Bennedsen, Hillebrand, and Koopman
A Multivariate Dynamic Statistical Model of the Global Carbon Budget 1959—2020

A note on the minimal representation of the linear state space model

January 11, 2023

In the paper, we have specified a state vector of 17 entries, many of them redundant in a mathematical sense ($G\_ATM^*, S\_LND^*, S\_OCN^*$ are all linear functions of C^*). Many constant states were added to the state vector as an implementation device for model parameters. It is often interesting to consider the representation of a state space model with a minimum number of states, see, for example, Hannan and Deistler (2012). This is the representation that we present in this note.

1 The state equation

The three key budget variables are placed in the 3×1 vector $B_t^* = (C_t^*, E_t^*, X_t^E)'$. The state space formulation requires Markovian (first-order autoregressive) updating equations for the 3×1 vector B_t^* and for the 4×1 vector X_t with its ith element given by $X_{i,t}$, for $i = 1, \dots, 4$. Given the established dynamic properties for the GCB variables, we can represent all variables as a first-order vector autoregressive model. In case of the concentration variable C_t^*, using the state equations defined in Section 2.4 in the paper, we can write $(1+\beta)C_{t+1}^* = k + C_t^* + E_t^* - \beta \mathrm{ENSO}_{t+1} + \beta_5 \Delta \mathrm{ECON}_{t+1} + X_t^E$, for constants $k = -c_1 - c_2$ and $\beta = (\beta_1 + \beta_2)/C_{1750}$. We recall that ENSO and ECON are treated as exogenous to the system. The autoregressive equation for CO_2 concentrations is then given by

$$C_{t+1}^* = \delta\, k + \delta C_t^* - \delta\beta \mathrm{ENSO}_{t+1} + \delta E_t^* + \delta\beta_5 \Delta \mathrm{ECON}_{t+1} + \delta X_t^E, \tag{1}$$

with $\delta := (1+\beta)^{-1} < 1$. The vector autoregressive equation for B_t^* is then given by

$$B_{t+1}^* = b + \Xi B_t^* + \tilde{\Gamma} U_t + V \kappa_t,$$

where $U_t = (\mathrm{ENSO}_{t+1}, \Delta \mathrm{ECON}_{t+1})'$ contains the exogenous variables, and

$$b = \begin{pmatrix} \delta\, k \\ d \\ 0 \end{pmatrix}, \quad \Xi = \begin{bmatrix} \delta & \delta & \delta \\ 0 & 1 & 1 \\ 0 & 0 & \phi_E \end{bmatrix}, \quad \tilde{\Gamma} = \begin{bmatrix} -\delta\beta & \delta\beta_5 \\ 0 & \beta_5 \\ 0 & 0 \end{bmatrix}, \quad V = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The updating equation for X_t is given by

$$X_{t+1} = \Phi X_t + \eta_t,$$

where Φ is a diagonal 4×4 matrix with entries $\phi_1, \ldots, \phi_4$ and η_t is a 4×1 vector with its ith element $\eta_{i,t}$ for $i = 1, \ldots, 4$.

The linear state space representation of the GCB model is obtained by defining the state vector $\alpha_t = (B_t^{*\prime}, X_t^\prime)^\prime$ that contains all unobserved variables in the GCB model and the innovation vector $\xi_t = (\kappa_t, \eta_t^\prime)^\prime$. The updating or state equation is then given by

$$\alpha_{t+1} = a + T\alpha_t + \Gamma U_t + R\,\xi_t, \quad a = \begin{pmatrix} b \\ 0 \end{pmatrix}, \quad T = \begin{bmatrix} \Xi & 0 \\ 0 & \Phi \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \tilde{\Gamma} \\ 0_{4\times 2} \end{bmatrix}, \quad R = \begin{bmatrix} V & 0 \\ 0 & I_4 \end{bmatrix}, \quad (2)$$

where I_4 is the 4×4 identity matrix, $0_{4\times 2}$ is a 4×2 matrix of zeroes, and U_t is treated as exogenous regressors (Durbin and Koopman, 2012, section 3.2.5). The variance matrix of the zero mean innovation vector ξ_t is given by

$$\text{Var}(\xi_t) = Q = D \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & r_{12} & r_{13} & r_{14} \\ 0 & r_{12} & 1 & r_{23} & r_{24} \\ 0 & r_{13} & r_{23} & 1 & r_{34} \\ 0 & r_{14} & r_{24} & r_{34} & 1 \end{bmatrix} D,$$

where D is a diagonal 5×5 matrix with elements $\sigma_\kappa, \sigma_1, \ldots, \sigma_4$. Conditional on the variables E_t^* and U_t, the variable C_t^* follows a stationary process, while all remaining variables in the vectors B_t^* and X_t are also stationary. This implies that the initial conditions for all variables in the state vector α_t are properly defined except the one for E_t^*, which can be treated using a *diffuse prior* (Durbin and Koopman, 2012, section 5.2).

2 The observation vector equation

The observation equation for $y_t = (C_t, S\_LND_t, S\_OCN_t, E_t)^\prime$ follows immediately from the definitions in Section 2.5 of the paper; it is given by

$$y_t = z + Z\alpha_t + PU_t, \qquad z = \begin{pmatrix} 0 \\ c_1 \\ c_2 \\ 0 \end{pmatrix}, \quad Z = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ \bar{\beta}_1 & 0 & 0 & 0 & 1 & 0 & 0 \\ \bar{\beta}_2 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 0 \\ \beta_3 & 0 \\ \beta_4 & 0 \\ 0 & 0 \end{bmatrix},$$

where $\bar{\beta}_j = \beta_j \,/\, C_{1750}$, for $j = 1, 2$. We emphasize that this observation equation has no additive disturbance vector: the *signal* is constructed from the elements in B_t^* and the *noise* is represented by the elements in X_t. Both the signal and noise variables are contained in the state vector α_t. Finally, we notice that for the identification (or estimation) of the signal, the observation variables C_t and E_t are sufficient. However, the land and ocean sink observation variables, S_LND_t and S_OCN_t, need to be included in the model, since they are instrumental for the budget equation to hold. The observation vector equation can easily be extended with exogenous effects such as explanatory and dummy variables (Durbin and Koopman, 2012, section 3.2.5).

The state space framework facilitates the signal extraction of the variables of interest such as those in vector B_t^*, but also $S\_LND_t^*$, $S\_OCN_t^*$ and other functions of B_t^*, the estimation of the parameters in the system matrices, and forecasting. Assuming that the innovation vector ξ_t in equation (2) is normally distributed, the log-likelihood function is evaluated by the Kalman filter via the prediction error decomposition. We adopt the method of maximum likelihood for parameter estimation, where the maximization of the log-likelihood function relies on a numerical optimization method. When the model is correctly specified, the standardized prediction errors should be normally, independently and identically distributed with zero mean and unit variance. Various residual diagnostics can be used to verify these properties. Finally, the smoothed estimates of the state vector (based on all data) are obtained from the Kalman filter and a related smoothing method.

References

Durbin, J. and S. J. Koopman, *Time series analysis by state space methods*, 2nd ed., Oxford University Press, 2012.

Hannan, E.J. and M. Deistler, *The statistical theory of linear systems*, SIAM, 2012.