_y dnmse upmse if time >=tq(1970q1), tline(1980q4)
```



Prediction uncertainty grows with *t*.

Further details of the forecasts can be achieved by the `forecast` command.