State-Space Models

Overview

1. State-space models (*a.k.a.*, dynamic linear models, DLM)
2. Regression Examples
3. AR, MA and ARMA models in state-space form

See S&S Chapter 6, which emphasizes fitting state-space models to data via the Kalman filter.

State-space models

Linear filtering The observed data $\{X_t\}$ is the output of a linear filter driven by white noise, $X_t = \sum \psi_j w_{t-j}$. This perspective leads to parametric ARMA models as concise expressions for the lag weights as functions of the underlying ARMA parameters, $\psi_j = \psi_j(\phi, \theta)$:

$$X_t = \sum_j \psi_j(\phi, \theta)\, w_{t-j} \ .$$

This expression writes each observation as a function of the entire history of the time series. State-space models represent the role of history differently in a finite-dimensional vector, the state. Very Markovian. Autoregressions are a special case in which the state vector — the previous p observations — is observed.

There are several reasons to look for other ways to specify such models. An illustration of the effects of measurement error motivates the need for an alternative to the ARMA representation.

Motivation 1 If the sought data (signal) $\{S_t\}$ and the interfering noise $\{w_t\}$ (which may not be white noise) are independent stationary processes, then the covariance function of the sum of these series $X_t = S_t + w_t$ is the sum of the covariance functions,

$$\mathrm{Cov}(X_{t+h}, X_t) = \mathrm{Cov}(S_{t+h}, S_t) + \mathrm{Cov}(w_{t+h}, w_t) \ ,$$

Recall that the covariance generating function of a stationary process $\{X_t\}$ is a polynomial $G_X(z)$ defined so that the coefficient of z^k is $\gamma(k)$,

$$G_X(z) = \sum_j \gamma(k) z^k \; .$$

Hence, the covariance generating function of X_t is the sum

$$G_X(z) = G_S(z) + G_w(z) \; .$$

Consider the effect of additive white noise: the process $\{X_t\}$ is no longer AR(p), but in general becomes a mixed ARMA process. This is most easily seen by considering the covariance generating process of the sum. The covariance generating function of white noise is a constant, σ_w^2. For the sum:

$$\begin{aligned} G_X(z) &= G_S(z) + G_w(z) \\ &= \frac{\sigma^2}{\phi(z)\phi(1/z)} + \sigma_w^2 \\ &= \tilde{\sigma}^2 \frac{\theta(z)\theta(1/z)}{\phi(z)\phi(1/z)} \end{aligned}$$

An ARMA(p, p) representation is not a *parsimonious* representation for the process: the noise variance contributes one more parameter, but produces p more coefficients in the model.

$\Rightarrow$ We ought to consider other classes of models that can handle common tasks like adding series or dealing with additive observation noise.

Motivation 2 Autoregressions simplify the prediction task. If we keep track of the most recent p cases, then we can 'forget' the rest of history when it comes time to predict. Suppose, though, that the process is not an AR(p). Do we have to keep track of the entire history? We've seen that the effects of past data eventually wash out, but might there be a finite-dimensional way to represent how the future depends on the past?

Motivation 3 We don't always observe the interesting process. We might only observe linear functionals of the process, with the underlying structure being hidden from view. State-space models are natural in this class of indirectly observed processes, such as an array in which we observe only the marginal totals.

State-space models The data is a linear function of an underlying Markov process (the "state") plus additive noise. The state is observed directly and only partially observable via the observed data. The resulting models

1. Make it easier to handle missing values, measurement error.
2. Provide recursive expressions for prediction and interpolation.
3. Support evaluation of Gaussian likelihood for ARMA processes.

The methods are related to hidden Markov models, except the state-space in the models we discuss is continuous rather than discrete.

Model equations S&S formulate the filter in a very general setting with lots of Greek symbols (Ch 6, page 325). I will use a simpler form that does not include non-stochastic "control" variables (*a.k.a.*, ARMAX models with exogenous inputs)

$$\text{State } (d_s \times 1): \quad \boldsymbol{X}_{t+1} = F_t \boldsymbol{X}_t + G_t V_t \tag{1}$$

$$\text{Observation } (d_o \times 1): \quad \boldsymbol{Y}_t = H_t \boldsymbol{X}_t + W_t \tag{2}$$

where the state $\boldsymbol{X}_t$ is a *vector-valued* stochastic process of dimension d_s and the observations $\boldsymbol{Y}_t$ are of dimension d_o. Notice that the time indices in the state equation often look like that shown here, with the error term V_t having a subscript less than the variable on the l.h.s. (X_{t+1}).

The state $\boldsymbol{X}_t$ retains all of the memory of the process; all of the dependence between past and future must "funnel" through the p dimensional state vector. The observation equation is the "lens" through which we observe the state. In most applications, the coefficient matrices are time-invariant: $F_t = F$, $G_t = G$, $H_t = H$. Also, typically $G_t = I$ unless we allow dependence between the two noise processes. (For instance, some models have $V_t = W_t$.)

The error processes in the state-space model are zero-mean white-noise processes with, in general, time-dependent variance matrices

$$\text{Var}(V_t) = Q_t, \quad \text{Var}(W_t) = R_t, \quad \text{Cov}(V_t, W_t) = S_t.$$

In addition,

$$\boldsymbol{X}_t = f_t(V_{t-1}, V_{t-2}, \ldots)$$

and

$$\boldsymbol{Y}_t = g_t(W_t, V_{t-1}, V_{t-2}, \ldots) .$$

Hence, using $X \perp Y$ to denote $\text{Cov}(X, Y) = 0$, past and current values of the state and observation vector are uncorrelated with the current errors,

$$\boldsymbol{X}_s, \boldsymbol{Y}_s \quad \perp \quad V_t \quad \text{for } s < t.$$

and

$$\boldsymbol{X}_s \perp W_t\ , \quad \forall s, t, \qquad \boldsymbol{Y}_s \perp W_t\ , \quad s < t.$$

As with the coefficient matrices, we typically fix $Q_t = Q$, $R_t = R$ and set $S_t = 0$.

Stationarity For models of stationary processes, we can fix $F_t = F$, $G_t = I$, and $H_t = H$. Then, back-substitution gives

$$\boldsymbol{X}_{t+1} = F\boldsymbol{X}_t + V_t = \sum_{j=0}^{\infty} F^j V_{t-j} .$$

Hence, stationarity requires powers of F to "decrease." That is, the eigenvalues of F must be less than 1 in absolute value. Such an F is causal. Causality implies that the roots of the characteristic equation satisfy:

$$0 = \det(F - \lambda I) = \lambda^p - f_1\lambda^{p-1} - f_2\lambda^{p-2} - \cdots - f_p \implies |\lambda| < 1\ . \quad (3)$$

No surprise: this condition on the eigenvalues is related to the condition on the zeros of the polynomial $\phi(z)$ associated with an ARMA process.

The covariances of the observations then have the form

$$\begin{aligned}
\text{Cov}(Y_{t+h}, Y_t) &= \text{Cov}(H(F\boldsymbol{X}_{n+h-1} + V_{t+h-1}) + W_{n+h}, H\boldsymbol{X}_n + W_n) \\
&= H\,F\ \text{Cov}(\boldsymbol{X}_{n+h-1}, \boldsymbol{X}_n)\ H' \\
&= H\,F^h\ P\ H'
\end{aligned}$$

where $P = \text{Var}(\boldsymbol{X}_t)$. In an important paper, Akaike (1975) showed that it works the other way: if the covariances have this form, then a Markovian representation exists.

Origin of model The state-space approach originated in the space program for tracking satellites. Computer systems of the time had limited memory, motivating a search for recursive methods of prediction. In this context, the state is the *actual* position of the satellite and the observation vector contains *observed* estimates of the location of the satellite.

Non-unique The representation is not unique. That is, one has many choices of F, G, and H that manifest the same covariances. It will be up to a model to identify a specific representation. Basically, its like a regression: the predictions lie in a subspace spanned by the predictors; as long as one chooses any vectors that span the subspace, the predictions are the same (though the coefficients may be very, very different).

For example, simply insert an orthonormal matrix M (non-singular matrices work as well, but alter the covariances)

$$\begin{array}{rcl} M\,\boldsymbol{X}_{t+1} = M\,F\,M^{-1}M\,\boldsymbol{X}_t + M\,V_t & \Rightarrow & X^*_{t+1} = F^* X^*_t + V^*_t \\ \boldsymbol{Y}_t = GM^{-1}M\boldsymbol{X}_t + W_t & \Rightarrow & Y_t = G^* X^*_t + W_t \end{array}$$

Dimension of state One can increase the dimension of the state vector without "adding information." In general, we want to find the *minimal* dimension representation of the process, the form with the smallest dimension for the state vector. (This is analogous to adding perfectly collinear variables in regression; the predictions remain the same though the coefficients become unstable.)

Regression Examples

Random walk plus noise & trend Consider the process $\{Z_t\}$ with $Z_1 = 0$ and

$$Z_{t+1} = Z_t + \beta + V_t = t\beta + \sum_{j=1}^{t} V_j$$

for $V_t \sim WN(0, \sigma^2)$. To get the state-space form, write

$$X_t = \begin{pmatrix} Z_t \\ \beta \end{pmatrix}, \quad F_t = F = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad V_t = \begin{pmatrix} v_t \\ 0 \end{pmatrix}.$$

Note that constants get embedded into F. The observations are $Y_t = (1\ 0)X_t + W_t$ with obs noise W_t uncorrelated with V_t. There are two important generalizations of this model in regression analysis.

Least squares regression In the usual linear model with k covariates $x_i = (x_{i1}, \ldots, x_{ik})'$ for each observation, let β denote the state (it is constant) with the regression equation $Y_i = x_i'\beta + \epsilon_i$ acting as the observation equation.

Random-coefficient regression Consider the regression model

$$Y_t = x_t'\beta_t + \epsilon_t, \quad t = 1, \ldots, n, \tag{4}$$

where x_t is a $k \times 1$ vector of fixed covariates and $\{\beta_t\}$ is a k-dimensional random walk,

$$\beta_t = \beta_{t-1} + V_t \ .$$

The coefficient vector plays the role of the state; the last equation is the state equation ($F = I_k$). Equation (4) is the observation equation, with $H_t = x_t$ and $W_t = \epsilon_t$.

AR and MA models in state-space form

AR(p) example Define the so-called *companion matrix*

$$F = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

The eigenvalues of F given by (3) are the reciprocals of the zeros of the AR(p) polynomial $\phi(z)$. (To find the characteristic equation $|F - \lambda I|$, add λ times the first column to the second, then λ times the new second column to the third, and so forth. The polynomial $\phi_p + \lambda\phi_{p-1} + ... + \lambda^p = \lambda^p\phi(1/\lambda)$ ends up in the top right corner. The cofactor of this element is 1.) Hence if the associated process is stationary, then the state-space model is causal. Define

$$\boldsymbol{X}_t \ = \ (X_t, X_{t-1}, \ldots, X_{t-p+1})'$$

$$\boldsymbol{W}_t = (w_t, 0, \ldots, 0)'$$

That is, any scalar AR(p) model can be written as a vector AR(1) process (VAR) with

$$\boldsymbol{X}_t = F\ \boldsymbol{X}_{t-1} + \boldsymbol{W}_t$$

The observation equation $y_t = H\boldsymbol{X}_t$ simply picks off the first element, so let $H = (1, 0, \ldots, 0)$.

MA(1) example Let $\{Y_t\}$ denote the MA(1) process. Write the observation equation as

$$Y_t = (1\ \theta)\boldsymbol{X}_t$$

with state $\boldsymbol{X}_t = (w_t\ w_{t-1})'$ and the state equation

$$\boldsymbol{X}_{t+1} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \boldsymbol{X}_t + \begin{pmatrix} w_{t+1} \\ 0 \end{pmatrix}$$

The state vector of an MA(q) process represented in this fashion has dimension $q + 1$. An alternative representation reduces the dimension of the state vector to q but implies that the errors W_t and V_t in the state and observation equations are correlated.

ARMA models in state-space form

Many choices As noted, the matrices of a state-space model are not fixed; you can change them in many ways while preserving the correlation structure. The choice comes back to the interpretation of the state variables and the relationship to underlying parameters. This section shows 3 representations along with a well-known author who uses them in this style, and ends with the canonical representation.

Throughout this section, the observations $\{y_t\}$ are from a stationary scalar ARMA(p, q) process ($d_o = 1$)

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + w_t + \sum_{j=1}^{q} \theta_j w_{t-j} \tag{5}$$

ARMA(p,q), Hamilton This "direct" representation relies upon the result that the lagged sum of an AR(p) process is an ARMA process. (For example, if z_t is AR(1), then $z_t + c\, z_{t-1}$ is an ARMA process.) The dimension of the state is $d_s = \max(p, q+1)$. Arrange the AR coefficients as a companion matrix ($\phi_j = 0$ for $j > p$)

$$F = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{d_s-1} & \phi_{d_s} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} & \boldsymbol{\phi}' & \\ I_{d_s-1} & & 0_{d_s-1} \end{pmatrix}$$

Then we can write ($\theta_j = 0$ for $j > q$)

$$\begin{aligned} \boldsymbol{X}_t &= F\,\boldsymbol{X}_{t-1} + \boldsymbol{V}_t, \quad \boldsymbol{V}_t = (w_t, 0, \ldots, 0)' \\ y_t &= (1\ \theta_1\ \cdots \theta_{d_s-1})\boldsymbol{X}_t \end{aligned}$$

(One can add measurement error in this form by adding uncorrelated white noise to the observation equation.)

To see that this formulation produces the original ARMA process, let x_t denote the leading scalar element of the state vector. Then it follows from the state equation that $\phi(B)x_t = w_t$. The observation equation is $y_t = \theta(B)x_t$. Hence, we have

$$y_t = \frac{\theta(B)}{\phi(B)} w_t \quad \Rightarrow \quad y_t \text{ is ARMA}(p, q).$$

ARMA(p,q), Harvey Again, let $d = \max(p, q+1)$, but this time use F',

$$F' = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \phi_3 & 0 & 0 & \ddots & 0 \\ \vdots & 0 & 0 & & 1 \\ \phi_d & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Write the state equation as

$$\boldsymbol{X}_t = F'\boldsymbol{X}_{t-1} + w_t \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{d_s-1} \end{pmatrix}$$

The observation equation picks off the first element of the state,

$$y_t = (1\ 0\ \cdots\ 0)' \boldsymbol{X}_t\ .$$

To see that this representation captures the ARMA process, work your way up the state vector from the bottom, by back-substitution. For example,

$$X_{t,d_s} = \phi_{d_s} X_{t-1,1} + \theta_{d_s-1} w_t$$

so that in the next element

$$\begin{aligned} X_{t,d_s-1} &= \phi_{d_s-1} X_{t-1,1} + X_{t-1,d_s} + \theta_{d_s-2} w_t \\ &= \phi_{d_s-1} X_{t-1,1} + (\phi_{d_s} X_{t-2,1} + \theta_{d_s-1} w_{t-1}) + \theta_{d_s-2} w_t \end{aligned}$$

and on up.

ARMA(p,q), Akaike This form is particularly interpretable because Akaike puts the conditional expectation of the ARMA process in the state vector.

The idea works like this. Define the conditional expectation $\widehat{y}_{t|s} = \mathbb{E}\, y_t | y_1, \ldots, y_s$. We want a recursion for the state, and the vector of conditional expectations of the next $d = d_s = \max(p, q+1)$ observations works well since $d > q$ implies we are "beyond" the influences of the moving average component:

$$\widehat{y}_{t+d|t} = \phi_1 \widehat{y}_{t+d-1|t} + \cdots + \phi_d \widehat{y}_{t+1|t}$$

as in an autoregression. Hence, define the state vector,

$$\boldsymbol{X}_t = (\hat{y}_{t|t} = y_t,\ \hat{y}_{t+1|t},\ \ldots, \hat{y}_{t+d-1|t})$$

The state equation updates the conditional expectations as new information becomes available when we observe y_{t+1}. This time write

$$F = \begin{pmatrix} 0_{d-1} & I_{d-1} \\ \multicolumn{2}{c}{\tilde{\boldsymbol{\phi}}'} \end{pmatrix}$$

with the reversed coefficients $\tilde{\phi} = (\phi_p, \phi_{p-1}, \ldots, \phi_1)'$ in the last row.

How should you "update" these predictions to the next time period when the value of w_{t+1} becomes available? Notice that

$$\begin{aligned} y_{t+f} &= \sum_{j=0}^{\infty} \psi_j w_{t+f-j} \\ &= w_{t+f} + \psi_1 w_{t+f-1} + \cdots + \psi_f w_t + \psi_{f+1} w_{t-1} + \cdots \ . \end{aligned}$$

The ψ_j's determine the impact of a given error w_t upon the future estimates. The state equation is then

$$\boldsymbol{X}_t = F\boldsymbol{X}_{t-1} + w_t \begin{pmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_{d-1} \end{pmatrix}$$

As in the prior form, the observation equation picks off the first element of the state,

$$y_t = (1\ 0\ \cdots\ 0)'\boldsymbol{X}_t \ .$$

Canonical representation. This form uses a different time lag. The dimension is possibly smaller, $d = \max(p, q)$, but it introduces dependence between the errors in the two equations – but using w_t in both equations. It's basically a compressed version of Akaike's form; the state again consists of predictions, without the data term $\boldsymbol{X}_t = (\hat{y}_{t+1|t}, \hat{y}_{t+2|t}, \ldots, \hat{y}_{t+d|t})'$. Define the state equation as

$$\boldsymbol{X}_{t+1} = F\boldsymbol{X}_t + w_t \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{d-1} \end{pmatrix}$$

where the ψ's are the infinite moving average coefficients, omitting $\psi_0 = 1$. The observation equation adds this term, reusing the "current" error:

$$Y_t = (1, \ldots, 0, 0)\boldsymbol{X}_t + w_t.$$

The value of the process is $y_t = \hat{y}_{t|t-1} + w_t$. This defines the observation equation. (Recent discussion of this approach appears in de Jong and Penzer (2004)).

Example ARMA(1,1) of the canonical representation. First observe that the moving average weights have a *recursive* form,

$$\psi_0 = 1, \psi_1 = \phi_1 + \theta_1,\ \psi_j = \phi\psi_{j-1}, \quad j = 2, 3, \ldots .$$

The moving average form makes it easy to recognize the minimum mean squared error predictor

$$\hat{X}_{t|t-1} = X_t - w_t = \psi_1 w_{t-1} + \psi_2 w_{t-2} + \cdots = \sum_{j=1}^{\infty} \psi_j w_{t-j} \ .$$

Thus we can write

$$\begin{aligned} X_{t+1} &= w_{t+1} + \psi_1 w_t + \psi_2 w_{t-1} + \psi_3 w_{t-2} + \cdots \\ &= w_{t+1} + \psi_1 w_t + \phi(\psi_1 w_{t-1} + \psi_2 w_{t-2} + \cdots) \\ &= w_{t+1} + \psi_1 w_t + \phi \hat{X}_{t|t-1} \ . \end{aligned} \tag{6}$$

Hence, we obtain the scalar state equation

$$\hat{X}_{t+1|t} = \phi(\hat{X}_{t|t-1}) + \psi_1 w_t;$$

and the observation equation

$$X_t = \hat{X}_{t|t-1} + w_t \ .$$

Comments on the canonical representation:

- Prediction from this representation is easy, simply compute $F^f \boldsymbol{X}_{t+1}$.
- Moving average representation is simple to obtain via back-substitution. Let $H = (\psi_v, \ldots, \psi_1)'$ and write

  $$\boldsymbol{X}_t = HZ_{t-1} + FHZ_{t-2} + F^2 HZ_{t-2} + \cdots$$

  so that

  $$Y_t = Z_t + G(HZ_{t-1} + FHZ_{t-2} + F^2 HZ_{t-2} + \cdots).$$

Equivalence of ARMA and state-space models

Equivalence Just as we can write an ARMA model in state space form, it is also possible to write a state-space model in ARMA form. The result is most evident if we suppress the noise in the observation equation (2) so that

$$\boldsymbol{X}_{t+1} = F\boldsymbol{X}_t + V_t, \quad y_t = H\,\boldsymbol{X}_t\ .$$

Back-substitute Assuming F has of dimension p, write

$$\begin{aligned} y_{t+p} &= H(F^p\boldsymbol{X}_t + V_{t+p-1} + FV_{t+p-2} + \cdots F^{p-1}V_t) \\ y_{t+p-1} &= H(F^{p-1}\boldsymbol{X}_t + V_{t+p-2} + FV_{t+p-3} + \cdots F^{p-2}V_t) \\ &\ldots \\ y_{t+1} &= H(F\boldsymbol{X}_t + V_t) \\ y_t &= H\boldsymbol{X}_t \end{aligned}$$

Referring to the characteristic equation (3), multiply the first equation by 1, the second by $-f_1$, the next by $-f_2$ etc, and add them up:

$$\begin{aligned} y_{t+p} - f_1 y_{t+p-1} - \cdots - f_p y_t &= H(F^p - f_1F^{p-1} - \cdots f_p) + \sum_{j=1}^{p} C_j V_{t+p-j} \\ &= \sum_{j=1}^{p} C_j V_{t+p-j} \end{aligned}$$

by the Cayley-Hamilton theorem (matrices satisfy their characteristic equation). Thus, the observed series y_t satisfies a difference equation of the ARMA form (with error terms V_t).

Further reading

1. Jaczwinski
2. Astrom
3. Harvey
4. Hamilton
5. Akaike papers