State-space methods and the Kalman filter

Òscar Jordà and Karel Mertens

AEA Continuing education 2023

January 4, 2023

Main references

- Hamilton, James D. 1994. *Time Series Analysis*. Chapter 13. Princeton University Press: Princeton, NJ
- Hamilton, James D. 1994. State-Space Models, c. 50. In *Handbook of Econometrics, v. 4*, Engle, Robert F. and Daniel L. McFadden (eds.). North-Holland, Elsevier: Amsterdam, the Netherlands.

- Harvey, Andrew C. 1989. *Forecasting Structural Time Series Models and the Kalman Filter.* Cambridge University Press: Cambridge, U.K.

- Durbin J. and Siem J. Koopman. 2001. *Time Series Analysis by State Space Methods.* Oxford Statistical Science Series 24. Oxford University Press: Oxford, U.K.
- Kim, Chang-Jin and Charles R. Nelson. 1999. *State-Space Models with Regime Switching.* MIT Press: Cambridge, MA.

State-Space representation

Set-up and my notation

Previously:

$$\begin{cases} \boldsymbol{s}_t = \mathcal{G}\boldsymbol{s}_{t-1} + \mathcal{F}\boldsymbol{\epsilon}_t & \text{state equation} \\ \boldsymbol{z}_t = \mathcal{A}\boldsymbol{s}_{t-1} + \mathcal{D}\boldsymbol{\epsilon}_t & \text{observation equation} \end{cases}$$

My notation and some changes to the specification:

$$\begin{cases} \underset{r\times 1}{\boldsymbol{\xi}_t} = \underset{r\times r}{F}\,\underset{r\times 1}{\boldsymbol{\xi}_{t-1}} + \underset{r\times 1}{\boldsymbol{v}_t} & \text{state equation} \\ \underset{n\times 1}{\boldsymbol{y}_t} = \underset{n\times r}{H}\,\underset{r\times 1}{\boldsymbol{\xi}_t} + \underset{n\times 1}{\boldsymbol{w}_t} & \text{observation equation} \end{cases}$$

i.e. $\boldsymbol{s}_t$ is $\boldsymbol{\xi}_t$ and $\boldsymbol{z}_t$ is $\boldsymbol{y}_t$ in my notation
also, note the observation equation has $\boldsymbol{\xi}_t$ rather than $\boldsymbol{\xi}_{t-1}$

Set-up

$$\begin{cases} \underset{r\times 1}{\boldsymbol{\xi}_t} = \underset{r\times r}{F}\,\underset{r\times 1}{\boldsymbol{\xi}_{t-1}} + \underset{r\times 1}{\mathbf{v}_t} & \text{state equation} \\ \underset{n\times 1}{\mathbf{y}_t} = \underset{n\times r}{H}\,\underset{r\times 1}{\boldsymbol{\xi}_t} + \underset{n\times 1}{\mathbf{w}_t} & \text{observation equation} \end{cases}$$

$$E\left[\begin{pmatrix} \mathbf{v}_t \\ \mathbf{w}_t \end{pmatrix} \begin{pmatrix} \mathbf{v}_t' & \mathbf{w}_t' \end{pmatrix}\right] = \begin{bmatrix} \underset{r\times r}{Q} & \underset{r\times n}{0} \\ \underset{r\times n}{0} & \underset{n\times n}{R} \end{bmatrix}$$

- $\boldsymbol{\xi}_t$: state, can be observed, usually unobserved
- obs. eqn. with exogenous variables (e.g. constant):

$$\mathbf{y}_t = \underset{n\times k}{A}\,\underset{k\times 1}{\mathbf{x}_t} + H\boldsymbol{\xi}_t + \mathbf{w}_t$$

- Parameter matrices can be time-varying

Key observation

State equation is an AR(1)

$$\boldsymbol{\xi}_t = F\,\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t$$

hence by iterating forecast is easy to compute:

$$E_t[\boldsymbol{\xi}_{t+h}] = F^h \boldsymbol{\xi}_t$$

Many models fit in the state-space representation

Examples

AR(1)

$$y_t = \rho\, y_{t-1} + \epsilon_t \quad \rightarrow \quad \begin{cases} \xi_t & = \rho \xi_{t-1} + \epsilon_t \\ y_t & = \xi_t \end{cases}$$

AR(2)

$$y_t = \rho_1\, y_{t-1} + \rho_2\, y_{t-2} + \epsilon_t \quad \rightarrow$$

$$\left\{ \begin{pmatrix} \xi_{1,t} \\ \xi_{2,t} \end{pmatrix} = \begin{pmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \xi_{1,t-1} \\ \xi_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix} \right.$$

An MA(1) example

$$y_t = \epsilon_t + \theta \epsilon_{t-1} \quad \rightarrow$$

$$\begin{pmatrix} \xi_{1,t} \\ \xi_{2,t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \xi_{1,t-1} \\ \xi_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}$$

$$y_t = \begin{pmatrix} 1 & \theta \end{pmatrix} \begin{pmatrix} \xi_{1,t} \\ \xi_{2,t} \end{pmatrix}$$

A VAR(p) in state-space form

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times n}{\Phi_1}\mathbf{y}_{t-1} + \ldots + \underset{n\times n}{\Phi_p}\mathbf{y}_{t-p} + \underset{n\times 1}{\boldsymbol{\epsilon}_t}$$

$$\underbrace{\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix}}_{\boldsymbol{\xi}_t} = \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \ldots & \Phi_{p-1} & \Phi_p \\ I & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \ldots & I & 0 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix}}_{\boldsymbol{\xi}_{t-1}} + \underbrace{\begin{bmatrix} \boldsymbol{\epsilon}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{v}_t}$$

$$\mathbf{y}_t = \begin{bmatrix} I & 0 & \ldots & 0 \end{bmatrix} \boldsymbol{\xi}_t; \qquad Q = \begin{bmatrix} \Omega & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}$$

More examples

A coincident indicator index by Stock and Watson, 2002 JASA

- Let c_t be a common, unobserved state variable common for $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ macro variables
- Let $\boldsymbol{m}_t = (m_{1,t}, \ldots, m_{n,t})'$ denote corresponding latent stated for each macro variable

Assume:

$$\boldsymbol{y}_t = \boldsymbol{\gamma} c_t + \boldsymbol{m}_t$$

$$\begin{aligned} c_t &= \phi_c c_{t-1} + v_{c,t} \\ \boldsymbol{m}_t &= \Phi_m \boldsymbol{m}_{t-1} + \boldsymbol{v}_{m,t} \end{aligned}$$

A coincident indicator index

Continued

$$\underbrace{\begin{bmatrix} c_t \\ m_{1,t} \\ \vdots \\ m_{n,t} \end{bmatrix}}_{\boldsymbol{\xi}_t} = \underbrace{\begin{bmatrix} \phi_c & 0 & \dots & 0 \\ 0 & \phi_1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \phi_n \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} c_{t-1} \\ m_{1,t-1} \\ \vdots \\ m_{n,t-1} \end{bmatrix}}_{\boldsymbol{\xi}_{t-1}} + \underbrace{\begin{bmatrix} v_{c,t} \\ v_{1,t} \\ \vdots \\ v_{n,t} \end{bmatrix}}_{\mathbf{v}_t}$$

$$\underbrace{\begin{bmatrix} y_{1,t} \\ \vdots \\ y_{n,t} \end{bmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{bmatrix} \gamma_1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \gamma_n & 0 & 0 & \dots & 1 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} c_t \\ m_{1,t} \\ \vdots \\ m_{n,t} \end{bmatrix}}_{\boldsymbol{\xi}_t}$$

Exercise: reproduce MA(1) results in Hamilton (1994): 382–384

Checking covariance-stationarity

AR(1) form of S-S representation makes this easy to check

$$\boldsymbol{\xi}_{t+h} = \mathbf{v}_{t+h} + F\mathbf{v}_{t+h-1} + \ldots + F^{h-1}\mathbf{v}_{t+1} + F^h\boldsymbol{\xi}_t$$

Covariance-stationarity means: $F^h \to 0$ as $h \to \infty$

In practice: if λ_j for $j = 1, \ldots, np$ are the eigenvalues of F then check if $|\lambda_j| < 1$

STATA Example: potential output

`sspace_gdp_trend.do`

One possible model (there are others):

state equations : $\begin{cases} y_t^* = y_{t-1}^* + g_{t-1} & \text{potential output} \\ g_t = g_{t-1} + v_{gt} & \text{growth rate of potential} \end{cases}$

observation equation: $y_t = y_t^* + w_t$

Note: w_t is a measure of the output gap

The Kalman Filter

Intuition

Projecting the multivariate Gaussian

Multivariate Gaussian

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}; \quad \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right]$$

Projecting $\boldsymbol{x}_1$ onto $\boldsymbol{x}_2$:

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_{1|2} \\ \boldsymbol{\mu}_{2|1} \end{pmatrix}; \quad \begin{pmatrix} \Sigma_{11|2} & \Sigma_{12|2} \\ \Sigma_{21|1} & \Sigma_{22|1} \end{pmatrix}\right]$$

where

$$\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 + \underbrace{\Sigma_{12}'\Sigma_{11}^{-1}}_{\textit{like OLS}}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1); \quad \Sigma_{22|1} = \Sigma_{22} - \underbrace{\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}_{\text{“+”}} < \Sigma_{22}$$

Kalman filter recursions

Notation

$$
\begin{aligned}
\boldsymbol{w}_{t|t-1} &= E_{t-1}(\boldsymbol{w}_t) & \quad \boldsymbol{w}_t = \boldsymbol{y}_t; \boldsymbol{\xi}_t \\
\boldsymbol{w}_{t|t} &= E_t(\boldsymbol{w}_t) & \\
\\
P_{t|t-1} &= E[(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1})(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1})'] & \quad MSE(\boldsymbol{\xi}_{t|t-1}) \\
P_{t|t} &= E[(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t})(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t})'] & \quad MSE(\boldsymbol{\xi}_{t|t}) \\
G_{t|t-1} &= E[(\boldsymbol{y}_t - \boldsymbol{y}_{t|t-1})(\boldsymbol{y}_t - \boldsymbol{y}_{t|t-1})'] & \quad MSE(\boldsymbol{y}_{t|t-1})
\end{aligned}
$$

Kalman filter recursions

Set-up

Recall:

$$\begin{cases} \underset{r\times 1}{\boldsymbol{\xi}_t} = F\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t & E(\boldsymbol{v}_t\,\boldsymbol{v}_t') = Q \\ \underset{n\times 1}{\boldsymbol{y}_t} = H\boldsymbol{\xi}_t + \boldsymbol{w}_t & E(\boldsymbol{w}_t\,\boldsymbol{w}_t') = R \end{cases}$$

hence

$$\begin{cases} \boldsymbol{\xi}_{t|t-1} = F\boldsymbol{\xi}_{t-1|t-1} \\ P_{t|t-1} = FP_{t-1|t-1}F' + Q \end{cases}$$

$$\begin{cases} \boldsymbol{y}_{t|t-1} = H\boldsymbol{\xi}_{t|t-1} \\ G_{t|t-1} = HP_{t|t-1}H' + R \end{cases}$$

so far, no surprises

Kalman filter recursions

Projection

Use Gaussian projection:

$$\begin{pmatrix} \mathbf{y}_t \\ \boldsymbol{\xi}_t \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{y}_{t|t-1} \\ \boldsymbol{\xi}_{t|t-1} \end{pmatrix}; \quad \begin{pmatrix} G_{t|t-1} & HP_{t|t-1} \\ P_{t|t-1}H' & P_{t|t-1} \end{pmatrix}\right]$$

we get:

$$\boldsymbol{\xi}_t = \boldsymbol{\xi}_{t|t-1} + P_{t|t-1}H'G^{-1}_{t|t-1}(\mathbf{y}_t - \mathbf{y}_{t|t-1}) + \mathbf{v}_t$$

from where:

$$\boldsymbol{\xi}_{t|t} = \boldsymbol{\xi}_{t|t-1} + P_{t|t-1}H'G^{-1}_{t|t-1}(\mathbf{y}_t - \mathbf{y}_{t|t-1}) \tag{1}$$

rule: when conditioning, smallest information set wins

The Kalman filter recursions

Equation (1) is the updating equation:

$$\boldsymbol{\xi}_{t|t} = \boldsymbol{\xi}_{t|t-1} + P_{t|t-1}H'G^{-1}_{t|t-1}(\boldsymbol{y}_t - \boldsymbol{y}_{t|t-1})$$

with conditional variance:

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H'G^{-1}_{t|t-1}HP_{t|t-1}$$

Remarks

- $\boldsymbol{y}_{t|t} = \boldsymbol{y}_t$ since $\boldsymbol{y}_t$ is observed $\rightarrow G_{t|t} = 0$
- for $t = 0$ "guess" $\boldsymbol{\xi}_{1|0}, P_{1|0}$ to generate $\boldsymbol{y}_{1|0}, G_{1|0}$
- for $t = 1$ use updating eq for $\boldsymbol{\xi}_{1|1}, P_{1|1}$ and then $\boldsymbol{\xi}_{2|1}, P_{2|1}, \boldsymbol{y}_{2|1}, G_{2|1}$
- keep iterating to get $\{\boldsymbol{\xi}_{t|t}\}_{t=1}^T, \{\boldsymbol{\xi}_{t|t-1}\}_{t=1}^T$

The Kalman filter recursions

Recap

from state-equation:

$$\boldsymbol{\xi}_{t|t-1} = F\boldsymbol{\xi}_{t-1|t-1}$$
$$P_{t|t-1} = FP_{t-1|t-1}F' + Q$$

from observation equation:

$$\boldsymbol{y}_{t|t-1} = H\boldsymbol{\xi}_{t|t-1}$$
$$G_{t|t-1} = HP_{t|t-1}H' + R$$

at time t we get $\boldsymbol{y}_t$. Hence the update is:

$$\boldsymbol{\xi}_{t|t} = \boldsymbol{\xi}_{t|t-1} + P_{t|t-1}H'G_{t|t-1}^{-1}(\boldsymbol{y}_t - \boldsymbol{y}_{t|t-1})$$
$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H'G_{t|t-1}^{-1}HP_{t|t-1}$$

The Kalman filter recursions

Key insight

recall:

$$\boldsymbol{\xi}_{t|t} = \boldsymbol{\xi}_{t|t-1} + P_{t|t-1}H'G_{t|t-1}^{-1}(\mathbf{y}_t - \mathbf{y}_{t|t-1})$$

$$\mathbf{y}_{t|t-1} = H\boldsymbol{\xi}_{t|t-1}$$

hence:

$$\boldsymbol{\xi}_{t+1|t} = F\boldsymbol{\xi}_{t|t} = F\left\{\boldsymbol{\xi}_{t|t-1} + P_{t|t-1}H'G_{t|t-1}^{-1}(\mathbf{y}_t - \mathbf{y}_{t|t-1})\right\}$$

$$\boldsymbol{\xi}_{t+1|t} = \underbrace{F\boldsymbol{\xi}_{t|t-1}}_{AR(1)} + \underbrace{FP_{t|t-1}H'G_{t|t-1}^{-1}}_{K_t}\underbrace{(\mathbf{y}_t - H\boldsymbol{\xi}_{t|t-1})}_{\text{forecast error}}$$

$$\boldsymbol{\xi}_{t+1|t} = F\boldsymbol{\xi}_{t|t-1} + K_t(\mathbf{y}_t - H\boldsymbol{\xi}_{t|t-1})$$

The Kalman recursions

Initialization

Suppose you knew F, H, Q + stationarity, then:

$$\underbrace{\boldsymbol{\xi}_t}_{E(\boldsymbol{\xi})} = F \underbrace{\boldsymbol{\xi}_{t-1}}_{E(\boldsymbol{\xi})} + \underbrace{\mathbf{v}_t}_{E(\mathbf{v}_t)=0} \rightarrow E(\boldsymbol{\xi}) = FE(\boldsymbol{\xi})$$

$$\rightarrow (I - F)E(\boldsymbol{\xi}) = 0 \rightarrow E(\boldsymbol{\xi}) = 0$$

Similarly:

$$\underbrace{E(\boldsymbol{\xi}_t \, \boldsymbol{\xi}_t')}_{\Sigma_\xi} = E[(F\boldsymbol{\xi}_{t-1} + \mathbf{v}_t)(F\boldsymbol{\xi}_{t-1} + \mathbf{v}_t)'] =$$

$$= F \underbrace{E(\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_{t-1}')}_{\Sigma_\xi} F' + \underbrace{E(\mathbf{v}_t \, \mathbf{v}_t')}_{Q} + \underbrace{\text{cross products}}_{=0}$$

$$\Sigma_\xi = F \, \Sigma_\xi F' + Q \rightarrow vec(\Sigma_\xi) = [I - (F \otimes F')]^{-1} vec(Q)$$

Recap so far

- Kalman filter uses Gaussian projection to break complex models into simpler recursive problems.
- Generalizations have been done along many dimensions:
  - time-varying parameters
  - non-Gaussian likelihood problems
  - nonlinear problems
- Next we will see how to construct likelihood
- Bayesian approach lends itself nicely: e.g., specify a prior for $\hat{\boldsymbol{\xi}}_{1|0}, \hat{P}_{1|0}$

Forecasting

Forecasting h-periods ahead

by recursive substitution on AR(1) for $\boldsymbol{\xi}_t$:

$$\boldsymbol{\xi}_{t+h} = F^h \boldsymbol{\xi}_t + F^{h-1}\mathbf{v}_{t+1} + F^{h-2}\mathbf{v}_{t+2} + \ldots + F\mathbf{v}_{t+h-1} + \mathbf{v}_{t+h}$$

hence:

$$E_t(\boldsymbol{\xi}_{t+h}) = \boldsymbol{\xi}_{t+h|t} = F^h \boldsymbol{\xi}_{t|t}$$

with forecast error:

$$\boldsymbol{\xi}_{t+h} - \boldsymbol{\xi}_{t+h|t} = F^h(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t}) + F^{h-1}\mathbf{v}_{t+1} + \ldots + \mathbf{v}_{t+h}$$

and MSE:

$$P_{t+h|t} = F^h P_{t|t}(F')^h + F^{h-1}Q(F')^{h-1} + \ldots + FQF' + Q$$

Forecasting h-periods ahead
Continued

bring state-variable forecast into observation equation:

$$\mathbf{y}_{t+h} = H\boldsymbol{\xi}_{t+h} + \mathbf{w}_{t+h}$$
$$\mathbf{y}_{t+h|t} = H\boldsymbol{\xi}_{t+h|t}$$

forecast error:

$$\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t} = H(\boldsymbol{\xi}_{t+h} - \boldsymbol{\xi}_{t+h|t}) + \mathbf{w}_{t+h}$$

MSE:

$$G_{t+h|t} = HP_{t+h|t}H' + R$$

Missing observations

Suppose $\boldsymbol{y}_s$ is missing from $\{\boldsymbol{y}_t\}_{t=1}^{T}$

Kalman filter offers a natural solution:
replace $\boldsymbol{y}_s$ with $H\boldsymbol{\xi}_{s|s-1}$

recall the Kalman recursion:

$$\boldsymbol{\xi}_{t+1|t} = F\boldsymbol{\xi}_{t|t-1} + K_t(\boldsymbol{y}_t - H\boldsymbol{\xi}_{t|t-1})$$

update the state at time s by simply setting:
$\boldsymbol{\xi}_{s|s} = \boldsymbol{\xi}_{s|s-1}$ and $P_{s|s} = P_{s|s-1}$ and hence:

$$\boldsymbol{\xi}_{s+1|s} = F\boldsymbol{\xi}_{s|s-1} + K_t(H\boldsymbol{\xi}_{s|s-1} - H\boldsymbol{\xi}_{s|s-1}) = F\boldsymbol{\xi}_{s|s-1}$$

Maximum Likelihood Estimation

Recall the MA(1) likelihood estimator

suppose the DGP is:

$$y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}; \qquad \epsilon_t \sim N(0, \sigma^2)$$

if ϵ_{t-1} were known, easy to set-up $y_t|\epsilon_{t-1} \sim N(\mu + \theta\epsilon_{t-1}, \sigma^2)$
supose you knew $\epsilon_0 = 0$, then $\epsilon_1 = y_1 - \mu$ and

$$f_{y_2|y_1,\epsilon_0=0}(y_2|y1, \epsilon_0 = 0; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_2 - \mu - \theta\epsilon_1)^2}{2\sigma^2}\right]$$

hence you could work your way through $t = 2, \ldots, T$:

$$\epsilon_t = (y_t - \mu) - \theta(y_{t-1} - \mu) - \ldots + (-1)^{t-1}\theta^{t-1}(y_1 - \mu) + (-1)^t\theta^t\epsilon_0$$

MA(1) MLE estimator

Remarks

- this delivers the conditional MLE (conditional on $\epsilon_0 = 0$). Depends on invertibility
- maximization of the likelihood requires numerical techniques (notice θ is raised to powers of t as we specify the likelihood for each observation in the sample
- exact likelihood can be constructed two ways: (1) take ϵ_0 as one more parameter to estimate; (2) write down exact joint likelihood

EM algorithm—an MA(1) example

suppose instead you had a "guess" $\{\epsilon_t^0\}_{t=1}^T \dots$
then life is simple: estimate by OLS

$$y_t = \mu^0 + \theta^0 \epsilon_{t-1}^0 + \epsilon_t \quad \rightarrow \quad \epsilon_t^1 = y_t - \hat{\mu}^0 - \hat{\theta}^0 \epsilon_{t-1}^0$$

rinse, and repeat until usual stopping rules for non-linear optimization:

1. $\hat{\mu}^j \approx \hat{\mu}^{j-1}$; $\hat{\theta}^j \approx \hat{\theta}^{j-1}$, or ...
2. $\left.\frac{\partial \mathcal{L}(\mu;\theta)}{\partial \mu}\right|_{\hat{\mu}^j} \approx 0$; $\left.\frac{\partial \mathcal{L}(\mu;\theta)}{\partial \mu}\right|_{\hat{\theta}^j} \approx 0$, or ...
3. $|\mathcal{L}(\hat{\mu}^j; \hat{\theta}^j) - \mathcal{L}(\hat{\mu}^{j-1}; \hat{\theta}^{j-1})| \approx 0$

MLE with the Kalman filter

The recursions

obs. eqn. $\begin{cases} \mathbf{y}_{t|t-1} & = H\boldsymbol{\xi}_{t|t-1} \\ G_{t|t-1} & = HP_{t|t-1}H' + R \end{cases}$

state eqn. $\begin{cases} \boldsymbol{\xi}_{t|t-1} & = F\boldsymbol{\xi}_{t-1|t-1} \\ P_{t|t-1} & = FP_{t-1|t-1}F' + Q \end{cases}$

updating equations:

$$\boldsymbol{\xi}_{t|t} = \boldsymbol{\xi}_{t|t-1} + P_{t|t-1}H'G_{t|t-1}^{-1}(\mathbf{y}_t - \mathbf{y}_{t|t-1})$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H'G_{t|t-1}^{-1}HP_{t|t-1}$$

Kalman filter MLE

Recursive formulation: like conditional MA(1) example

start with $\{\boldsymbol{y}_t\}_{t=1}^{T}$, and initial guess for $\boldsymbol{\xi}_{1|0}$ and $P_{1|0}$

$$\boldsymbol{y}_{t|t-1} \sim N(H\boldsymbol{\xi}_{t|t-1};\ G_{t|t-1})$$

$$f(\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\dots,\boldsymbol{\theta}) = (2\pi)^{-n/2}|G_{t|t-1}|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{y}_t - H\boldsymbol{\xi}_{t|t-1})' G_{t|t-1}^{-1}(\boldsymbol{y}_t - H\boldsymbol{\xi}_{t|t-1})\}$$

the log-likelihood hence becomes:

$$\mathcal{L}(\boldsymbol{\theta}) = \log f(\boldsymbol{y}_1;\boldsymbol{\theta}) + \sum_{t=1}^{T} \log f(\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\dots,\boldsymbol{\theta})$$

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{Tn}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|G_{t|t-1}| - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{y}_t - H\boldsymbol{\xi}_{t|t-1})' G_{t|t-1}^{-1}(\boldsymbol{y}_t - H\boldsymbol{\xi}_{t|t-1})$$

EM approach

Similar to MA(1) example

ingredients:

- $\{\mathbf{y}_t\}_{t=1}^T$ sample of observed data
- $\{\boldsymbol{\xi}_t^0\}_{t=1}^T$ an initial guess for $\boldsymbol{\xi}_t$

estimate by OLS:

$$\boldsymbol{\xi}_t^0 = F\boldsymbol{\xi}_{t-1}^0 + \mathbf{v}_t \rightarrow \hat{F}^0; \{\hat{\mathbf{v}}_t^0\}_{t=1}^T \rightarrow \hat{Q}^0$$

$$\mathbf{y}_t = H\boldsymbol{\xi}_t^0 + \mathbf{w}_t \rightarrow \hat{H}^0, \{\hat{\mathbf{w}}_t^0\}_{t=1}^T \rightarrow \hat{R}^0$$

with $\hat{F}^0$, $\hat{Q}^0$, $\hat{H}^0$, and $\hat{R}^0$ and Kalman recursions generate $\{\boldsymbol{\xi}_{t|t}^1\}$
rinse and repeat until convergence

EM algorithm references

- Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum Likelihood Estimation from Incomplete Data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39:1—38
- Shumway, R. H. and D.S. Stoffer. 1982. An Approach to Time Series Smoothing and Forecasting using the EM Algorithm. *Journal of Time Series Analysis*, 3(4):253–264
- Watson, M. W. and R. F. Engle. 1983. Alternative Algorithms for the Estimation of MIMIC and Varying Coefficient Models. *Journal of Econometrics*, 23(3): 385-400
- McLachlan G. J. and Krishnan T. 2007. *The EM Algorithm and Extensions—2nd ed.* John Wiley & Sons: Hoboken, NJ

Identification

For some models, different combinations of F, Q, H, R generate identical likelihood values. How can you tell?

- likelihood is flat at the max $\rightarrow$ poor convergence
- poorly identified parameters $\rightarrow$ near singular 2nd derivatives (similar to colinearity) $\rightarrow$ large S.E.s

Solution: no systematic ex-ante check available

use good judgement: more flexibility $\rightarrow$ lack of identification

Asymptotic properties of MLE

Typical conditions for identification:

- eigenvalues of F inside unit circle
- $\boldsymbol{\theta}_0$ not on the boundary of Θ

then:

$$\sqrt{T}\ \mathcal{I}_{2D,T}^{1/2}\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} N(0,1)$$

$$\mathcal{I}_{2D,T} = -\frac{1}{T}E\left[\sum_{t=1}^{T} \frac{\partial^2 \log f(\mathbf{y}_t|\mathbf{y}_{t-1},\ldots;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$$

$\hat{\mathcal{I}}_{2D,T}$ by evaluating at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_T$ since $\hat{\mathcal{I}}_{2D,T} \xrightarrow{p} \mathcal{I}_{2D,T}$

QMLE results

So far assumed $\boldsymbol{y}_t$ and $\boldsymbol{\xi}_t$ are jointly Gaussian

What if they are not? White (1982), Watson (1989):

$$\sqrt{T}\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} N\left(0,\ \left(\mathcal{I}_{2D,T}\,\mathcal{I}_{OP}^{-1}\,\mathcal{I}_{2D,T}\right)^{-1}\right)$$

$$\hat{\mathcal{I}_{OP}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{s}_t(\hat{\boldsymbol{\theta}})\boldsymbol{s}_t(\hat{\boldsymbol{\theta}})'; \quad s_t(\hat{\boldsymbol{\theta}}) = \frac{\partial \log f(\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\ldots,\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}$$

i.e. usual "sandwich" estimator of the covariance matrix

Kalman filter output

Steady-state Kalman filter

Propositions 13.1 + 13.2. Assume:

- F is $r \times r$ and eigenvalues inside unit circle
- H is arbitrary $n \times r$
- Q and R positive (semidefinite) symmetric
- $vec(P_{1|0}) = [I_{r^2} - F \otimes F]^{-1} vec(Q)$

then:

- $P_{t+1|t} \leq P_{t|t-1}$ and $P_{t+1|t} \rightarrow P$ as $T \rightarrow \infty$
- if either Q, R or both positive definite, then P unique
- the Kalman gain is such that $K_t \rightarrow K$ as $T \rightarrow \infty$
- eigenvalues of $(F - KH)$ inside unit circle (if P unique)

how is this useful...?

VAR(∞) representation

From before, as $T \to \infty$:

$$\boldsymbol{\xi}_{t+1|t} = F\boldsymbol{\xi}_{t|t-1} + K(\boldsymbol{y}_t - H\boldsymbol{\xi}_{t|t-1}) = F\,L\boldsymbol{\xi}_{t+|t} + K(\boldsymbol{y}_t - H\,L\boldsymbol{\xi}_{t+1|t})$$
$$\boldsymbol{\xi}_{t+1|t} = [I_r - (F - KH)\,L]^{-1}K\boldsymbol{y}_t$$

note: $E(\boldsymbol{y}_{t+1}|\boldsymbol{y}_t, \ldots) = H\boldsymbol{\xi}_{t+1|t} = H[I_r - (F - KH)\,L]^{-1}K\boldsymbol{y}_t$ define:
$\boldsymbol{\epsilon}_{t+1} \equiv \boldsymbol{y}_{t+1} - E(\boldsymbol{y}_{t+1}|y_t, \ldots)$

VAR(∞) representation is easily seen to be:

$$\boldsymbol{y}_{t+1} = H[I_r - (F - KH)\,L]^{-1}K\boldsymbol{y}_t + \boldsymbol{\epsilon}_{t+1}$$
$$\boldsymbol{y}_{t+1} = H(F - KH)K\boldsymbol{y}_t + H(F - KH)^2K\boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\epsilon}_{t+1}$$
$$E(\boldsymbol{\epsilon}_{t+1}\,\boldsymbol{\epsilon}'_{t+1}) = HPH' + R$$

Wold representation: MA(∞)

Wold thm: every C-S process has MA(∞) representation

invert VAR(∞) since (F - KH) invertible:

$$\mathbf{y}_{t+1} = \left\{I_n - H[I_r - (F - KH)L]^{-1} K\, L\right\}^{-1} \boldsymbol{\epsilon}_{t+1}$$

can show (see Hamliton p. 393) that:

$$\mathbf{y}_{t+1} = \left\{I_n + H[I - F\, L]^{-1} K\, L\right\} \boldsymbol{\epsilon}_{t+1}$$
$$\mathbf{y}_{t+1} = \boldsymbol{\epsilon}_{t+1} - HK\boldsymbol{\epsilon}_t - HFK\boldsymbol{\epsilon}_{t-1} - HF^2K\boldsymbol{\epsilon}_{t-2} - \dots$$

useful to find impulse responses for models

Smoothed inference

when ξ_t is itself of interest

e.g. $\boldsymbol{\xi}_t$ is natural rate of interest

idea: use the entire sample for best estimate $\boldsymbol{\xi}_{t|T}$

1. Run KF $\rightarrow \{P_{t|t}\}_{t=1}^T$; $\{P_{t|t-1}\}_{t=1}^T$; $\{\boldsymbol{\xi}_{t|t}\}$; $\{\boldsymbol{\xi}_{t|t-1}\}_{t=1}^T$
2. Work backwards from $\boldsymbol{\xi}_{T|T}$ as follows:

$$
\begin{aligned}
\boldsymbol{\xi}_{t|T} &= \boldsymbol{\xi}_{t|t} + J_t(\boldsymbol{\xi}_{t+1|T} - \boldsymbol{\xi}_{t+1|t}) \\
J_t &= P_{t|t} F' P_{t+1|t}^{-1} \\
P_{t|T} &= P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t}) J_t'
\end{aligned}
$$

for $t = T-1, T-2, \ldots$

see Hamilton p. 394

Confidence intervals for $\xi_{t|T}$

Let $\xi_{t|T}(\hat{\boldsymbol{\theta}}_T)$ denote the best estimate from:

1. MLE estimates of $\hat{\boldsymbol{\theta}}_T$ using state-space
2. using backwards filter to obtain $\xi_{t|T}(\hat{\boldsymbol{\theta}}_T)$

Note (Hamilton, 1986):

$$E[(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|T}(\hat{\boldsymbol{\theta}}_T))(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|T}(\hat{\boldsymbol{\theta}}_T))'] = \underbrace{E[(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|T}(\boldsymbol{\theta}_0))(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|T}(\boldsymbol{\theta}_0))']}_{\text{filter uncertainty}} + \underbrace{E[(\boldsymbol{\xi}_{t|T}(\boldsymbol{\theta}_0) - \boldsymbol{\xi}_{t|T}(\hat{\boldsymbol{\theta}}_T)(\boldsymbol{\xi}_{t|T}(\boldsymbol{\theta}_0) - \boldsymbol{\xi}_{t|T}(\hat{\boldsymbol{\theta}}_T)']}_{\text{parameter uncertainty}}$$

in practice use Monte Carlo draws or Bayesian MCMC methods

$r^\star$ in the HLW model

The model

$$
\begin{array}{rcll}
\tilde{y}_t &=& y_t - y_t^\star & \text{output gap} \\
y_t^\star &=& y_{t-1}^\star + g_{t-1} + \epsilon_{y,t}^\star & \text{potential output} \\
g_t &=& g_{t-1} + \epsilon_{g,t} & \text{growth of potential} \\
\tilde{y}_t &=& \alpha_1^y \tilde{y}_{t-1} + \alpha_2^y \tilde{y}_{t-2} - \gamma(r_{t-1} - r_{t-1}^\star) + \epsilon_t^{IS} & \text{IS curve} \\
\pi_t &=& \alpha_\pi \pi_{t-1} + (1-\alpha_\pi)\frac{\sum_2^4 \pi_{t-j}}{3} + \kappa \tilde{y}_{t-1} + \epsilon_{\pi,t} & \text{Phillips curve} \\
r_t^\star &=& 4g_t + z_t & \text{r-star equation} \\
z_t &=& z_{t-1} + \epsilon_{z,t} & \text{unobserved factors}
\end{array}
$$

observations: $\boldsymbol{y}_t = (y_t,\ r_t,\ \pi_t)'$

states: $\boldsymbol{\xi}_t = (y_t^\star,\ \tilde{y}_t,\ g_t,\ r_t^\star,\ z_t)'$

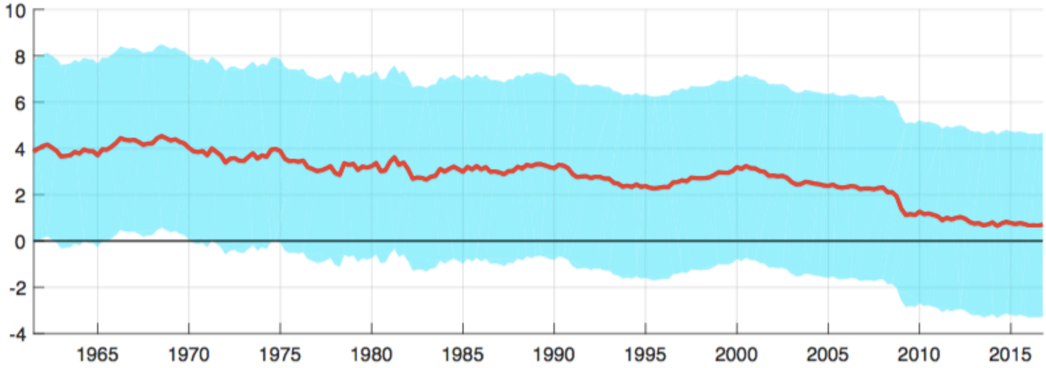
errors: $\boldsymbol{v}_t = (\epsilon_t^\star,\ \epsilon_{g,t},\ \epsilon_{z,t})'$; $\boldsymbol{w}_t = (\epsilon_t^{IS},\ \epsilon_{\pi,t})'$ serially and mutually uncorrelated

parameters: $\alpha_1^y,\ \alpha_2^y,\ \gamma,\ \alpha_\pi,\ \kappa$ and $\sigma_y^2,\ \sigma_g^2,\ \sigma_{IS}^2,\ \sigma_\pi^2,\ \sigma_z^2$

Confidence intervals for $\xi_{t|T}$ in HLW model

Estimates of $r^{\star}$

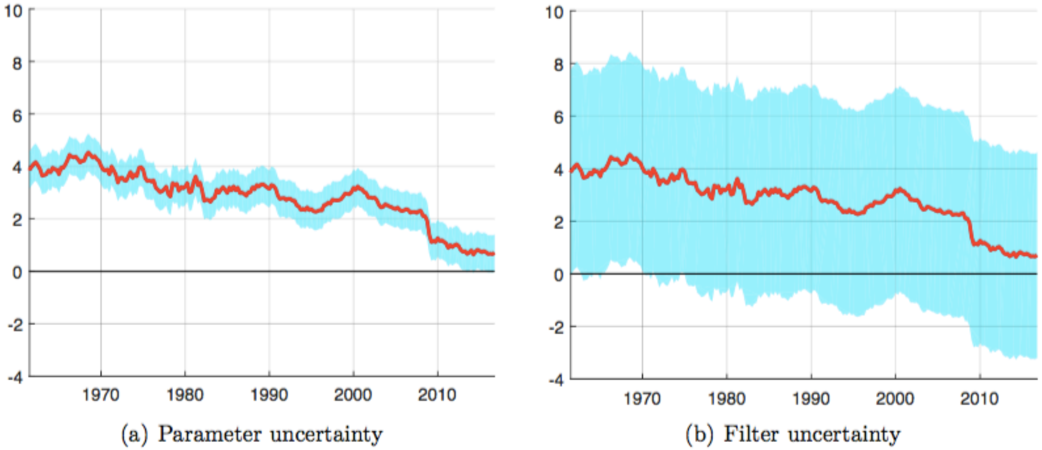
Figure 1: U.S. natural rate of the HLW model: median estimates and 90% bands



Notes: quarterly data, 1961Q2-2016Q3. Estimated r^{*} joint with 90% confidence bands. Bands reflect both filter and parameter uncertainty. Bands are computed using the Hamilton's (1986) approach with 2000 Montecarlo replications of the parameter vector.

Filter vs. parameter confidence intervals for $r^\star$ in HLW model

Figure 2: Filter and parameter uncertainty in the HLW model



Notes: quarterly data, 1961Q2-2016Q3. r^* joint with 90% confidence bands. Bands in the left-hand side chart reflect parameter uncertainty; bands in the right-hand side reflect filter uncertainty.

STATA example: simplified HLW

HLW_example.do

$$
\begin{aligned}
\tilde{y}_t &= y_t - y_t^\star && \text{output gap} \\
y_t^\star &= y_{t-1}^\star + g_{t-1} + \epsilon_{y,t}^\star && \text{potential output} \\
g_t &= g_{t-1} + \epsilon_{g,t} && \text{growth of potential} \\
\tilde{y}_t &= 0.75^y \tilde{y}_{t-1} - \gamma(r_{t-1} - r_{t-1}^\star) + \epsilon_t^{IS} && \text{IS curve} \\
\pi_t &= 0.95\pi_{t-1} + (1 - 0.95)\pi_t^* + \kappa \tilde{y}_{t-1} + \epsilon_{\pi,t} && \text{Phillips curve} \\
i_t &= r_t^* + \pi_t^* + 0.75 i_{t-1} + \epsilon_t^i && \text{nominal interest rate} \\
r_t^\star &= g_t + z_t && \text{r-star equation} \\
z_t &= z_{t-1} + \epsilon_{z,t} && \text{unobserved factors}
\end{aligned}
$$

Fama and Gibbons 1982

Definitions:

- i_t: 3-m T-Bill for month 3 of quarter t as annual rate
- π_t inflation between month 3 of quarter t and $t+1$, as $400 \log CPI$
- $y_t = i_t - \pi_t$, ex-post real rate

State-space:

$$
\begin{aligned}
y_t &= \mu + \xi_t + w_t \\
\xi_t &= f\,\xi_{t-1} + v_t
\end{aligned}
$$

ex-ante real rate: $\hat{r}^{\star}_t = i_t - \hat{\pi}^e_t = \hat{\mu} + \xi_{t|T}$

- estimate $\hat{r}^{\star}_t$
- compare estimates with HLW - why the difference?
- compute error bands for $\hat{r}^{\star}_t$

Observability

Given the state-space:

$$\begin{cases} \underset{r\times 1}{\boldsymbol{\xi}_t} & = \underset{r\times r}{F}\,\underset{r\times 1}{\boldsymbol{\xi}_{t-1}} + \underset{r\times 1}{\mathbf{v}_t}\,; \quad E(\mathbf{v}_t\,\mathbf{v}_t') = Q \\ \underset{n\times 1}{\mathbf{y}_t} & = \underset{n\times r}{H}\,\underset{r\times 1}{\boldsymbol{\xi}_t} + \underset{n\times 1}{\mathbf{w}_t}\,; \quad E(\mathbf{w}_t\,\mathbf{w}_t') = R \end{cases}$$

observability: when can we learn about the dynamics of the states from the observables and disturbances?

$$B = \begin{bmatrix} H \\ H\,F \\ H\,F^2 \\ \vdots \\ H\,F^{r-1} \end{bmatrix} \quad \text{observability requires:} rank(B) = r = dim(\boldsymbol{\xi}_t)$$

if $rank(B) \leq r$ then states not well identified
(wide confidence bands for states)

Observability

An application to HLW

Examine the *observability* of HLW when:

1. IS curve is flat (i.e. $\gamma = 0$)
2. Phillips curve is flat (i.e. $\kappa = 0$)

note:

$$B = \begin{bmatrix} 1-\alpha_y & 1+4\gamma & \gamma \\ -\kappa & 0 & 0 \\ 1-\alpha_y & 2+4\gamma-\alpha_y & \gamma \\ -\kappa & -\kappa & 0 \\ 1-\alpha_y & 3+4\gamma-2\alpha_y & \gamma \\ -\kappa & -2\kappa & 0 \end{bmatrix}$$

check $rank(B) = 3$