A Flexible State Space Model and its Applications

Qian, Hang

Iowa State University

April 2012

A Flexible State Space Model and its Applications

Hang Qian[1]

Abstract

The standard state space model (SSM) treats observations as imprecise measures of the Markov latent states. Our flexible SSM treats the states and observables symmetrically, which are simultaneously determined by historical observations and up to first-lagged states. The only distinction between the states and observables is that the former are latent while the latter have data. Despite the conceptual difference, the two SSMs share the same Kalman filter. However, when the flexible SSM is applied to the ARMA model, mixed frequency regression and the dynamic factor model with missing data, the state vector is not only parsimonious but also intuitive in that low-dimension states are constructed simply by stacking all the relevant but unobserved variables in the structural model.

Keywords: State Space Model, Kalman Filter, ARMA, Mixed Frequency, Factor Model

1. Introduction

Starting with the path-breaking paper of Kalman (1960), the state space model (SSM) has been widely applied in engineering, statistics and eco-

[1]We would like to thank Brent Kreider for helpful comments on this paper. Corresponding author: Department of Economics, Iowa State University, Ames, IA, 50010. Email: hqi@iastate.edu

nomics. See Harvey (1989), Hamilton (1994), Durbin and Koopman (2001) for comprehensive presentations on the SSM and its applications in time series analysis. Basdevant (2003) surveys various applications in macroeconomics. For practitioners, the art consists in the model building, that is, to cast a structural model into its state space form. Once an SSM is built, the likelihood function as well as the smoothed states can be routinely evaluated by the Kalman filter. The state space representation is not unique, for one can increase the dimensions of the state vector but equally represent the same data generating process. Two aspects of a representation, namely parsimony and intuitiveness, are of major concern. A parsimonious model with minimum length of the state vector avoids large matrix manipulations, saves overheads in computation (say $0 * 0$) and thus accelerates the Kalman filter. An intuitive form with the economically interpretable state vector enhances attractiveness of the representation, for both predicted latent states and smoothed historical states bear economic significances. Furthermore, intuitiveness also means a practitioner can straightforwardly rewrite a structural model into its state space form.

The SSM derived its name because the system is driven by unobserved states that have a Markov dependence structure. The observed variables are imprecise measures of the states in each period. Based on this structure and Gaussian disturbances, the Kalman filter first obtains the joint predictive distribution of the current states and observables, conditional on the previous information set (historical observables). Then the states are updated by further conditioning on the current observables. Through recursive predicting and updating at each date, the filter gradually assimilates information conveyed by the observed data.

Our argument is that the filtering procedure does not necessarily require the model structure implied by the standard SSM. The recursion is valid as long as no higher than first-lagged states are in the dynamic system, without restrictions on how the lagged observations affect current states and observables. In other words, the Markov transition of states is suitable but not required for the forward recursion. That motivates us to bring into SSM more symmetry and two-way dynamics between the states and observables. The flexible model allows dynamic dimensions of the state and measurement vectors, lagged observations in the equations and first-lagged state vector in the measure equation. Examined individually, each new feature seems trivial. Combining these features, however, will lead to non-trivial simplification of the state space representation of many time series models. The idea of our flexible SSM is to put all the relevant but unobserved variables in the state vector at each date and all the observables in the measurement vector. Therefore, our state vector always bears structural interpretations. Furthermore, the simplification is not only conceptual but also computational in that the state vector typically has low dimensions under our flexible SSM.

The rest of the paper is organized as follows. Section 2 sets up the flexible SSM and Section 3 explains the filtering procedure. Section 3 to 5 applies our approach to the ARMA model, the mixed frequency vector autoregression and the dynamic factor model with missing data. Our state space representations are distinct from those in the literature and fewer variables are put in the state vector. Section 6 concludes the paper.

2. A flexible state space model

First consider a standard SSM. Let $\boldsymbol{\xi}_t$ be a $m \times 1$ latent state vector and $\boldsymbol{Y}_t$ be a $n \times 1$ measurement vector. The dynamic system consists of a transition equation of states and a measurement equation that bridges the observables and the unobservables.

$$\begin{aligned} \boldsymbol{\xi}_t &= \boldsymbol{c}_t + \boldsymbol{F}_t \boldsymbol{\xi}_{t-1} + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{Y}_t &= \boldsymbol{d}_t + \boldsymbol{H}_t \boldsymbol{\xi}_t + \boldsymbol{u}_t. \end{aligned} \tag{1}$$

where the Gaussian white noises $\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{u}_t \end{pmatrix} \sim N \left[\mathbf{0}, \begin{pmatrix} \boldsymbol{Q}_t & \boldsymbol{S}_t \\ \boldsymbol{S}_t' & \boldsymbol{R}_t \end{pmatrix} \right]$. The coefficients $\boldsymbol{c}_t, \boldsymbol{F}_t, \boldsymbol{d}_t, \boldsymbol{H}_t, \boldsymbol{Q}_t, \boldsymbol{R}_t, \boldsymbol{S}_t$ are time-varying but deterministic. The system starts from Date 1 and runs through Date T with the observations $\boldsymbol{Y}_1^T \equiv \{\boldsymbol{Y}_1, ..., \boldsymbol{Y}_T\}$, which is the information set by Date T. The initial state $\boldsymbol{\xi}_0$ has a known distribution, say the stationary distribution when $\boldsymbol{c}_t, \boldsymbol{F}_t, \boldsymbol{Q}_t$ are not time-varying and satisfy stability conditions.

The flexible SSM is a moderate generalization of the standard model. Let $\boldsymbol{\xi}_t$ be a $m_t \times 1$ latent state vector and $\boldsymbol{Y}_t$ be a $n_t \times 1$ measurement vector. They are simultaneously determined by lagged observations and up to first-lagged states:

$$\begin{aligned} \boldsymbol{\xi}_t &= f_t\left(\boldsymbol{Y}_1^{t-1}\right) + \boldsymbol{F}_t \boldsymbol{\xi}_{t-1} + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{Y}_t &= g_t\left(\boldsymbol{Y}_1^{t-1}\right) + \boldsymbol{H}_t \boldsymbol{\xi}_t + \boldsymbol{J}_t \boldsymbol{\xi}_{t-1} + \boldsymbol{u}_t. \end{aligned} \tag{2}$$

where $f_t(\cdot), g_t(\cdot)$ are two linear or non-linear functions that maps the information set of Date $t-1$ into $\mathbb{R}^{m_t}$ and $\mathbb{R}^{n_t}$ respectively. In some applications of the flexible SSM, the contemporaneous correlation between $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ is essential. Note that if we included $\boldsymbol{Y}_t$ in the transition equation, the state and measurement vectors would be perfectly symmetric and distinguished solely by their observability. Doing so will not change the Kalman filter, but we are not aware of an application of that, so we do not put $\boldsymbol{Y}_t$ in the transition equation.

The flexible model has three features, though each of them seems trivial at the first glance.

First, both the state and measurement vectors can change dimensions at each date. The time-varying dimensions (TVD) of $\boldsymbol{Y}_t$ is well understood and implemented in practice. For example, the missing values in $\boldsymbol{Y}_t$ components lead to reduced the size of the measurement vector at Date t. If $\boldsymbol{Y}_t$ is completely missing, the updating step in the Kalman filter is effectively skipped (see Jones, 1980; Harvey and Pierse, 1984). The TVD of $\boldsymbol{\xi}_t$ has been under-appreciated in the literature until recently. Jungbacker et al. (2011) consider a dynamic factor model with missing data. Common factors and idiosyncratic disturbances corresponding to missing data at Date t and/or $t-1$ are put in the state vector. Since the number of missing data varies over time, the state vector has TVD. Chan et al. (2011) explore TVD in a different setting. The model switches to a more parsimonious representation at random dates controlled by hidden Markov-switching regimes. This is a dynamic mixture model with stochastic dimension changes. Our paper is closer to Jungbacker et al. (2011) in that the dimension changes at deterministic dates. The model per se does not involve dynamic dimensions, we only rewrite it into a parsimonious form with a TVD state vector.

Second, historical observations $\boldsymbol{Y}_{t-1}, ..., \boldsymbol{Y}_1$ can affect $\boldsymbol{\xi}_t, \boldsymbol{Y}_t$. It is a well-understood that this feature will not change the Kalman filter. The setup of the state space model in Hamilton (1994, p.372 - 373) includes an $A'x_t$ term in the measurement equation. Hamilton mentions "x_t could include lagged values of $\boldsymbol{y}$ ...", though no application of such feature is provided in the book. In fact, lagged variables in the system are most useful when they are used together with TVD feature. Suppose we intend to write $g_t(\cdot)$ as a function of p lagged values $\boldsymbol{Y}_{t-1}, ..., \boldsymbol{Y}_{t-p}$, we will encounter problems handling the initial observations since $\boldsymbol{Y}_0, \boldsymbol{Y}_{-1}, ...$ are not observed. TVD offers two solutions. One is to adjust the size of the state vector. Put unobserved lagged variables in the state vector and remove them when they become available. The other is to adjust the size of the measurement vector. No measurement variables are in use at Date $1, ..., p-1$ but they are used together at Date p.

Allowing lagged observations in the states transition equation is rarely seen in the literature. Some may argue that the modeling philosophy of the SSM is to keep the state vector Markovian – summarizing the entire history into states of the last period. This argument is not entirely relevant for our model, for we never introduce high-order lagged states $\boldsymbol{\xi}_{t-2}, \boldsymbol{\xi}_{t-3}, ...$ in the system, but only allow observables $\boldsymbol{Y}_1^{t-1}$ affecting $\boldsymbol{\xi}_t$. In the Kalman filter, the state $\boldsymbol{\xi}_t$ updated conditional on $\boldsymbol{Y}_1^{t-1}$ as well as the new information $\boldsymbol{Y}_t$. Technically, introducing $f_t\left(\boldsymbol{Y}_1^{t-1}\right)$ does not change the filter since it is treated as a constant conditional on $\boldsymbol{Y}_1^{t-1}$. However, this feature substantially enriches the dependence structure of the SSM. In the standard SSM, $\boldsymbol{\xi}_t$ has a law of motion independent of $\boldsymbol{Y}_t$. If we cast a time series model into Eqs. (1), we must ensure the state vector can evolve in a self-sufficient manner. This often entails larger size of the state vector by including variables that we do observe. However, in the flexible SSM the state vector may temporarily disappear, but reappear later relying on $f_t\left(\boldsymbol{Y}_1^{t-1}\right)$. Missing data and mixed frequency regressions illustrate this feature, which will be discussed in Section 5 and 6.

Third, $\boldsymbol{Y}_t$ is determined not only by the current states $\boldsymbol{\xi}_t$ but also by first-lagged states $\boldsymbol{\xi}_{t-1}$. This feature effectively downsizes the state vector without affecting the Kalman filter. A simple application of this feature is a local-location model such that

$$\mu_t = \mu_{t-1} + \varepsilon_t,$$
$$Y_t - \mu_t = \phi\left(Y_{t-1} - \mu_{t-1}\right) + v_t,$$

where μ_t is the latent local location. Rewrite this model into the standard SSM requires a two-dimension state vector, say $(\mu_t, \mu_{t-1})'$ with the measurement variable Y_t. However, the local-location model itself is readily a flexible SSM with the single state μ_t.

Another immediate application of the third feature is the dynamic factor model. Let $\boldsymbol{Y}_t$ be a vector of time series observations, determined by a vector of common factors $\boldsymbol{f}_t$ and idiosyncratic terms $\boldsymbol{v}_t$ such that

$$\boldsymbol{Y}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{v}_t.$$

Suppose both common factors and idiosyncratic components follow AR(1) processes

$$\boldsymbol{f}_t = \boldsymbol{F}\boldsymbol{f}_{t-1} + \boldsymbol{\varepsilon}_t,$$
$$\boldsymbol{v}_t = \boldsymbol{\Phi}\boldsymbol{v}_{t-1} + \boldsymbol{u}_t.$$

The measurement equation can be rewritten as

$$\boldsymbol{Y}_t = \boldsymbol{\Phi Y}_{t-1} + \boldsymbol{\Lambda f}_t - \boldsymbol{\Phi\Lambda f}_{t-1} + \boldsymbol{u}_t.$$

Clearly, it is already in the flexible state space form, though a standard SSM requires doubling the length of the state vector by stacking $\left(\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}\right)'$.

We want to emphasis the fact that each single feature is trivial and has limited usage, but when these features are combined together, the state space representation can take on a parsimonious and intuitive form.

3. The filtering procedure

The procedure presented below is essentially the Kalman filter. We focus on why the three features of the flexible SSM does not change the filter but further extension will modify the filter. The forward recursion consists of the prediction step and update step in a recursive manner. The starting point is an assumption on the distribution of the initial state. Assume $\boldsymbol{\xi}_0 \sim N\left(\boldsymbol{c}_0, \boldsymbol{Q}_0\right)$. Before the information of Date 1 comes in, the information set $\boldsymbol{Y}_1^0$ is empty, so that $\boldsymbol{\xi}_0 \left|\boldsymbol{Y}_1^0 \sim N\left(\widehat{\boldsymbol{\xi}}_{0|0}, \boldsymbol{P}_{0|0}\right)\right.$, where $\widehat{\boldsymbol{\xi}}_{0|0} = \boldsymbol{c}_0$, $\boldsymbol{P}_{0|0} = \boldsymbol{Q}_0$.

At Date t $(t = 1, ..., T)$, we first predict $\boldsymbol{\xi}_t$ and $\boldsymbol{Y}_t$ conditional on the information set of Date $t-1$. Rewrite Eqs. (2) as

$$\begin{pmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{Y}_t \end{pmatrix} = \begin{bmatrix} f_t\left(\boldsymbol{Y}_1^{t-1}\right) \\ g_t\left(\boldsymbol{Y}_1^{t-1}\right) + \boldsymbol{H}_t f_t\left(\boldsymbol{Y}_1^{t-1}\right) \end{bmatrix} + \begin{pmatrix} \boldsymbol{F}_t \\ \boldsymbol{H}_t\boldsymbol{F}_t + \boldsymbol{J}_t \end{pmatrix} \boldsymbol{\xi}_{t-1} + \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{H}_t\boldsymbol{\varepsilon}_t + \boldsymbol{u}_t \end{pmatrix}.$$

Clearly, introducing the term $\boldsymbol{J}_t\boldsymbol{\xi}_{t-1}$ into the measure equation (i.e., the third feature of the flexible model) does not add complexity to the SSM in that $\boldsymbol{J}_t\boldsymbol{\xi}_{t-1}$ is merged into $\boldsymbol{H}_t\boldsymbol{F}_t\boldsymbol{\xi}_{t-1}$. It follows that

$$\begin{pmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{Y}_t \end{pmatrix} \Big| \boldsymbol{Y}_1^{t-1} \sim N\left[\begin{pmatrix} \widehat{\boldsymbol{\xi}}_{t|t-1} \\ \widehat{\boldsymbol{Y}}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \boldsymbol{P}_{t|t-1} & \boldsymbol{L}_{t|t-1} \\ \boldsymbol{L}'_{t|t-1} & \boldsymbol{D}_{t|t-1} \end{pmatrix}\right],$$

where

$$\begin{aligned}
\widehat{\boldsymbol{\xi}}_{t|t-1} &= f_t\left(\boldsymbol{Y}_1^{t-1}\right) + \boldsymbol{F}_t\widehat{\boldsymbol{\xi}}_{t-1|t-1}, \\
\widehat{\boldsymbol{Y}}_{t|t-1} &= g_t\left(\boldsymbol{Y}_1^{t-1}\right) + \boldsymbol{H}_t\widehat{\boldsymbol{\xi}}_{t|t-1} + \boldsymbol{J}_t\widehat{\boldsymbol{\xi}}_{t-1|t-1}, \\
\boldsymbol{P}_{t|t-1} &= \boldsymbol{F}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{F}'_t + \boldsymbol{Q}_t, \\
\boldsymbol{D}_{t|t-1} &= \boldsymbol{H}_t\boldsymbol{P}_{t|t-1}\boldsymbol{H}'_t + \boldsymbol{R}_t + \boldsymbol{J}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{J}'_t + \boldsymbol{H}_t\boldsymbol{F}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{J}'_t \\
&\quad + \boldsymbol{J}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{F}'_t\boldsymbol{H}'_t + \boldsymbol{H}_t\boldsymbol{S}_t + \boldsymbol{S}'_t\boldsymbol{H}'_t, \\
\boldsymbol{L}_{t|t-1} &= \boldsymbol{P}_{t|t-1}\boldsymbol{H}'_t + \boldsymbol{F}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{J}'_t + \boldsymbol{S}_t.
\end{aligned}$$

Clearly, introducing the terms $f_t\left(\boldsymbol{Y}_1^{t-1}\right)$ and $g_t\left(\boldsymbol{Y}_1^{t-1}\right)$ into the model (i.e., the second feature of the flexible model) does not add complexity to the SSM in that they are predetermined conditional on the information set of Date $t-1$. Then we update $\boldsymbol{\xi}_t$ conditional on $\boldsymbol{Y}_t$ and $\boldsymbol{Y}_1^{t-1}$. It follows that $\boldsymbol{\xi}_t \left| \boldsymbol{Y}_1^t \right. \sim N\left(\widehat{\boldsymbol{\xi}}_{t|t}, \boldsymbol{P}_{t|t}\right)$, where

$$\begin{aligned}
\widehat{\boldsymbol{\xi}}_{t|t} &= \widehat{\boldsymbol{\xi}}_{t|t-1} + \boldsymbol{L}_{t|t-1}\left(\boldsymbol{D}_{t|t-1}\right)^{-1}\left(\boldsymbol{Y}_t - \widehat{\boldsymbol{Y}}_{t|t-1}\right), \\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{L}_{t|t-1}\left(\boldsymbol{D}_{t|t-1}\right)^{-1}\boldsymbol{L}'_{t|t-1}.
\end{aligned}$$

This completes a recursion cycle and the filter proceeds to the next period. One can also rewrite the recursion formulas in terms of the Kalman gain and Riccati equation by plugging $\widehat{\boldsymbol{\xi}}_{t|t}$ and $\boldsymbol{P}_{t|t}$ back into $\widehat{\boldsymbol{\xi}}_{t+1|t}$ and $\boldsymbol{P}_{t+1|t}$. Once the filter goes through the entire sample periods, we obtain the likelihood function in its prediction error decomposition form, namely $\prod_{t=1}^{T}\phi\left(\boldsymbol{Y}_t;\widehat{\boldsymbol{Y}}_{t|t-1},\boldsymbol{D}_{t|t-1}\right)$, where $\phi\left(\boldsymbol{x};\mu,\boldsymbol{\Sigma}\right)$ is the the density of $N\left(\mu,\boldsymbol{\Sigma}\right)$.

The TVD state and measurement vectors only reflect in the varying size of matrixes at each date, while the recursion formula itself does not change. It is also possible that at some date we have no state or measurement vector, which can be interpreted as a zero dimension column vector (i.e., a 0×1 vector). As long as a programming platform adopts the conformable matrix algebra for empty matrixes[2], the above formula remains the same, though it can be expressed in a simplified manner.

If $\boldsymbol{\xi}_t$ has zero dimension, $\widehat{\boldsymbol{\xi}}_{t|t-1}$, $\boldsymbol{P}_{t|t-1}$,$\boldsymbol{L}_{t|t-1}$, $\widehat{\boldsymbol{\xi}}_{t|t}$, $\boldsymbol{P}_{t|t}$ are empty while $\widehat{\boldsymbol{Y}}_{t|t-1} = g_t\left(\boldsymbol{Y}_1^{t-1}\right) + \boldsymbol{J}_t\widehat{\boldsymbol{\xi}}_{t-1|t-1}$ and $\boldsymbol{D}_{t|t-1} = \boldsymbol{R}_t + \boldsymbol{J}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{J}_t'$. In other words, the prediction and update on $\boldsymbol{\xi}_t$ are skipped. Note that in the next period, the predicting and updating steps can be conducted normally since $\boldsymbol{Y}_t$ may pass on its value to $\boldsymbol{\xi}_{t+1}$, that is, $\boldsymbol{\xi}_{t+1} = f_{t+1}\left(\boldsymbol{Y}_1^t\right) + \boldsymbol{\varepsilon}_{t+1}$.

If $\boldsymbol{Y}_t$ has zero dimension, $\widehat{\boldsymbol{Y}}_{t|t-1}$, $\boldsymbol{D}_{t|t-1}$, $\boldsymbol{L}_{t|t-1}$ are empty while $\widehat{\boldsymbol{\xi}}_{t|t} = \widehat{\boldsymbol{\xi}}_{t|t-1}$ and $\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1}$. In other words, due to no information at Date t, we can only update the latent states by making a one-period-ahead prediction.

---

[2] An $m \times n$ matrix is said to be empty if either $m = 0$ or $n = 0$ (or both). The matrix algebra for empty matrixes is defined as follows: i) a $0 \times m$ matrix times an $m \times n$ matrix yields a $0 \times n$ matrix. ii) a $m \times 0$ matrix times a $0 \times n$ matrix yields a $m \times n$ matrix of zeros; iii) the summation of two $0 \times m$ matrixes yields a $0 \times m$ matrix. For example, let $\boldsymbol{\xi}_{t-1}$ be a $m \times 1$ vector, $\boldsymbol{\xi}_t$ and $\boldsymbol{\varepsilon}_t$ be 0×1 vectors, $\boldsymbol{F}_t$ be $0 \times m$ matrix. It follows that $\boldsymbol{F}_t\boldsymbol{\xi}_{t-1}$ has the dimension 0×1 and $\boldsymbol{F}_t\boldsymbol{\xi}_{t-1} + \boldsymbol{\varepsilon}_t$ leads to a 0×1 vector, which is conformable with $\boldsymbol{\xi}_t$. Further assume $\boldsymbol{Y}_t$ is a $n \times 1$ vector and $\boldsymbol{H}_t$ is a $n \times 0$ matrix. It follows that $\boldsymbol{H}_t\boldsymbol{\xi}_t$ is a $n \times 1$ vector of zeros, whose size is conformable with $\boldsymbol{Y}_t$.

In the likelihood evaluation, $\boldsymbol{Y}_t$ of zero dimension is omitted.

Lastly, despite the innocuous inclusion of $\boldsymbol{J}_t\boldsymbol{\xi}_{t-1}$ in the measurement equation, attempting to include more lags such as $\boldsymbol{\xi}_{t-2}, \boldsymbol{\xi}_{t-3}$ in the transition and/or measurement equation will non-trivially alter the forward recursion. This is because the forward recursion only keeps track of $\boldsymbol{\xi}_{t-1}\left|\boldsymbol{Y}_1^{t-1}\right.$ but not $\boldsymbol{\xi}_{t-2}\left|\boldsymbol{Y}_1^{t}\right.,\ \boldsymbol{\xi}_{t-3}\left|\boldsymbol{Y}_1^{t}\right.$. It does not mean we cannot apply the filter, for we can modify the filter either by adding a backward recursion (smoothing) at each date or by tracking the joint distribution of $\boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_{t-2}, \boldsymbol{\xi}_{t-3}\left|\boldsymbol{Y}_1^{t-1}\right.$, which is equivalent to tripling the dimension of the state vector. Either solution increases the computational complexity of the filter and thus is not further pursued. If a practitioner does encounter high-order lags in their model, a quick solution is simply to stack multiple-period states into a big state vector.

4. The state space form of ARMA

One prominent application of the Kalman filter in statistics is to evaluate the likelihood function of an ARMA process. Let $\{Z_t\}$ be a univariate $ARMA\,(p, q)$ process

$$Z_t = c + \sum_{i=1}^{p} \phi_i Z_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$

where the disturbances are Gaussian white noises $N\left(0, \sigma^2\right)$. There are various ways to write an ARMA model into its state space form. In Akaike (1973, 1974) and Jones (1980), the state vector is chosen as the projection of $Z_t, Z_{t+1,\ldots,}Z_{t+r-1}$ on the information set of Date t, where $r \equiv \max\left(p, q+1\right)$. The measurement equation is simply an extraction of the first element of the state vector. Hamilton (1994) explores the fact the the lagged sum of an AR process is an ARMA process. The state vector keeps track of r recent values of a latent $AR(p)$ process with coefficients $\phi_1, ..., \phi_p$. The measurement variable Z_t is the sum of these r recent values weighed by $1, \theta_1, ..., \theta_q, 0, ..., 0$. In the state space representation of Harvey and Phillips (1979), the transition matrix is the transpose of Hamilton's. By a backward substitution from the last element to the first element of the state vector, one can see the representation captures the ARMA process. de Jong and Penzer (2004) extend the idea of Pearlman (1980) and discuss a canonical form of the state space model in which the length of the state vector is reduced to $\max(p, q)$.

Our flexible state space representation of the ARMA model distinguishes from the above well-known SSMs in three aspects. First, it is more parsimonious. The state vector only has q dimensions except for the initial p periods when the state vector has dynamic dimensions. In most applications, $T - p$ is much larger than p, handling initial distributions accounts for a fraction of the total computation. Second, it is more general. The well-known SSMs are mostly suitable for stationary ARMA processes and the initial values typically come from the steady states. However, our representation can more conveniently handle other types of initial distribution and time-varying parameters. Third, it is more intuitive. Latent states simply consist of the disturbance terms ε_t and some unobserved initial values in the structural model.

Suppose the observables are Z_1^T. Let $\boldsymbol{W}_t = (Z_t, ..., Z_{t-p+1}, \varepsilon_t, ..., \varepsilon_{t-q+1})'$, $t = 0, ..., T$. Since the data generating process of Z_1 depends on the unobservable $\boldsymbol{W}_0$, we must first specify the distribution of $\boldsymbol{W}_0$. The ARMA literature distinguishes the exact likelihood and the conditional likelihood. The exact likelihood approach assumes $\boldsymbol{W}_0$ is conformable with the stationary distribution of ARMA process. The conditional likelihood treats either $\boldsymbol{W}_0$ or $\boldsymbol{W}_p$ as deterministic. See Hamilton (1994, p.132) and Box and Jenkins (1976, p.211). The well-known SSMs are all suitable for exact likelihood evaluation, but apparently have difficulty handling the conditional likelihood since the states are not expressed in terms of Z_t or ε_t. The flexible SSM accommodates both exact and conditional likelihood as special cases by properly specifying the initial distributions. There are two methods to cast an ARMA in the flexible SSM. One explores the TVD state vector, the other mainly resorts to the TVD measurement vector. We refer to them model 1 and 2 respectively.

Denote $\boldsymbol{\Phi} = (\phi_1, ..., \phi_p)$, $\boldsymbol{\Theta} = (\theta_1, ..., \theta_q)$, and construct an $i \times (i+1)$ matrix $\boldsymbol{E}_i = \begin{pmatrix} \boldsymbol{I}_i & \mathbf{0}_{i,1} \end{pmatrix}$.

Model 1: Assume $\boldsymbol{W}_0 \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Let the state vector be $\boldsymbol{\xi}_t = (Z_0, ..., Z_{t-p+1}, \varepsilon_t, ..., \varepsilon_{t-q+1})'$. By assumption, $\boldsymbol{\xi}_0 \equiv \boldsymbol{W}_0 \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Note that the length of the state vector decreases every period until Date p. After that the state vector only contains structural disturbances $\boldsymbol{\xi}_t = (\varepsilon_t, ..., \varepsilon_{t-q+1})'$.

For Date $t = 1, ..., p$, the transition equation is given by

$$\boldsymbol{\xi}_t = \begin{pmatrix} \boldsymbol{E}_{p-t} & \mathbf{0}_{p-t,q} \\ \mathbf{0}_{1,p-t+1} & \mathbf{0}_{1,q} \\ \mathbf{0}_{q-1,p-t+1} & \boldsymbol{E}_{q-1} \end{pmatrix} \boldsymbol{\xi}_{t-1} + \begin{pmatrix} \mathbf{0}_{p-t,1} \\ \varepsilon_t \\ \mathbf{0}_{q-1,1} \end{pmatrix},$$

and the measurement equation is given by

$$Z_t = c + \sum_{i=1}^{t-1} \phi_i Z_{t-i} + \begin{pmatrix} \mathbf{0}_{1,p-t} & 1 & \mathbf{0}_{1,q-1} \end{pmatrix} \boldsymbol{\xi}_t + (\phi_t, ..., \phi_p, \boldsymbol{\Theta})\, \boldsymbol{\xi}_{t-1}.$$

Note that at Date $t = p$, $\boldsymbol{E}_{p-t}, \mathbf{0}_{p-t,q}, \mathbf{0}_{p-t,p-t}, \mathbf{0}_{1,p-t}$ are empty, but the formula still applies.

For Date $t = p+1, ..., T$, the dynamic system becomes simpler

$$\boldsymbol{\xi}_t = \begin{pmatrix} \mathbf{0}_{1,q} \\ \boldsymbol{E}_{q-1} \end{pmatrix} \boldsymbol{\xi}_{t-1} + \begin{pmatrix} \varepsilon_t \\ \mathbf{0}_{q-1,1} \end{pmatrix},$$
$$Z_t = c + \sum_{i=1}^{p} \phi_i Z_{t-i} + \begin{pmatrix} 1 & \mathbf{0}_{1,q-1} \end{pmatrix} \boldsymbol{\xi}_t + \boldsymbol{\Theta}\boldsymbol{\xi}_{t-1}.$$

Suppose the initial distribution of $\boldsymbol{W}_0$ is known (as in the case of the conditional likelihood), we can immediately apply Model 1. However, we often do not explicitly specify an initial distribution but require $\boldsymbol{W}_0$ coming from the stationary distribution (as in the case of the exact likelihood). Unlike the fixed-dimension SSM, Model 1 cannot automatically generate a stationary initial distribution due to the shrinking size of the state vector. The easiest way to enable stationary distribution generation is to slightly modify Model 1 by temporarily expanding $\boldsymbol{\xi}_1$ by one dimension. To be exact, let $\boldsymbol{\xi}_1 = \boldsymbol{W}_1$ and $\boldsymbol{\xi}_1 = \boldsymbol{c}_1 + \boldsymbol{F}_1\boldsymbol{\xi}_0 + \widetilde{\varepsilon}_1$, where $\boldsymbol{c}_1 = \begin{pmatrix} c \\ \mathbf{0}_{p+q-1,1} \end{pmatrix}$, $\boldsymbol{F}_1 = \begin{pmatrix} \boldsymbol{\Phi} & \boldsymbol{\Theta} \\ \boldsymbol{E}_{p-1} & \mathbf{0}_{p-1,q} \\ \mathbf{0}_{1,p} & \mathbf{0}_{1,q} \\ \mathbf{0}_{q-1,p} & \boldsymbol{E}_{q-1} \end{pmatrix}$, $\widetilde{\varepsilon}_1 = \begin{pmatrix} \varepsilon_t & \mathbf{0}_{1,p-1} & \varepsilon_t & \mathbf{0}_{1,q-1} \end{pmatrix}'$. Then the stationary distribution can be generated by

$$E(\boldsymbol{\xi}_0) = \left(\boldsymbol{I}_{(p+q)} - \boldsymbol{F}_1\right)^{-1} \boldsymbol{c}_1, \qquad (3)$$
$$vec\left[Var(\boldsymbol{\xi}_0)\right] = \left(\boldsymbol{I}_{(p+q)^2} - \boldsymbol{F}_1 \otimes \boldsymbol{F}_1\right)^{-1} vec(\boldsymbol{Q}_1).$$

where $\boldsymbol{Q}_1$ is the covariance matrix of $\widetilde{\varepsilon}_1$, that is, a $(p+q) \times (p+q)$ matrix of zeros except for $(1,1)$, $(1,p+1)$, $(p+1,1)$,$(p+1,p+1)$ elements being σ^2.

Model 1 takes advantage of the TVD state vector by only including those relevant but unobserved variables at each date, but the measurement variable is always the scalar Z_t. There is an alternative way to represent an ARMA process with the TVD measurement vector. Here the initial values are specified in terms of $\boldsymbol{W}_p$ instead of $\boldsymbol{W}_0$. The alternative representation is ideal for two scenarios. First, we intend to evaluate the exact likelihood and have obtained the distribution of $\boldsymbol{W}_0$ from Eqs. (3). Stationarity implies $\boldsymbol{W}_p$ has the same distribution as $\boldsymbol{W}_0$. Second, we intend to find the conditional likelihood for a given distribution of $\boldsymbol{W}_p$ such as a deterministic one. The idea of this representation is to treat the initial values $Z_p, ..., Z_1$ as a whole, so that there is no need to keep track of $Z_0, Z_{-1}, ...$ as latent states. To see this, let the measurement variable $\boldsymbol{Y}_t$ be empty for $t = 1, ..., p-1$, and at Date p let $\boldsymbol{Y}_p = (Z_p, ..., Z_1)'$ and the state vector be $\boldsymbol{\xi}_p = (\varepsilon_p, ..., \varepsilon_{p-q+1})'$. At Date p the filter starts from the predictive distribution of $\begin{pmatrix} \boldsymbol{Y}_p \\ \boldsymbol{\xi}_p \end{pmatrix} \big| \boldsymbol{Y}_1^{p-1}$, which has the same distribution as $\boldsymbol{W}_p$. As long as we properly specify $\boldsymbol{Q}_p$, $\boldsymbol{R}_p$ and $\boldsymbol{S}_p$ so as to replicate the covariance of $\boldsymbol{W}_p$, the recursion from Date 1 to $p-1$ becomes irrelevant. This method leads to greater parsimony of the state vector. The details are specified below.

Model 2: Assume $\boldsymbol{W}_p \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The flexible SSM is given by Eqs. (2), with the following state and measurement vector and coefficients:

For $t = 1, ..., p-1$, let $\boldsymbol{\xi}_t, \boldsymbol{Y}_t$ be empty.

For $t = p$, let $\boldsymbol{\xi}_t = (\varepsilon_p, ..., \varepsilon_{p-q+1})'$, $\boldsymbol{Y}_t = (Z_p, ..., Z_1)'$.

Partition $\boldsymbol{\mu}$ into $\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$ with the length p and q respectively. Similarly partition $\boldsymbol{\Sigma}$ into $\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$. Let $f_t\left(\boldsymbol{Y}_1^{t-1}\right) = \boldsymbol{\mu}_2$, $\boldsymbol{Q}_t = \boldsymbol{\Sigma}_{22}$, $g_t\left(\boldsymbol{Y}_1^{t-1}\right) = \boldsymbol{\mu}_1$, $\boldsymbol{R}_t = \boldsymbol{\Sigma}_{11}$, $\boldsymbol{S}_t = \boldsymbol{\Sigma}_{21}$ and $\boldsymbol{F}_t, \boldsymbol{H}_t, \boldsymbol{J}_t$ be empty.

For $t = p+1, ..., T$, the state variables, measurement variables and coefficients are the same as those in Model 1.

In summary, Model 1 and 2 have the same specification from Date $p+1$ to T, which are the main body of the state space model. The main body has fixed-length state and measurement vectors as well as time-invariant parameters. The state vector only includes recent q disturbance terms, keeping track of the MA part of the series. The AR part is predetermined and thus treated as if it were a constant in the measurement equation. The TVD state and measurement vectors are only employed to handle the initial distribution.

In the flexible SSM, the predicted and smoothed latent states have structural interpretations, even the distribution of the initial states are of theoretical interest since it provides an exact solution to the autocovariance function of an ARMA process.

Pick an arbitrary t, let $\mu = E(Z_t)$, $\gamma_j = E[(Z_t - \mu)(Z_{t-j} - \mu)]$, $\delta_j = E[(Z_t - \mu)\varepsilon_{t-j}]$. Clearly $\delta_j = 0$, $\forall j < 0$. Note that $(\gamma_0, ..., \gamma_{p-1}, \delta_0, ..., \delta_{q-1})$ can be read directly from the first row of $Var(\boldsymbol{\xi}_0)$ in Eqs. (3). It follows that the analytic expression of the ARMA autocovariance function is

$$\gamma_j = \sum_{i=1}^{p} \phi_i \gamma_{j-i} + \delta_{-j} + \sum_{i=1}^{q} \theta_i \delta_{-j+i}, \forall j \geq p.$$

5. Mixed frequency regression

One feature of the flexible state space model is that lagged observations can affect current states, allowing richer dynamics between the states and observables. We illustrate its usage by a mixed frequency Vector Autoregression (VAR) model. Macroeconomic data are not observed at a uniformed frequency. For example, the best available data of GDP is quarterly, while that of the unemployment rate is monthly. If a VAR includes both variables, we may interpret the quarterly GDP data as the sum of latent "monthly GDP". Temporal aggregation in the state space framework has been explored by Zadrozny (1988), Mittnik and Zadrozny (2004), Mariano and Murasawa (2003, 2010), Hyung and Granger (2008). For illustration purposes, consider a bivariate $VAR(1)$ model operated at the semi-annual frequency

$$\boldsymbol{Z}_t = \boldsymbol{c} + \boldsymbol{\Phi}\boldsymbol{Z}_{t-1} + \boldsymbol{\varepsilon}_t,$$

or in the expanded form

$$\begin{pmatrix} Z_{1,t} \\ Z_{2,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}.$$

where $\boldsymbol{\varepsilon}_t$ are Gaussian white noises with the variances $\boldsymbol{\Sigma} \equiv \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$. Assume the initial values come from the stationary distribution: $\boldsymbol{Z}_0 \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega})$, where $\boldsymbol{\mu} = (\boldsymbol{I}_2 - \boldsymbol{\Phi})^{-1}\boldsymbol{c}$, $\boldsymbol{\Omega} = (\boldsymbol{I}_4 - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi})^{-1} vec(\boldsymbol{\Sigma})$.

Though $\{Z_{2,t}\}$ is fully observed, we do not have semi-annual data on $\{Z_{1,t}\}$. Instead we observe annual data $\overline{Z}_{1,t} = Z_{1,t-1} + Z_{1,t}$, $t = 2, 4, 6, ..., T$. For simplicity, T is assumed to be an even number.

To write this model into a standard SSM, we need a four-dimension state vector keeping track of the two variates in recent two periods. Let $\boldsymbol{\xi}_t = (Z_{1,t}, Z_{2,t}, Z_{1,t-1}, Z_{2,t-1})'$. The transition equation can be written as

$$\boldsymbol{\xi}_t = \begin{pmatrix} \boldsymbol{c} \\ \mathbf{0}_{2,1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Phi} & \mathbf{0}_{2,2} \\ \boldsymbol{I}_2 & \mathbf{0}_{2,2} \end{pmatrix} \boldsymbol{\xi}_{t-1} + \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0}_{2,1} \end{pmatrix},$$

and $E(\boldsymbol{\xi}_0) = \begin{pmatrix} \boldsymbol{\mu}' & \boldsymbol{\mu}' \end{pmatrix}'$, $vec\,[Var\,(\boldsymbol{\xi}_0)] = (\boldsymbol{I}_{16} - \boldsymbol{F}_1 \otimes \boldsymbol{F}_1)^{-1}\, vec\,(\boldsymbol{Q}_1)$.

The measurement equation (with fixed-dimension observations) in Date $t = 1, 3, ..., T-1$ is given by[3]

$$\begin{pmatrix} 0 \\ Z_{2,t} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \boldsymbol{\xi}_t,$$

and in Date $t = 2, 4, ..., T$ is given by

$$\begin{pmatrix} \overline{Z}_{1,t} \\ Z_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \boldsymbol{\xi}_t.$$

Though this is a valid representation, the state vector is lengthy in that some observed variables are put as states. The flexible SSM only admits unobserved variables in the state vector. Let $\boldsymbol{\xi}_0 = (Z_{1,0}, Z_{2,0})$ and $\xi_t = Z_{1,t}$ for all $t = 1, ..., T$.

[3]The first element of measurement vector is set to zero as that the measurement vector has fixed length. Alternatively, one can fill in the first element by some exogenous random variable whose data generating processe is unrelated with model parameters so that the likelihood is only shifted by a constant (see Mariano and Murasawa, 2003). The only advantage of introducing such artificial random variables is to keep constant the size of the measurement vector.

For $t = 1$, the transition and measurement equations are given by

$$\xi_1 = c_1 + \begin{pmatrix} \phi_{11} & \phi_{12} \end{pmatrix} \boldsymbol{\xi}_0 + \varepsilon_{1,1},$$
$$Z_{2,1} = c_2 + \begin{pmatrix} \phi_{21} & \phi_{22} \end{pmatrix} \boldsymbol{\xi}_0 + \varepsilon_{2,1}.$$

For $t = 3, 5..., T-1$, the dynamic equations are

$$\xi_t = c_1 + \phi_{12} Z_{2,t-1} + \phi_{11}\xi_{t-1} + \varepsilon_{1,t},$$
$$Z_{2,t} = c_2 + \phi_{22} Z_{2,t-1} + \phi_{21}\xi_{t-1} + \varepsilon_{2,t}.$$

For $t = 2, 4, ..., T$, the transition takes the same form as that in odd-numbered dates, but the measurement equations have two dimensions

$$\begin{pmatrix} \overline{Z}_{1,t} \\ Z_{2,t} \end{pmatrix} = \begin{pmatrix} 0 \\ c_2 + \phi_{22} Z_{2,t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \xi_t + \begin{pmatrix} 1 \\ \phi_{21} \end{pmatrix} \xi_{t-1} + \begin{pmatrix} 0 \\ \varepsilon_{2,t} \end{pmatrix}.$$

In the standard SSM, the state vector has four dimensions and the coefficient matrixes contain many zeros and ones, which slows down the filter due to excessive overheads such as $0 * 0$. Worse still, to compute the covariance matrix of the initial state, we need to work on a 16×16 matrix and its inversion. However, the flexible SSM only keeps track of the scalar $Z_{1,t}$ as the state vector. The state and measurement equations simply replicate the original $VAR(1)$ process and the aggregation constraints.

6. Dynamic factor model with missing data

Factor models have wide applications in macroeconomic forecasting (e.g., Stock and Watson, 2002; Forni et al., 2003; Schumacher, 2007), monetary policy analysis (Bernanke et al., 2005; Stock and Watson, 2005) and business cycle transmission study (Eickmeier, 2007). We adopt the likelihood-based inference on a dynamic factor model where large amount of observations are driven by a few common factors. Economic data are not perfect and possibly a fraction of observations are missing. We consider a factor model with randomly missing data similar to Jungbacker et al. (2011), but propose a more parsimonious state space representation.

Let $\boldsymbol{Y}_t$ be a $n \times 1$ vector of time series observations, determined by a $m \times 1$ vector of common factors $\boldsymbol{f}_t$ and idiosyncratic terms $\boldsymbol{v}_t$ such that

$$\boldsymbol{Y}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{v}_t. \tag{4}$$

Both common factors and idiosyncratic components follow AR(1) processes such that

$$\boldsymbol{f}_t = \boldsymbol{F} \boldsymbol{f}_{t-1} + \boldsymbol{\varepsilon}_t,$$
$$\boldsymbol{v}_t = \boldsymbol{\Phi} \boldsymbol{v}_{t-1} + \boldsymbol{u}_t,$$

where $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{Q})$ and $\boldsymbol{u}_t \sim N(\mathbf{0}, \boldsymbol{R})$ are white noises.

The term $\boldsymbol{v}_t$ can be squeezed out of the measurement equation so that $\boldsymbol{Y}_t$ is determined by its lagged values and lagged factors:

$$\boldsymbol{Y}_t = \boldsymbol{\Phi} \boldsymbol{Y}_{t-1} + \boldsymbol{\Lambda} \boldsymbol{f}_t - \boldsymbol{\Phi} \boldsymbol{\Lambda} \boldsymbol{f}_{t-1} + \boldsymbol{u}_t, \tag{5}$$

We follow the notations of Jungbacker et al. (2011) in handling missing data in $\boldsymbol{Y}_t$. Consider some $n \times 1$ vector $\boldsymbol{Z}_t$. The vector $\boldsymbol{Z}_t(\boldsymbol{o}_s)$ contains all elements of $\boldsymbol{Z}_t$ that correspond to observed entries in $\boldsymbol{Y}_s$ $(t, s = 1, ..., T)$. In other words, $\boldsymbol{o}_s$ is a logical index indicating the observed entries in $\boldsymbol{Y}_s$ and we use $\boldsymbol{o}_s$ to select corresponding elements in $\boldsymbol{Z}_t$. Similarly, $\boldsymbol{Z}_t(\boldsymbol{m}_s)$ contains all elements of $\boldsymbol{Z}_t$ that correspond to missing entries in $\boldsymbol{Y}_s$. We can also use the logical indexes to extract corresponding rows and/or columns of a $n \times n$ matrix $\boldsymbol{A}$. For example, $\boldsymbol{A}(\boldsymbol{o}_s, :)$ denotes row selections, $\boldsymbol{A}(:, \boldsymbol{o}_s)$ denotes column selections, and $\boldsymbol{A}(\boldsymbol{m}_s, \boldsymbol{o}_s)$ denotes both row and column selections.

In principle, we can track both $\boldsymbol{\xi}_t$ and $\boldsymbol{v}_t$ as latent states and straightforwardly write the model with missing data into the state space form. However, $\boldsymbol{v}_t$ is of length n, which is typically much larger than m. It is unfavorable to work on an SSM with a high-dimension state vector. Jungbacker et al. (2011) solve this problem by putting a fraction of $\boldsymbol{v}_t$ into the state vector. For those entries observed in both $\boldsymbol{Y}_t$ and $\boldsymbol{Y}_{t-1}$, Eq. (5) is employed to characterize the measurement equation. Otherwise, the measurement equation is switched to Eq. (4).

Our flexible SSM only relies on Eq. (5) as the measurement equation and $\boldsymbol{v}_t$ never enters the state vector. Recall the idea of our flexible SSM is solely including those relevant but unobserved variables in the state vector. Whenever an element in $\boldsymbol{Y}_t$ is observed, it is put in the measurement equation. Whenever it is missing, it enters the state vector. It follows that the state vector consists of $\boldsymbol{f}_t$ and $\boldsymbol{Y}_t(\boldsymbol{m}_t)$. The measurement vector is simply $\boldsymbol{Y}_t(\boldsymbol{o}_t)$. To find out the transition and measurement equations, we first rewrite Eq. (5) as

$$\boldsymbol{Y}_t = \boldsymbol{\Phi}\boldsymbol{Y}_{t-1} + \boldsymbol{J}\boldsymbol{f}_{t-1} + \boldsymbol{w}_t, \tag{6}$$

where $\boldsymbol{J} = \boldsymbol{\Lambda}\boldsymbol{F} - \boldsymbol{\Phi}\boldsymbol{\Lambda}$, $\boldsymbol{w}_t = \boldsymbol{\Lambda}\boldsymbol{\varepsilon}_t + \boldsymbol{u}_t$. $\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{w}_t \end{pmatrix} \sim N\left[\boldsymbol{0}, \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{Q}\boldsymbol{\Lambda}' \\ \boldsymbol{\Lambda}\boldsymbol{Q} & \boldsymbol{\Lambda}\boldsymbol{Q}\boldsymbol{\Lambda}' + \boldsymbol{R} \end{pmatrix}\right]$.

Note that $\boldsymbol{Y}_{t-1}$ can be decomposed into $\boldsymbol{Y}_{t-1}(\boldsymbol{o}_{t-1})$ and $\boldsymbol{Y}_{t-1}(\boldsymbol{m}_{t-1})$. Eq. (6) implies that $\boldsymbol{Y}_t$ is determined by $\boldsymbol{Y}_{t-1}(\boldsymbol{o}_{t-1})$, $\boldsymbol{Y}_{t-1}(\boldsymbol{m}_{t-1})$ and $\boldsymbol{f}_{t-1}$.

The first one is predetermined, while the last two are exactly the state vector of Date $t-1$. Furthermore, $\boldsymbol{Y}_t$ can be decomposed into observed $\boldsymbol{Y}_t(\boldsymbol{o}_t)$ and unobserved $\boldsymbol{Y}_t(\boldsymbol{m}_t)$. In a symmetric manner, we put $\boldsymbol{Y}_t(\boldsymbol{o}_t)$ in the measurement equation and $\boldsymbol{Y}_t(\boldsymbol{m}_t)$ in the transition equation. It follows that the measurement equation is given by

$$\begin{aligned}\boldsymbol{Y}_t(\boldsymbol{o}_t) &= \boldsymbol{\Phi}(\boldsymbol{o}_t, \boldsymbol{o}_{t-1})\boldsymbol{Y}_{t-1}(\boldsymbol{o}_{t-1}) \\ &\quad + \begin{bmatrix} \boldsymbol{J}(\boldsymbol{o}_t, :) & \boldsymbol{\Phi}(\boldsymbol{o}_t, \boldsymbol{m}_{t-1}) \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{Y}_{t-1}(\boldsymbol{m}_{t-1}) \end{bmatrix} + \boldsymbol{w}_t(\boldsymbol{o}_t),\end{aligned}$$

and the transition equation is given by

$$\begin{aligned}\begin{pmatrix} \boldsymbol{f}_t \\ \boldsymbol{Y}_t(\boldsymbol{m}_t) \end{pmatrix} &= \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\Phi}(\boldsymbol{m}_t, \boldsymbol{o}_{t-1})\boldsymbol{Y}_{t-1}(\boldsymbol{o}_{t-1}) \end{pmatrix} \\ &\quad + \begin{bmatrix} \boldsymbol{F}_t & \boldsymbol{0} \\ \boldsymbol{J}(\boldsymbol{m}_t, :) & \boldsymbol{\Phi}(\boldsymbol{m}_t, \boldsymbol{m}_{t-1}) \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{Y}_{t-1}(\boldsymbol{m}_{t-1}) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{w}_t(\boldsymbol{m}_t) \end{bmatrix}.\end{aligned}$$

In this application, we critically explore the third feature of the flexible SSM. Introduction of the first-lagged state vector in the measurement equation not only avoids tracking $\boldsymbol{f}_t, \boldsymbol{f}_{t-1}$ as latent states but also grants $\boldsymbol{Y}_t(\boldsymbol{o}_t)$ access to $\boldsymbol{Y}_{t-1}(\boldsymbol{m}_{t-1})$. Compared with the state space representation of Jungbacker et al. (2011), our flexible SSM represents the same process but has some advantages. First, our state vector is shorter. Suppose $\boldsymbol{Y}_t$ has k_1 missing entries, $\boldsymbol{Y}_{t-1}$ has k_2 distinct missing entries (entries that are missing in both periods are counted once). Our state vector is of length $m+k_1$ while that in Jungbacker et al. (2011) is $2m+k_1+k_2$. Second, our formulation puts no restriction on $\boldsymbol{\Phi}$. The transition equation presented in Jungbacker et al. (2011) is based on a diagonal $\boldsymbol{\Phi}$ so that, say, $\boldsymbol{v}_t(\boldsymbol{m}_{t-1})$ only depends on $\boldsymbol{v}_{t-1}\,(\boldsymbol{m}_{t-1})$ rather than the whole $\boldsymbol{v}_{t-1}$. For a non-diagonal $\boldsymbol{\Phi}$, the states transition would become cumbersome. Third, elements in our state vector need not to be reshuffled in the states transition. In Jungbacker et al. (2011), a selection matrix is employed to re-order the states to facilitate transition. Fourth, our representation is intuitive. The transition and measurement equation are symmetric and they largely resemble Eq. (4) and Eq. (6). The elements in $\boldsymbol{Y}_t$, no matter as the states or observables, always fetch all elements of $\boldsymbol{Y}_{t-1}$ partially from the past observations and partially from the previous states.

7. Conclusion

In the standard SSM, the state vector is detached from the measurement vector due to its own autoregressive law of motion. The measure vector is viewed as a noise-ridden representation of the latent states. The asymmetric treatment of the states and observations often entails a lengthy state vector when a structural model is cast into the state space form.

In this paper, the SSM is examined from a new angle. Our SSM is flexible mainly because of the symmetry of the state and measurement vectors as well as two-way dynamics. This feature merits concise translation from a structure model to its state space form. Relevant but unobserved variables in the structural model are placed in the state vector while all observables are in the measurement vector. The number of unobserved/observed variables often varies over time, so the length of the state/measurement vector is also time-varying. Intuitive representation is the main attraction of the flexible state space form.

Despite different interpretations of system dynamics between the standard and flexible SSM, the same Kalman filter can be applied to both. In the flexible SSM, the state vector is shorter in length and the parameter matrixes have fewer axillary elements such as zeros and ones. Therefore, the Kalman filter is expected to run faster. Computational efficiency offers another attraction of the flexible state space form.

Akaike, H., 1973. Maximum likelihood identification of gaussian autoregressive moving average models. Biometrika 60 (2), 255–265.

Akaike, H., 1974. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. Annals of the Institute of Statistical Mathematics 26, 363–387.

Basdevant, O., 2003. On applications of state-space modelling in macroeconomics. Reserve Bank of New Zealand Discussion Paper Series.

Bernanke, B., Boivin, J., Eliasz, P. S., 2005. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. The Quarterly Journal of Economics 120 (1), 387–422.

Box, G. E. P., Jenkins, G. M., 1976. Time Series Analysis: Forecasting and Control. Holden Day, San Francisco.

Chan, J., Koop, G., Leon-Gonzalez, R., Strachan, R., 2011. Time varying dimension models. Working papers, University of Strathclyde Business School, Department of Economics.

de Jong, P., Penzer, J., 2004. The arma model in state space form. Statistics and Probability Letters 70 (1), 119–125.

Durbin, J., Koopman, S. J., 2001. Time Series Analysis by State Space Methods. Oxford University Press.

Eickmeier, S., 2007. Business cycle transmission from the us to germany–a structural factor approach. European Economic Review 51 (3), 521–551.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2003. Do financial variables help forecasting inflation and real activity in the euro area? Journal of Monetary Economics 50 (6), 1243–1255.

Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press: Princeton.

Harvey, A. C., 1989. Time Series Analysis. Forecasting, Structural Time Series Models and the Kalman Filter.

Harvey, A. C., Phillips, G. D. A., 1979. Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika 66 (1), 49–58.

Harvey, A. C., Pierse, R. G., 1984. Estimating missing observations in economic time series. Journal of the American Statistical Association 79 (385), 125–131.

Hyung, N., Granger, C. W., 2008. Linking series generated at different frequencies. Journal of Forecasting 27 (2), 95–108.

Jones, R. H., 1980. Maximum likelihood fitting of arma models to time series with missing observations. Technometrics 22 (3), 389–395.

Jungbacker, B., Koopman, S., van der Wel, M., 2011. Maximum likelihood estimation for dynamic factor models with missing data. Journal of Economic Dynamics and Control 35 (8), 1358–1368.

Kalman, R. E., 1960. A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME,Journal of Basic Engineering, 35–45.

Mariano, R. S., Murasawa, Y., 2003. A new coincident index of business cycles based on monthly and quarterly series. Journal of Applied Econometrics 18 (4), 427–443.

Mariano, R. S., Murasawa, Y., 2010. A coincident index, common factors, and monthly real gdp. Oxford Bulletin of Economics and Statistics 72 (1), 27–46.

Mittnik, S., Zadrozny, P. A., 2004. Forecasting quarterly german gdp at monthly intervals using monthly ifo business conditions data. CESifo Working Paper Series.

Pearlman, J. G., 1980. An algorithm for the exact likelihood of a high-order autoregressive-moving average process. Biometrika 67 (1), 232–233.

Schumacher, C., 2007. Forecasting german gdp using alternative factor models based on large datasets. Journal of Forecasting 26 (4), 271–302.

Stock, J. H., Watson, M. W., 2002. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics 20 (2), 147–62.

Stock, J. H., Watson, M. W., 2005. Implications of dynamic factor models for var analysis. NBER Working Papers 11467, National Bureau of Economic Research, Inc.

Zadrozny, P., 1988. Gaussian likelihood of continuous-time armax models when data are stocks and flows at different frequencies. Econometric Theory 4 (01), 108–124.