HG ECON 5101

Fifth lecture – 12 Feb. 2014

1. Introduction to VAR(p) - (section 4 in fourth lecture)

See lecture notes 4.

2. Some notes on prediction and estimation for VAR(p)

Prediction: We will briefly consider the VAR(1) case only.

Suppose $Y_t = \nu + AY_{t-1} + u_t$ is causal stationary and stable (i.e., the eigenvalues of A are less than 1 in absolute values.), where $u_t \sim WN(0, \Sigma_u)$.

The causal stationary solution is given by Theorem 4 in LN4 (page 17):

(1) $Y_t = \mu + u_t + Au_{t-1} + A^2 u_{t-2} + \cdots$ where $A^j \underset{j\to\infty}{\to} \underline{0}$, and

$\mu = E(Y_t) = (I - A)^{-1}\nu$

We want to predict

(2) $Y_{t+h} = \mu + u_{t+h} + Au_{t+h-1} + A^2 u_{t+h-2} + \cdots + A^{h-1}u_{t+1} + A^h u_t + A^{h+1}u_{t-1} + \cdots$

Using the general principle in Theorem 3 in LN4 (page 6), we get the best linear predictor (which will be same as the conditional expectation under the stronger white noise assumption, $E(u_t \mid u_{t-1}, u_{t-2}, \ldots) = 0$)

(3) $\hat{Y}_{t+h|t} = \mu + A^h u_t + A^{h+1}u_{t-1} + \cdots$

Since $\hat{Y}_{t+h-1|t} = \mu + A^{h-1}u_t + A^h u_{t-1} + \cdots$, we get

$$\nu + A\hat{Y}_{t+h-1|t} = \nu + A\mu + A^h u_t + A^{h+1} u_{t-1} + \cdots = \nu + A\mu + \hat{Y}_{t+h|t} - \mu =$$
$$= \nu - (I-A)\mu + \hat{Y}_{t+h|t} = \nu - (I-A)(I-A)^{-1}\nu + \hat{Y}_{t+h|t} = \nu - \nu + \hat{Y}_{t+h|t} = \hat{Y}_{t+h|t}$$

So the best linear predictor satisfies

(4) $\hat{Y}_{t+h|t} = \nu + A\hat{Y}_{t+h-1|t}$ (where, in particular, $\hat{Y}_{t+1|t} = \nu + AY_t$)

Similarly we get in the VAR(p) case

$$\hat{Y}_{t+1|t} = \nu + A_1 Y_t + \cdots + A_p Y_{t-p}$$
$$\hat{Y}_{t+2|t} = \nu + A_1 \hat{Y}_{t+1|t} + A_2 Y_t + \cdots + A_p Y_{t-p+1}$$

and in general

$$\hat{Y}_{t+h|t} = \nu + A_1 \hat{Y}_{t+h-1|t} + A_2 \hat{Y}_{t+h-2|t} + \cdots + A_p \hat{Y}_{t+h-p|t}$$

The mean reversion phenomenon still holds under stability: In the VAR(1) this follows directly from (3) since $A^h \underset{h\to\infty}{\to} \underline{0}$. Rewriting (3), this becomes even more clear:

$$\hat{Y}_{t+h|t} = \mu + A^h u_t + A^{h+1} u_{t-1} + \cdots = \mu + A^h (u_t + Au_{t-1} + \cdots) = \mu + A(Y_t - \mu) =$$
$$= (I - A^h)(I-A)^{-1}\nu + A^h Y_t$$

which converges to $(I-A)^{-1}\nu = \mu = E(Y_t)$ as $h \to \infty$.

The mean squared error, MSE(*h*), of $\hat{Y}_{t+h|t}$:

From (2) and (3) we get

(5) $Y_{t+h} - \hat{Y}_{t+h|t} = u_{t+h} + Au_{t+h-1} + \cdots + A^{h-1} u_{t+1}$

And hence

$$MSE(h) = E\left(Y_{t+h} - \hat{Y}_{t+h|t}\right)\left(Y_{t+h} - \hat{Y}_{t+h|t}\right)' =$$

$$(6) \qquad = E\left(u_{t+h} + Au_{t+h-1} + \cdots + A^{h-1}u_{t+1}\right)\left(u_{t+h} + Au_{t+h-1} + \cdots + A^{h-1}u_{t+1}\right)' =$$

$$= \Sigma_u + A\Sigma_u A' + \cdots + A^{h-1}\Sigma_u (A^{h-1})'$$

From theorem 4 in LN4 (page 17) we have a general formula for the covariance matrix, Σ_Y, of the vector Y_t

$$\Sigma_Y = \Gamma_Y(0) = \sum_{j=0}^{\infty} A^j \Sigma_u (A^j)'$$

Hence MSE(h) increases with h and converges, $\lim_{h\to\infty} MSE(h) = \Sigma_Y = \text{cov}(Y_t)$.

If $\{u_t\}$ is Gaussian, we can calculate prediction intervals for the individual variables $Y_{k,t+h}$ in Y_{t+h}, since then $Y_{t+h} - \hat{Y}_{t+h|t} \sim N(0, MSE(h))$, implying that $Y_{k,t+h} - \hat{Y}_{k,t+h|t} \sim N(0, \tau_k^2)$, where τ_k^2 is the main diagonal element no. k in MSE(h). Therefore, $\dfrac{Y_{k,t+h} - \hat{Y}_{k,t+h|t}}{\tau_k} \sim N(0,1)$, from which, $\hat{Y}_{k,t+h|t} \pm z_{\alpha/2}\tau_k$ becomes a $1-\alpha$ prediction interval for Y_{t+h} (where $z_{\alpha/2}$ is the upper $\alpha/2$-quantile in N(0,1)).

More details can be found in Hamilton chap. 11.

Estimation.

The estimation problem of VAR(p) is similar to the univariate AR(1) case summarized in LN3 page 7. Some points

- The VAR(p) can be estimated consistently by OLS (in various variants as indicated in LN3). This is in contrast to the more complicated estimation problem for VARMA models. However, for moderate sample sizes, these estimates “suffer” from some biasedness (Hurwitz-bias).
- Nowadays it is more common to use MLE (maximum likelihood) based on the Gaussian likelihood, that, to a certain extent compensates for the

biasedness. This is, for example, the default estimation method used in Stata for most stable time series models.

- If the error terms $\{u_t\}$ are truly Gaussian, the estimators are true MLE with asymptotic normal distribution, and where the (asymptotic) standard errors are determined by the classical Fisher information matrix.
- If the error terms $\{u_t\}$ are *not* Gaussian, it is still common to use the (wrong) Gaussian likelihood function for constructing estimators. Estimators based on a wrong likelihood function are usually called *quasi ML estimators*. The theory for using wrong models (and wrong likelihood functions) was developed in the 1980's, and it was proven that in many cases (including the stable VAR case), the quasi ML-estimators are still consistent and asymptotic normally distributed with the same estimates as before – only the standard errors needed modification (the so called "sandwich estimates" for standard errors were developed). Most modern softwares have implemented these modifications of the standard errors. In Stata the option `vce(robust)` is implemented for most estimations routines, and calculates "sandwich" standard errors that to a certain degree compensates for the use of a wrong likelihood.

3. 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 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).
- 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.

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, $\begin{pmatrix} v_{t+1} \\ w_t \end{pmatrix}$ are white noise, uncorrelated with the history,
$D_t = \{y_1, y_2, \ldots, y_t, x_1, \ldots, x_t\}$ (i.e., independent under normality).
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*).

Note 1. 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*.

Note 2. 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 3. Note that many authors formulate the **SE** with t instead of t+1. Note that Hamilton's **SE** is equivalent to saying

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

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:

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

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 :

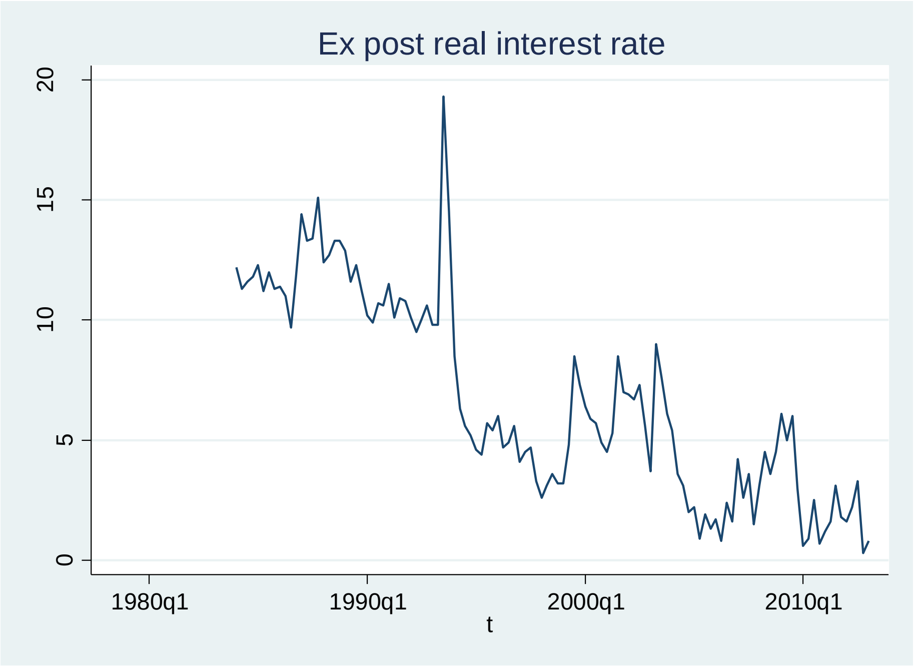
SE: $\xi_{t+1} = \varphi \xi_t + v_{t+1}$ where $v_t \sim WN(0, \sigma_s^2)$ ($\Rightarrow$ $r = 1$ og $F = \varphi$)

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* (see an exercise on RW's possibility to model trends in the seminar III, 24 Feb.)).
- 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:

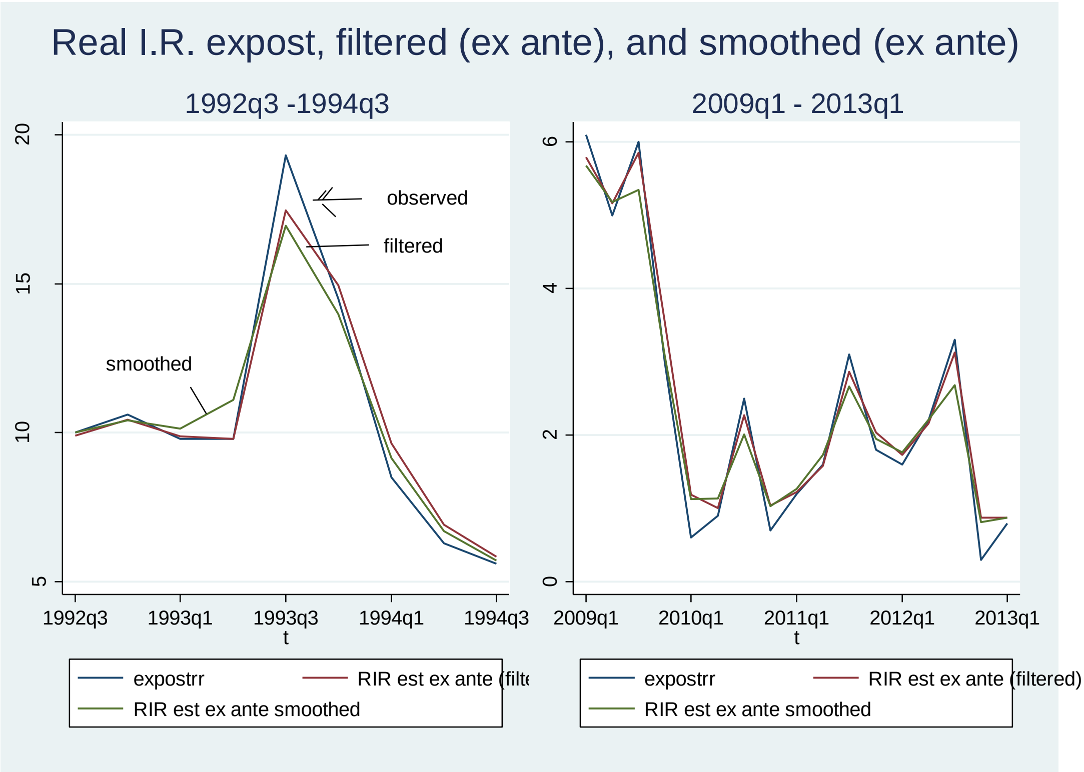
```
. gen rrfilter1= exfilter1+6.6617
(25 missing values generated)

. 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,



We will return to this example in the next lecture 26 Feb.- together with further discussion and examples of the Kalman filter methods.