State Space Model

Abstract

This methodological framework examines the State Space Model (SSM), which consists of equations describing the relationships between observable and unobservable variables. In particular, the unknown parameters characterising these relationships are estimated using maximum likelihood in conjunction with the Kalman filter. The Kalman filter is a recursive algorithm that enables the estimation of unobservable state variables based on observed data. The objective of this article is to assess the applicability of an additional modelling approach for estimating the fundamental value of housing prices and to support a comprehensive analysis of Uzbekistan's real estate market based on international experience.

Keywords: Bayes theorem, Kalman filter, covariance matrix, maximum likelihood, density function.

Prepared by the Financial Stability Department of the Central Bank of Uzbekistan.

If you have any comments, please send them to udjumanazarov@cbu.uz.

State Space Model

The SSM[1] serves to theoretically estimate the value of a endogenous variable by determining the linear correlation function between the variables. This model consists of an observable or measurement equation representing the level of interdependence of observable and unobservable variables and a transition equation representing dynamic state variables:

$$y_t = H_t z_t + G_t x_t + v_t$$
$$z_t = B_{t-1} z_{t-1} + F_{t-1} x_{t-1} + w_{t-1}$$

Here, y_t is a vector of observable endogenous variables, x_t is a vector of observable exogenous variables, z_t is a state vector, v_t is a vector of measurement errors, w_t is a vector of transition equation errors, H_t, G_t, B_t and F_t are matrices of estimated coefficients.

The estimated coefficients representing the relationships between endogenous, exogenous and state variables in this model are estimated by maximum likelihood using the Kalman filter.

Kalman filter

A Kalman filter is a mathematical algorithm that recursive estimates unobservable variables with observable ones. In this case, the values of conditional expected unobserved state variables $z_{t|t}$ and conditional covariance matrix $\Omega_{t|t}$ depending on the data up to this period are determined for each t period through the Kalman filter.

The following additional notation in stating the Kalman filter recursions:

$$z_{t|s} := E(\, z_t |\, y_1, \ldots, y_s),$$
$$\textstyle\sum_z (t|s) := Cov(\, z_t |\, y_1, \ldots, y_s),$$
$$y_{t|s} := E(\, y_t |\, y_1, \ldots, y_s),$$
$$\textstyle\sum_y (t|s) := Cov(\, y_t |\, y_1, \ldots, y_s),$$
$$(z|y) \sim N(\mu, \textstyle\sum).$$

Under the previously stated conditions, the normality assumption implies:

[1] Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer Science & Business Media.

$$(z_t|\, y_1, \ldots, y_{t-1}) \sim N(z_{t|t-1}, \Sigma_z(t|t-1)),\ t = 2, \ldots, T$$
$$(z_t|\, y_1, \ldots, y_t) \sim N(z_{t|t}, \Sigma_z(t|t)),\ t = 1, \ldots, T$$
$$(y_t|\, y_1, \ldots, y_{t-1}) \sim N(y_{t|t-1}, \Sigma_y(t|t-1)),\ t = 2, \ldots, T$$
$$(z_t|\, y_1, \ldots, y_T) \sim N(z_{t|T}, \Sigma_z(t|T))$$
$$(y_t|\, y_1, \ldots, y_T) \sim N(y_{t|T}, \Sigma_y(t|T)),\ t > T.$$

The unobservable variables and covariance matrices are determined by the prediction, correction and forecasting steps after running the Kalman filter.

A Kalman filter is run by entering the initial values of the unobservable variables and covariance matrices:

$$z_{0|0} := \mu_0\ ,\ \Sigma_z\ (0|0) := \Sigma_0$$

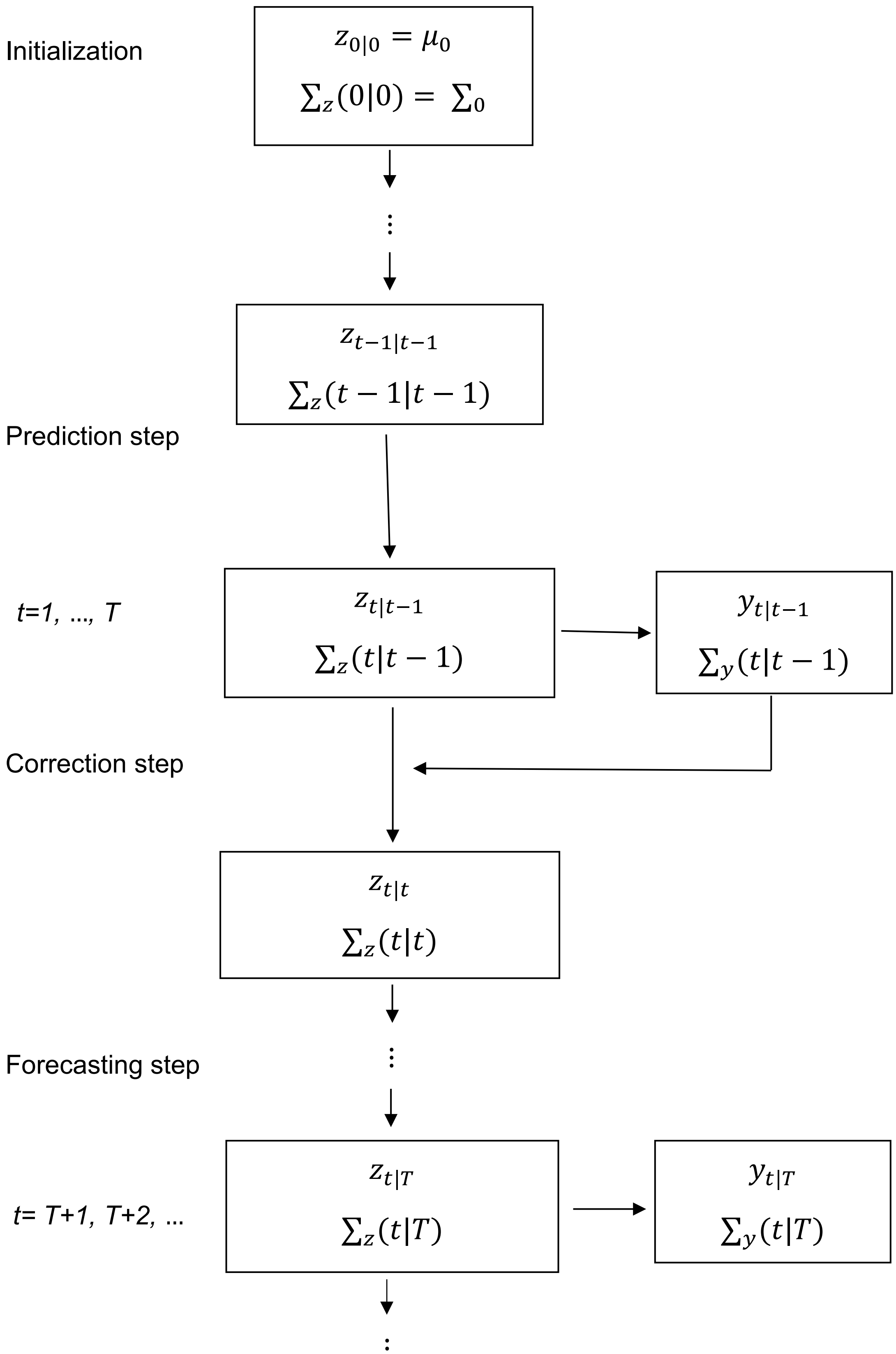
The prediction step of the Kalman filter ($1 \leq t \leq T$):

$$z_{t|t-1} = \mathrm{B} z_{t-1|t-1} + \mathrm{F} x_{t-1},$$
$$\Sigma_z(t|t-1) = \mathrm{B}\, \Sigma_z(t-1|t-1)\mathrm{B}' + \Sigma_w,$$
$$y_{t|t-1} = \mathrm{H}_t\, z_{t|t-1} + \mathrm{G} x_t,$$
$$\Sigma_y(t|t-1) = \mathrm{H}_t\, \Sigma_z(t|t-1)\mathrm{H}'_t + \Sigma_v.$$

The correction step of the Kalman filter ($1 \leq t \leq T$):

$$z_{t|t} = z_{t|t-1} + \mathrm{P}_t(y_t - y_{t|t-1}),$$
$$\Sigma_z(t|t) = \Sigma_z(t|t-1) - \mathrm{P}_t\, \Sigma_y(t|t-1)\mathrm{P}'_t,$$
$$\mathrm{P}_t := \Sigma_z(t|t-1)\mathrm{H}'_t\, \Sigma_y(t|t-1)^{-1}.$$

Figure 1. Kalman filter recursions



Source: Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer Science & Business Media.

After running the Kalman filter, the prediction and correction steps are performed for the period $t = 1$. The steps of prediction and correction are repeated successively for subsequent periods as well.

The last forecasting step of the Kalman filter is $t > T$:

$$
\begin{aligned}
z_{t|T} &= \mathrm{B} z_{t-1|T} + \mathrm{F} x_{t-1}, \\
\textstyle\sum_z(t|T) &= \mathrm{B} \textstyle\sum_z(t-1|T)\mathrm{B}' + \textstyle\sum_w, \\
y_{t|T} &= \mathrm{H}_t \, z_{t|T} + \mathrm{G} x_t, \\
\textstyle\sum_y(t|T) &= \mathrm{H}_t \textstyle\sum_z(t|T)\mathrm{H}'_t + \textstyle\sum_v.
\end{aligned}
$$

The forecasting step may be carried out recursively for $t = T+1, T+2, \dots$.

The vectors of estimated coefficients used in the steps of the Kalman filter are estimated using the maximum likelihood function.

Maximum likelihood estimation

The time invariant vector $\delta(B, F, G, H_t, \sum_w, \sum_v, \sum_0, \mu_0)$ are uniquely determined and at least twice continuously differentiable with respect to elements of δ.

To determine the unknown parameters, a log-likelihood function is first created and the maximum value of this function is estimated.

According to Bayes' theorem, the sample density function has the following form:

$$
\begin{aligned}
f(y_1, \dots, y_T; \delta) &= f(y_1; \delta) f(y_2, \dots, y_T | y_1; \delta) \\
&\vdots \\
&= f(y_1; \delta)\, f(y_1|y_2; \delta) \, \cdots \, f(y_T | y_1, \dots, y_{T-1}; \delta).
\end{aligned}
$$

Also, the Gaussian log-likelihood function for K dimensional y_t is expressed by the logarithmic value of the multiplication of density functions as follows:

$$
\begin{aligned}
\ln l\,(\delta | y_1, \dots, y_T) &= \ln f(y_1, \dots, y_T; \delta) \\
&= \ln f(y_1; \delta) + \textstyle\sum_{t=2}^{T} \ln f(y_t | y_1, \dots, y_{t-1}; \delta) \\
&= -\tfrac{KT}{2} \ln(2\pi) - \tfrac{1}{2} \textstyle\sum_{t=1}^{T} \ln |\, \Sigma_y(t|t-1)| \\
&\quad -\tfrac{1}{2} \sum_{t=1}^{T} \left(y_t - y_{t|t-1}\right)' \Sigma_y(t|t-1)^{-1} \left(y_t - y_{t|t-1}\right)
\end{aligned}
$$

This log-likelihood function uses the following as starting values:

$$y_{1|0} \coloneqq E(y_1)$$
$$\Sigma_y(1|0) \coloneqq Cov(y_1)$$
$$(y_t | y_1, \dots, y_{t-1}) \sim \mathcal{N}\left(y_{t|t-1}, \Sigma_y(t|t-1)\right)$$
$$t = 1, \dots, T$$

After determining the parameter vector δ, all the quantities in the log-likelihood function can be computed with the Kalman filter recursions.

To simplify the representation of the log-likelihood function, the following notation is introduced:

$$e_t(\delta) \coloneqq y_t - y_{t|t-1}$$
$$\Sigma_t(\delta) \coloneqq \Sigma_y(t|t-1)$$

The log-likelihood function will have the following form through the given notations:

$$ln\, l\,(\delta) = -\frac{KT}{2}\, ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}[ln|\Sigma_t(\delta)| + e_t(\delta)'\Sigma_t(\delta)^{-1}e_t(\delta)]$$

There are different ways to estimate the unknown parameters at their maximum values of the log-likelihood function. In particular, the values of the parameters that maximize the log-likelihood can be determined by taking special derivatives with respect to the unknown parameters and equalizing them to zero.