State Space Models and the Kalman Filter

Paul Pichler
Seminar paper prepared for
40461 Vektorautoregressive Methoden
by Prof. Robert Kunst

Januaray 2007

Contents

1 Introduction

In many applications, the driving forces behind the evolution of economic variables are (at least partially) not observable or measurable. For example, on an individual level a person's income may depend on its intelligence, special abilities, social skills, and so on. Obviously, these variables cannot be measured directly. Similarly, on an aggregate level, economic theory suggests that macroeconomic variables such as economic growth are driven by unobservable factors, e.g. technological change or human capital accumulation.

When explanatory variables are not observable, standard VAR models can no longer be applied to study the evolution of the endogenous variables. However, it is easy to extend the VAR framework to analyze scenarios with unobservable explanatory variables by using *state space models*.

2 State Space models

State space models allow the researcher to model an observed (multiple) time series, $\{y_t\}_{t=1}^T$, as being explained by a vector of (possibly unobserved) state variables, $\{z_t\}_{t=1}^T$, which are driven by a stochastic process. A basic linear state space model takes the following form:

$$\begin{aligned} y_t &= Hz_t + v_t, \qquad v_t \sim N(0, \Sigma_v) \qquad \text{measurement equation} \qquad (1) \\ z_t &= Bz_{t-1} + w_t, \quad w_t \sim N(0, \Sigma_w) \qquad \text{transition equation} \qquad (2) \end{aligned}$$

The first equation, called *measurement equation*, describes the relation between the observed time series, y_t, and the (possibly unobserved) state z_t. In general, it is assumed that the data y_t are measured with error, which is reflected in the measurement error v_t that enters (1). The standard approach is to model v_t as a Gaussian error term, $v_t \sim N(0, \Sigma_v)$. The second equation, the *transition equation*, describes the evolution of the state variables as being driven by the stochastic process of innovations w_t. Typically one assumes normal innovations, such that $w_t \sim N(0, \Sigma_w)$.
State space models can be formulated much more general than the specification (1)/(2). For example, the system matrices H and B could depend explicitly on time, or one could introduce policy variables and constants in the specification. However, mostly for notational simplicity we will discuss state space models using the basic specification (1)/(2).

Let us briefly present two simple examples. First, consider a linear Real

Business Cycle model. In state space form, this model could be written as

$$\begin{pmatrix} \mathcal{Y}_t \\ \mathcal{C}_t \\ \mathcal{I}_t \end{pmatrix} = H \begin{pmatrix} \mathcal{K}_t \\ \mathcal{Z}_t \end{pmatrix} + v_t, \quad v_t \sim N(0, \Sigma_v) \quad \text{measurement equ.} \tag{3}$$

$$\begin{pmatrix} \mathcal{K}_t \\ \mathcal{Z}_t \end{pmatrix} = B \begin{pmatrix} \mathcal{K}_{t-1} \\ \mathcal{Z}_{t-1} \end{pmatrix} + w_t, \quad w_t \sim N(0, \Sigma_w) \quad \text{transition equ.} \tag{4}$$

where $\mathcal{Y}_t$ denotes aggregate output, $\mathcal{C}_t$ denotes private consumption, $\mathcal{I}_t$ denotes investment, $\mathcal{K}_t$ denotes the capital stock, and $\mathcal{Z}_t$ denotes an (exogenously evolving) technology level. The idea is thus to model output, consumption and investment as being driven by the capital stock and by the technology level. Naturally, we would assume that the technology level cannot be directly observed. This is not obvious for the capital stock. Although data on the capital stock are available for many countries, these data are known to have only a low quality, such that a researcher may want to consider capital an unobservable variable.
Finally, consider the following VAR(p) model:

$$Y_t = AY_{t-1} + U_t \tag{5}$$

It can be written in state space form by using the following specification:

$$y_t = Hz_t + v_t, \qquad z_t = Bz_{t-1} + w_t$$

with $y_t := Y_t$, $H := I$, $z_t := Y_t$, $v_t := 0$, $B := A$ and $w_t = U_t$. Similarly, many econometric models (in particular time series models) can be written in state space form. This obvious flexibility of the state space approach has contributed much to its popularity in recent years.

3 Estimation of State Space models

In practical applications, the system matrices H and B together with the variances Σ_v and Σ_w are unknown and have to be estimated. Obviously, whenever the explanatory variables are not observable, Least Squares estimation is not a way to go. However, even in this case, one can apply likelihood based inference, since the so-called *Kalman filter* allows to construct the likelihood function associated with a state space model.

3.1 Kalman filtering

Assume that we observe data $\{y_t\}_{t=1}^T$, that are to be described by the model (1)/(2). Assume that reasonable (but not necessarily the true) values for the model's parameters are available, and equal to $H^*, B^*, {\Sigma_v}^*, \Sigma_w^*$. Summarize these values by $\delta = \{H*, B*, \Sigma_v^*, \Sigma_w^*\}$. Let the sample density (or likelihood) function associated with a state space model for given parameters δ be denoted by $f(y_1, y_2, ..., y_T; \delta)$. By Bayes theorem, we can factor the likelihood as

$$\begin{aligned} f(y_1, y_2, ..., y_T; \delta) &= f(y_1, \delta) f(y_2|y_1, \delta) f(y_3|y_2, y_1, \delta) ... f(y_T|y_{T-1}, ..., y_1, \delta) \\ &= \prod_{t=1}^{T} f(y_t|y^{t-1}, \delta) \end{aligned} \quad (6)$$

where $y^0 = \emptyset$ and $y^{t-1} = (y_1, y_2, ..., y_{t-1})$, for $t \geqslant 2$. The log-likelihood function is thus given by

$$\ln L(y^T, \delta) = \sum_{t=1}^{T} \ln f(y_t|y^{t-1}, \delta)$$

Obviously, to construct the likelihood function, we need to derive the densities

$$f(y_t|y^{t-1}, \delta), t = 1, 2, ..., T$$

We can achieve this by using filtering techniques, in particular - when the system is linear and errors are Gaussian - by using the Kalman (1960) filter. The Kalman filter is a recursive procedure that involves the following steps:

1. Initialization
2. Prediction
3. Correction
4. Likelihood construction

In the following, we will discuss each of these steps in greater detail. Before, let us introduce some notation: in the remainder of this paper, we will use $X_{t|s}$ to denote the prediction of the variable X at time t, conditional upon information available at time s.

3.1.1 Initialization

The Kalman filter is initialized by deriving the best predictor of the initial state, $z_{0|0}$, and an estimate of its covariance matrix, $\Sigma_{0|0}^z = E[(z_0 - z_{0|0})(z_0 - z_{0|0})']$. If the process is stationary, this is straightforward, since we can build on the steady state of the system. More precisely, we can set $z_{0|0} = z^*$ and $\Sigma_{0|0}^z = \Sigma^*$ such that

i) $z^* = Bz^*$

ii) $\Sigma^* = B\Sigma^* B' + \Sigma_w = [I - B \otimes B]^{-1} vec(\Sigma_w)$.

Before moving to the next step, we set $t = 1$ such that, consequently, $z_{t-1|t-1} = z_{0|0}$ and $\Sigma^z_{t-1|t-1} = \Sigma^z_{0|0}$.

3.1.2 Prediction

At time t, we can use $z_{t-1|t-1}$ and $\Sigma^z_{t-1|t-1}$ together with the transition equation to compute

$$
\begin{aligned}
z_{t|t-1} &= Bz_{t-1|t-1} && (7)\\
\Sigma^z_{t|t-1} &= B\Sigma^z_{t-1|t-1}B' + \Sigma_w && (8)
\end{aligned}
$$

We can then use $z_{t|t-1}$ to construct the forecast $y_{t|t-1} = Hz_{t|t-1}$. Having observed y_t, we can construct the forecast error

$$u_t = y_t - y_{t|t-1} = y_t - Hz_{t|t-1} = v_t + H(z_t - z_{t|t-1}) \tag{9}$$

Because of Gaussian errors, it follows that $u_t \sim N(0, \Sigma_v + H\Sigma^z_{t|t-1}H')$. Furthermore, since $y_t = u_t + y_{t|t-1}$, it follows that $f(y_t|y^{t-1}; \delta) = f(u_t; \delta)$.

Let us briefly summarize our main results so far. In order to construct the likelihood we need $f(y_t|y^{t-1}, \delta)$ for all $t = 1, 2, ..., T$; given $z_{t-1|t-1}$ and $\Sigma^z_{t-1|t-1}$, we can compute $f(y_t|y^{t-1}, \delta)$ from the normal density function,

$$f(y_t|y^{t-1}, \delta) = \frac{1}{\sqrt{(2\pi)^{n_y}|\Sigma_v + H\Sigma^z_{t|t-1}H'|}} \exp\left(-\frac{u_t{}'(\Sigma_v + H\Sigma^z_{t|t-1}H')^{-1}u_t}{2}\right) \tag{10}$$

Consequently, to compute $f(y_{t+1}|y^t, \delta)$ we need $z_{t|t}$ and $\Sigma^z_{t|t}$. In other words, we need to correct our state predictions using the new information at time t, y_t. We can do this in the way described in the next step.

3.1.3 Correction

Having observed the data y_t, we can update (correct) the predictions $z_{t|t-1}$ and $\Sigma^z_{t|t-1}$ according to the Kalman (1960) formulae:

$$
\begin{aligned}
z_{t|t} &= z_{t|t-1} + K_t(y_t - y_{t|t-1}) = z_{t|t-1} + K_t(y_t - Hz_{t|t-1}), && (11)\\
\Sigma^z_{t|t} &= \Sigma^z_{t|t-1} - K_t(\Sigma_v + H\Sigma^z_{t|t-1}H')K_t{}', && (12)
\end{aligned}
$$

where

$$K_t = \Sigma^z_{t|t-1}H'(H\Sigma_{t|t-1}H' + \Sigma_\nu)^{-1}. \tag{13}$$

The intuition behind these formulae is quite straightforward. The corrected prediction is a linear combination between the *old prediction*, $z_{t|t-1}$, and the current prediction error, $(y_t - y_{t|t-1})$. Given the linear form, K_t is chosen such that it minimizes the prediction error variance. We do not provide a formal proof of these formulae, since they can be found in many econometrics textbooks.

Unless $t = T$, we increase t and return to the prediction step. Otherwise, we continue and construct the likelihood as described in the following step.

3.1.4 Likelihood construction

The two previous steps recursively compute $f(y_t|y^{t-1}, \delta)$ for $t = 1, 2, ..., T$. Obviously, these densities can be used to construct the likelihood according to

$$L(y^T, \delta) = \prod_{t=1}^{T} f(y_t|y^{t-1}, \delta) \tag{14}$$

3.2 Maximizing the Likelihood Function

The previous subsection has demonstrated how to compute the likelihood of a data sample conditional on parameters δ. This likelihood will in general be a complex nonlinear function of the parameters, such that often, maximizing the likelihood function will be conducted numerically. Several methods are available to conduct this numerical maximization. If the likelihood function is smooth and continuous, Gradient-based methods (e.g. Newton's method) will allow to derive the Maximum Likelihood methods in a straightforward way.

4 Forecasting with State Space models

Having obtained the maximum likelihood estimates, δ_{ML}, one can use the state space model to forecast the observables. In particular, one can use the final state predictor $z_{T|T}$ implied by the Maximum likelihood estimates δ_{ML} together with the measurement and transition equation to construct $y_{T+h|T}$ for $h = 1, 2, ...$ according to

$$y_{T+h|T} = HB^h z_{T|T} \tag{15}$$

5 Conclusion

This short seminar paper has provided a brief introduction to state space models. In particular, we have illustrated how to estimate the parameters of such models using the Kalman filter. The outlined methods allow to model the evolution of observable variables that are driven by (partially) unobservable forces.

References

KALMAN, R. E. (1960): "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D), 35–45.