A Practical Guide to State Space Modeling

Jin-Lung Lin

Institute of Economics, Academia Sinica
Department of Economics, National Chengchi University

March 2006

1 Introduction

State Space Model (SSM) has been a very powerful framework for the analysis of dynamical systems. While linear regression models use exogenous variables to distinguish the explained variation from the unexplained variation, SSM relies the dynamics of the state variables and the linkage between the observed variables and state variables to draw statistical inference about the unobserved states. The complete path of its conditional mean and variance are the output of Kalman Filtering.

SSM is particularly useful for models involving unobserved variables. Potential output, natural rate of unemployment, and common business are three handy examples in economics. Furthermore, the conventional univariate as well multivariate ARMA models, linear regression models, and spline models can be converted into a SSM where forecasting, missing value and testing structural breaks can be easily handled.

The theory of SSM was proposed in 1960's and been heavily used by economists and other social scientists for a long time. Yet, a general and easy-to-use statistical software has not been around until recently. *SsfPack* for *Ox*, to my mind, is the best software for SSM. In this note, I shall review the state space models, the Kalman Filtering, smoothing, forecasting and initialization issues. Then, a brief introduction of the *SsfPack* and *Ox* will be given. Application examples includes local trend models, airline model, structural break tests, spline, missing observations, and seasonal adjustment. Finally, I shall discuss various SSM models to estimate the potential GDP and/or NAIRU in Taiwan.

In addition to this introduction, Section 2 list an easy but important lemma for deriving the Kalman Filter. The Kalman Filter for the local level model is discussed in details in Section 3 and Section 4 summarizes the recursion equation for the general SSM. The SsfPack implementation is given in Section 5 and applications in Section 6. Section 7 concludes.

2 An important lemma for deriving the Kalman Filter

Let $(x, y, z)'$ are jointly normally distributed with $\mu_z = 0$, and $\Sigma_{yz} = 0$. Then

$$\begin{aligned} E(x|y,z) &= E(x|y) + \Sigma_{xz}\Sigma_{zz}^{-1}z \\ Var(x|y,z) &= Var(x|y) - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}' \end{aligned}$$

The proof of the lemma is straightforward and thus omitted. It reads as below. When a new independent piece of information is added, the conditional mean and variance can be obtained by updating the previous ones. This formula is of fundamental importance deriving the Kalman Filter.

3 Kalman Filtering for local level model

The local level model described below has a simple structure and serves as an excellent framework for understanding the recursion mechanism of the Kamlan Filter.

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2) & (1) \\ \alpha_{t+1} &= \alpha_t + \eta_t, \quad \eta_t \sim N(0, \eta_t^2) & (2) \\ \alpha_1 &\sim N(a_1, P_1) & (3) \end{aligned}$$

In this simple model y_t consists of a signal, α_t, measuring the stochastic trend and a measurement error, ε_t. While exogenous variables are brought in to discriminate the signal from the noise in linear regression analysis, it is the dynamics, that does the job in state space model. In other words, the different dynamics for signals and noise which latter is usually assumed to follow a white noise process enables us to decompose the observed variable into two parts: the signal related to the state variables and disturbance terms. Kalman Filter fully explores this dynamic structure for filtering, smoothing and forecasting. When up to current period $(y_1, \ldots, y_t)$, last period $(y_1, \ldots, y_{t-1})$ and whole sample $(y_1, \ldots, y_n)$ information are used to estimate state variables at current time $t, (\alpha_t)$, they are called filtering, forecasting and smoothing respectively.

3.1 Kalman filtering

Let $\boldsymbol{Y_t} = \sigma\{y_1, \cdots, y_t\}$ and defined

$$\begin{aligned} a_{t+1} &= E(\alpha_{t+1}|\boldsymbol{Y}_t), \\ P_{t+1} &= Var(\alpha_{t+1}|\boldsymbol{Y}_t). \end{aligned}$$

then from (1, 2), we have

$$\begin{aligned} a_{t+1} &= E(\alpha_t|\boldsymbol{y}_t) \\ P_{t+1} &= Var(\alpha_t|\boldsymbol{y}_t) + \sigma_\eta^2 \end{aligned}$$

Define $v_t = y_t - a_t$ and $F_t = Var(v_t)$. Then $E(v_t|\boldsymbol{Y}_{t-1}) = 0$ or v_t is a martingale difference process. So

$$\begin{aligned} E(v_t) &= E(E(v_t|\boldsymbol{Y}_{t-1})) = 0 \\ Cov(v_t, y_j) &= E(v_t y_j) = E(E(v_t|y_{t-1})y_j) = 0, \text{ for } j = 1, 2, \cdots, t-1. \end{aligned}$$

By the lemma,

$$\begin{aligned} E(\alpha_t|\boldsymbol{Y}_t) &= E(\alpha_t|\boldsymbol{Y}_{t-1}, v_t) \\ &= E(\alpha_t|\boldsymbol{Y}_{t-1}) + Cov(\alpha_t, v_t)Var(v_t)^{-1}v_t \end{aligned}$$

where

$$\begin{aligned} Cov(\alpha_t, v_t) &= E(\alpha_t(y_t - a_t)) \\ &= E(\alpha_t(\alpha_t + \varepsilon_t - a_t)) \\ &= E(\alpha_t(\alpha_t - a_t)) \\ &= Var(\alpha_t|\boldsymbol{Y}_{t-1}) \\ &= P_t \end{aligned}$$

Since

$$Var(v_t) = F_t = P_t + \sigma_\varepsilon^2$$

we have

$$E(\alpha_t|\boldsymbol{Y}_t) = a_t + K_t v_t$$

where

$$K_t = \frac{P_t}{F_t} = \frac{P_t}{P_t + \sigma_\epsilon^2}$$

Further,

$$\begin{aligned} Var(\alpha_t|\boldsymbol{Y}_t) &= Var(\alpha_t|\boldsymbol{Y}_{t-1}, v_t) \\ &= Var(\alpha_t|\boldsymbol{Y}_{t-1}) - Cov(\alpha_t, v_t)^2 Var(v_t)^{-1} \\ &= P_t - \frac{P_t^2}{F_t} \\ &= P_t(1 - K_t) \end{aligned}$$

Eqs. (4) and (4) say that the forecast error, v_t is used to update the estimate of mean and variance of α_5 when it becomes available. The complete updating equations are:

$$
\begin{aligned}
v_t &= y_t - a_t \\
F_t &= P_t + \sigma_\varepsilon^2 \\
a_{t+1} &= a_t + K_t v_t \\
P_{t+1} &= P_t(1 - K_t) + \sigma_\eta^2 \\
K_t &= \frac{P_t}{F_t}
\end{aligned}
$$

3.2 Initial conditions

Initial condition has to be given to complete the recursion and diffuse prior is the most commonly used one. Let a_1 be any number, and $P_1 \to \infty$ as in $\alpha_1 \sim N(a_1, P_1)$. In other words, we do not have any information on α_1. It is easy to see:

$$
\begin{aligned}
a_2 &= a_1 + \frac{P_1}{P_1 + \sigma_\varepsilon^2}(y_1 - a_1) \to y_1 \\
P_2 &= \frac{P_1 \sigma_\varepsilon^2}{P_1 + \sigma_\varepsilon^2} + \sigma_\eta^2 \to \sigma_\varepsilon^2 + \sigma_\eta^2
\end{aligned}
$$

The diffuse prior is equivalent to starting the recursion from $t = 2$ and using y_1 as initial conditions.

3.3 Kalman smoothing

We now turn to state smoothing. Let $\mathbf{y} = \sigma(y_1, \cdots, y_n)$. That is, y consists the whole sample information.

$$
\begin{aligned}
\hat{a}_t &= E(\alpha_t | y) \\
&= E(\alpha_t | \boldsymbol{Y}_{t-1}, v_t, \cdots, v_n) \\
&= a_t + P_t r_{t-1}
\end{aligned}
$$

where

$$
\begin{aligned}
r_{t-1} &= \frac{v_t}{F_t} + L_t r_t \\
L_t &= 1 - K_t = \frac{\sigma_\epsilon^2}{F_t}
\end{aligned}
$$

with $r_n = 0$

For the conditional variance,

$$\begin{aligned} V_t &= Var(\alpha_t|y) \\ &= Var(\alpha_t|\boldsymbol{Y}_{t-1}, v_t, \cdots, v_n) \\ &= P_t - P_t^2 N_{t-1} \end{aligned}$$

where $N_{t-1} = \frac{1}{F_t} + L_t^2 N_t$ with $N_n = 0$.

The Kalman Filter can be summarized as below. Starting with initial condition $(a_1, P_1), (a_2, P_2, F_2)$ are computed. When y_2 becomes available, $v_2 = y_2 - a_2$ is used to update the estimate of conditional mean and variance of α_2. Also, a_3, P_3 are computed and then updated when y_3 comes in. The process is repeated until at the end of the sample period. a_n, P_n, F_n are computed and updated. Now, using the smoothing recursion, $\hat{a}_{n-1}, V_{n-1}$ are computed. With the latter, $\hat{a}_{n-2}, V_{n-2}$ can be computed. The process is repeated until at the beginning of the period and $\hat{a}_1, V_1$ are computed.

3.4 Missing observations

Missing observations can be easily handled. Let $y_t,\ j = \tau, \cdots, \tau^* - 1$, are missing. Then for $t = \tau, \cdots, \tau^* - 1$

$$\begin{aligned} E(\alpha_t|\boldsymbol{Y}_{t-1}) &= E(\alpha_t|\boldsymbol{Y}_{\tau-1}) \\ &= E(\alpha_i + \sum_{j=i}^{t-1} \eta_\tau|\boldsymbol{Y}_{\tau-1}) \\ &= a_\tau \\ Var(\alpha_t|\boldsymbol{Y}_{t-1}) &= Var(\alpha_t|\boldsymbol{Y}_{\tau-1}) \\ &= Var(\alpha_\tau + \sum_{j=\tau}^{t-1} \eta_j|\boldsymbol{Y}_{\tau-1}) \\ &= P_\tau + (t-\tau)\sigma_\eta^2 \end{aligned}$$

As no new information is available during the missing periods, $v_t = 0$, conditional mean remains unchanged and conditional variance increases linearly with time.

3.5 Forecasting

Forecasting future values, $y_{n+1}, \ldots, y_{n+h}$ is equivalent to treating $y_{n+1}, \cdots, y_{n+h}$ as missing and using the Kalman Filtering.

3.6 Likelihood function and parameter estimation

Log-likelihood can be obtained as a side product of Kalman filtering. Since $P(y_1, \cdots, y_t) = P(y_t|\boldsymbol{Y}_{t-1})P(\boldsymbol{Y}_{t-1})$, we have

$$\begin{aligned} P(y) &= P(y_1, \cdots, y_t) \\ &= \Pi_{t=1}^{n} P(y_t|\boldsymbol{Y}_{t-1}) \end{aligned}$$

with $P(y_1|y_0) = P(y_1)$ and

$$\begin{aligned} logL &= logP(y) \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}(\log F_t + \frac{v_t^2}{F_t}) \end{aligned}$$

To achieve computation efficiency, we can substitute $(\sigma_\eta^2 = q\sigma_\varepsilon^2)$ and the system becomes

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \\ \alpha_{t+1} &= \alpha_t + \eta_t,\ \eta(t) \sim N(0, q\sigma_\varepsilon^2) \end{aligned}$$

The concentrated log likelihood becomes

$$\begin{aligned} P_t^* &= \frac{P_t}{\sigma_\varepsilon^2},\ F_t^* = \frac{F_t}{\sigma_\varepsilon^2} \\ v_t &= y_t - a_t,\ F_t^* = P_t^* + 1 \\ a_{t+1} &= a_t + K_t v_t,\ P_{t+1}^* = P_t^*(1 - K_t) + q \\ K_t &= \frac{P_t^*}{F_t^*} = \frac{P_t}{F_t} \\ logL_d &= -\frac{n}{2}log(2\pi) - \frac{n-1}{2}log\sigma_\varepsilon^2 - \frac{1}{2}\sum_{t=1}^{n}(logF_t^* + \frac{v_t^2}{\sigma_\varepsilon^2 F_t^*}) \\ \hat{\sigma}_\varepsilon^2 &= \frac{1}{n-1}\sum_{t=2}^{n}\frac{v_t^2}{F_t^*} \\ logL_{dc} &= -\frac{n}{2}log(2\pi) - \frac{n-1}{2} - \frac{n-1}{2}log\hat{\sigma}_\varepsilon^2 - \frac{1}{2}\sum_{t=2}^{n}logF_t^* \end{aligned}$$

3.7 Diagnostic checking

Diagnostic checking is based upon the assumption that $v_t \sim iid.N(0, F_t)$. Thus,

$$e_t = \frac{v_t}{\sqrt{F_t}} \sim N(0,1)$$

Normality can be checked by examining the skewness and kurtosis Normality: skewness and kurtosis

$$
\begin{aligned}
S &= \frac{m_3}{\sqrt{m_2^3}} \sim N(0, \frac{6}{n}) \\
K &= \frac{m_4}{m_2^2} \sim N(3, \frac{24}{n}) \\
N &= \{\frac{S^2}{6} + \frac{(K-3)^2}{24}\} \sim \chi^2(2)
\end{aligned}
$$

The Ljung-Box Q statistics is useful for checking serial correlation and there are a bundle of statistics for heteroscedasticity test.

4 Kalman Filtering for general models

The general state space model can be written as:

$$
\begin{aligned}
y_t &= Z_t \alpha_t + \varepsilon_t,\ \varepsilon_t \sim N(0, H_t) \\
\alpha_{t+1} &= T_t \alpha_t + R_t \eta_t,\ \eta_t \sim N(0, Q_t) \\
\alpha_1 &\sim N(a_1, P_1)
\end{aligned}
$$

Table 1: Dimensions for state space models

vector		matrix	
y_t	$p \times 1$	Z_t	$p \times m$
α_t	$m \times 1$	T_t	$m \times m$
ε_t	$p \times 1$	H_t	$p \times p$
η_t	$r \times 1$	R_t	$m \times r$
a_1	$m \times 1$	Q_t	$r \times r$
v_t	$p \times 1$	P_1	$m \times m$
a_t	$m \times 1$	P_t	$m \times m$

4.1 Initial conditions

Diffuse prior

$$P = P_* + \kappa P_\infty$$

P_∞ is $m \times m$ matrix with zeros and ones. κ is large, say $\kappa = 10^1 0$.

4.2 Filter recursion

As before, define

$$\begin{aligned} a_{t+1} &= E(\alpha_{t+1} | \boldsymbol{Y}_t) \\ P_{t+1} &= Var(\alpha_{t+1} | y_t) \\ v_t &= y_t - Z_t a_t \end{aligned}$$

The recursions become

$$\begin{aligned} F_t &= Z_t P_t Z_t^{'} + H_t \\ K_t &= T_t P_t Z_t^{'} F_t^{-1} \\ L_t &= T_t - K_t Z_t \\ a_{t+1} &= T_t a_t + K_t v_t \\ P_{t+1} &= T_t P_t L_t^{'} + P_t Q_t P_t^{'} \end{aligned}$$

Denote $a_{t|t} = E(\alpha_t | \boldsymbol{Y}_t), P_{t|t} = Var(\alpha_t | \boldsymbol{Y}_t)$, we have

$$a_{t|t} = a_t + M_t F_t^{-1} v_t$$

$$
\begin{aligned}
a_{t+1} &= T_t a_{t|t} \\
F_t &= Z_t P_t Z_t^{'} + H_t \\
M_t &= P_t Z_t^{'} \\
P_{t|t} &= P_t - M_t F_t^{-1} M_t^{'} \\
P_{t+1} &= T_t P_{t|t} T_t^{'} + R_t Q_t R_t^{'}
\end{aligned}
$$

4.3 State smoothing

Let $y = \sigma(y_1, \ldots, y_n)$. Then,

$$
\begin{aligned}
\hat{\alpha}_t &= E(\alpha_t|y) \\
&= a_{t|t} + P_{t|t} T_t^{'} P_{t+1}^{-1} (\hat{\alpha}_{t+1} - a_{t+1}) \\
&= a_t + P_t Z_t^{'} F_t^{-1} v_t + P_t L_t^{'} P_{t+1}^{-1} (\hat{\alpha}_{t+1} - a_{t+1})
\end{aligned}
$$

Let

$$
r_t = P_{t+1}^{-1} (\hat{\alpha}_{t+1} - a_{t+1})
$$

then

$$
r_{t-1} = Z_t^{'} F_t^{-1} v_t + L_t^{'} r_t
$$

with $r_n = 0$. For the variance,

$$
\begin{aligned}
V_t &= Var(\alpha_t|y) \\
&= P_t - P_t N_{t-1} P_t \\
N_{t-1} &= Z_t^{'} F_t^{-1} Z_t + L_t^{'} N_t L_t \\
r_{t-1} &= Z_t^{'} F_t^{-1} v_t + L_t^{'} t_t \\
\hat{\alpha}_t &= a_t + P_t r_{t-1}
\end{aligned}
$$

5 SsfPack notations

$$
\begin{aligned}
\alpha_{t+1} &= d_t + T_t \alpha_t + H_t \varepsilon t \\
Q_t &= c_t + Z_t \alpha_t
\end{aligned}
$$

$$
\begin{aligned}
y_t &= Q_t + G_t \varepsilon_t \\
\begin{pmatrix} a_{t+1} \\ y_t \end{pmatrix} &= \delta_t + \Phi_t \alpha_t + u_t \\
\delta_t &= \begin{pmatrix} d_t \\ c_t \end{pmatrix} \quad u_t = \begin{pmatrix} H_t \\ G_t \end{pmatrix} \varepsilon_t \\
\Omega_t &= \begin{bmatrix} H_t H_t' & H_t G_t' \\ G_t H_t' & G_t G_t' \end{bmatrix} \quad \Phi_t = \begin{bmatrix} T_t \\ Z_t \end{bmatrix} \\
\alpha_1 &\sim N(a_1, P_1) \quad \Sigma = \begin{bmatrix} P_1 \\ a_1' \end{bmatrix}
\end{aligned}
$$

Table 2: Dimension of SSM matrices

$\alpha_{t+1}, d_t, a : m \times 1,$	$y_t, Q_t, c_t : N \times 1$
$T_t, P : m \times m,$	$Z_t : N \times m$
$H_t : m \times r,$	$G_t : N \times r$
$\Phi : (m+N) \times m,$	$\delta : (m+N) \times 1$
$\Omega : (m+N) \times (m+N),$	$\Sigma : (m+1) \times 1$

Four possible model specifications:

nPhi	mOmega						
mPhi	mOmega	mSigma					
mPhi	mOmega	mSigma	mDelta				
mPhi	mOmega	mSigma	mDelta	mJ-Phi	mJ-Omega	mJ-Delta	mXt

where the last specification works for time-varying Z_t and T_t with mJ-Phi, mJ-Omega, mJ-Delta giving the corresponding row numbers in mXt, the data matrix. What is more, exogenous regressors can be added to the model. Below are summarized modules in *SsfPack*.

Model in static space form

AddSsfreg:	add regressor effect
GetSsfARMA:	put ARMA in static space.
GetSsfspline:	put regression in state space
GetSsfstsm:	put cubic spline in state space
SsfCombine:	combine two model
SsfCombineSym:	Combine two symmetric models

General state space algorithm

KalmanFil:	Kalman Filter
KalmanSmo:	Smoothing
SimSmoDraw:	simulation smoother
SimSmoWgt:	Covariance output of simulation smoother

Ready-to-use Function

SsfConDens:	mean of a draw of conditional density
SsfLik:	log-likelihood function
SsfLikeConc:	Profile log-likelihood function
SsfLikeSco:	Score vector
SsfMomentEst:	predictor, forecasting and smoothing
SsfRecursion:	State space recursion

6 Applications:estimating potential GPP and/or NAIRU

Potential output and non-accelerating inflation rate unemployment (NAIRU) are defined as the level of output and unemployment rate consistent with a stable rate of inflation. Here, I discuss three models to estimate potential output and/or NAIRU. They are Watson (1986), Kuttner (1994), and Apel & Jansson(1998).

Watson (1986)

$$\begin{aligned} y_t &= y_t^p + z_t \\ y_t^p &= y_{t-1}^p + \mu_y + e_{yt},\ e_{pt} \sim N(0, \sigma_p^2) \\ z_t &= \phi_1 z_{t-1} + \phi_2 z_{t-2} + e_{zt},\ e_{zt} \sim N(0, \sigma_z^2) \end{aligned}$$

where y_t: observed output, and y_t^p: potential GDP. In term of *Ox* notations. the model is :

$$\begin{pmatrix} y_{t+1}^p \\ z_{t+1} \\ z_t \\ y_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_t^p \\ z_t \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} e_{pt} \\ e_{zt} \end{pmatrix}$$

$$\delta = \begin{pmatrix} \mu_y \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Phi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

$$
\Omega = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
$$

Kuttner (1994)

$$
\begin{aligned}
y_t^p &= y_{t-1}^p + \mu_y + e_{yt} \\
y_t &= y_t^p + z_t \\
\Delta\pi_t &= \mu_\pi + r\Delta y_{t-1} + \beta z_{t-1} + v_t + \delta_1 v_{t-1} + \delta_2 v_{t-2} + \delta_3 v_{t-3}
\end{aligned}
$$

Apel and Jansson (1999)

$$
\begin{aligned}
y_t^p &= \alpha + y_{t-1}^p + \varepsilon_t^p \\
u_t^n &= u_{t-1}^n + \varepsilon_t^n \\
y_t &= y_t^p + \phi_0(u_t - u_t^n) + \phi_1(u_{t-1} - u_{t-1}^n) + \varepsilon_t^y \\
u_t - u_t^n &= \delta(u_{t-1} - u_{t-1}^n) + \varepsilon_t^c \\
\Delta\pi_t &= \rho_1\pi_{t-1} + \eta_1(u_t - u_t^n) + \omega X_t + \varepsilon_\tau
\end{aligned}
$$

7 Conclusions

State space models are very useful in econometric modeling, especially when unobserved variables are involved. SsfPack is a very general, easy-to-use and efficient package for SSM computation.

References

Durbin, J. and S.J. Koopman (2001), *Time Series Analysis by State Space Methods*, New York: Oxford University Press

Koopman, S.J., N. Shephard, and J. A. Doornik (1998), ”Statistical algorithms for models in state space model using SsfPack 2.2,” *Econometric Journals*, 2, 113-66.

- *Ox* is an object-oriented statistical system. Console Ox (no graphics) can be obtained freely via
  http://www.doornik.com
- *SsfPack* is a suite of C routines for carrying out computations involving the statistical analysis of univariate and multivariate models in state space form. The fully implemented link is to Ox and can be obtained freely at
  http://www.ssfpack.com
- *GnuDraw* is an Ox package meant for creating GnuPlot graphics. Both can be obtained at
  http://www.tinbergen.nl/ cbos/index.html?content=/ cbos/gnudraw.html
- Professional Ox is bundled of Console *Ox* and *GiveWin* provides complete support for graphical environment. It can purchased at
  www.doornik.com