Recursive online estimation for time-varying VAR paramaters - a short introduction

Suhasini Subba Rao

1 Time-varying AR processes

Before we discuss the asymptotic properties of your algorithm, we start by considering the time-varying AR(1) process. We explain the rescalar $'N'$ and show that for a given realisation (given set of observations) it can be ignored; it is simply a tool to help with the asymptotics. Suppose $\{Y_t\}_t$ satisfies

$$Y_t = aY_{t-1} + \varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\}$ are iid random variables and $|a| < 1$. Since a is constant over time $\{X_t\}$ is a stationary process. We can easily "nonstationarify" the process by making a time-dependent. Suppose $\{Y_t\}$ satisfies the model

$$Y_t = a_t Y_{t-1} + \varepsilon_t, \tag{2}$$

where $\sup_t |a_t| < (1 - \delta)$ $(\delta > 0)$. Now $\{Y_t\}$ is a nonstationary process. This is clear if we consider the covariances;

$$\mathrm{cov}\{Y_t, Y_{t+s}\} = \sum_{k=0}^{\infty} a_t^k a_{t+s}^{k+s} \qquad s > 0.$$

If $t \neq t'$, then $\mathrm{cov}\{Y_t, Y_{t+s}\} \neq \mathrm{cov}\{Y_{t'}, Y_{t'+s}\}$. Since the covariance is not invariant over time (that is $\mathrm{cov}(Y_t, Y_{t+s})$ depends on the position t) then $\{Y_t\}$ is a nonstationary process.

Suppose that a_t changes slowly with time. That is $a_t \approx a_{t+1}$. In which case, we assume there exists a function $a : [0, 1] \to \mathbb{R}$ such that for each t, $a_t = a(\frac{t}{N})$, where N is the rescaler. Therefore we can rewrite (2) as

$$Y_t = a(\frac{t}{N})Y_{t-1} + \varepsilon_t.$$

To make life easier, we let $Y_t = Y_{t,N}$ and

$$Y_{t,N} = a(\frac{t}{N})Y_{t-1,N} + \varepsilon_t. \tag{3}$$

However (2) and (3) are the same.

1.1 The time-varying vector AR process

Now we consider the tvVAR(p) process. Suppose the q-dimensional stochastic vector process $\{\underline{X}_t\}$ satisfies the model

$$\begin{aligned}\underline{X}_t &= \sum_{i=1}^{p} A_{t,i}\underline{X}_{t-i} + \Sigma_t\underline{\varepsilon}_t \\ &= \mathcal{A}_t\underline{Y}_{t-1} + \Sigma_t\underline{\varepsilon}_t, \end{aligned} \tag{4}$$

where $\underline{Y}_t^T = (\underline{X}_t^T, \ldots, \underline{X}_{t-p+1}^T)$ and $\mathcal{A}_t = (A_{t,1}, \ldots, A_{t,p})$ and $\{\underline{\varepsilon}_t\}$ are iid zero mean random vectors with $\mathrm{var}(\underline{\varepsilon}_t) = \Delta$.

As in Section 1.1, we introduce N into the system. We assume that $A_{t,i}$ changes slowly with t, that is $A_{t,i} \approx A_{t+1,i}$. And there exists a matrix function $A_i : [0.1] \to \mathbb{R}^{q\times q}$, such that $A_{t,i} = A_i(\frac{t}{N})$. Hence

$$\underline{X}_t = \sum_{i=1}^{p} A_i(\frac{t}{N})\underline{X}_{t-i} + \Sigma(\frac{t}{N})\underline{\varepsilon}_t.$$

As in Section 1.1, we let $\underline{X}_t = \underline{X}_{t,N}$ and

$$\begin{aligned}\underline{X}_{t,N} &= \sum_{i=1}^{p} A_i(\frac{t}{N})\underline{X}_{t-i,N} + \Sigma(\frac{t}{N})\underline{\varepsilon}_t \\ &= \mathcal{A}(\frac{t}{N})\underline{Y}_{t-1,N} + \Sigma(\frac{t}{N})\underline{\varepsilon}_t,\end{aligned}$$

where $A(\frac{t}{N}) = (A_1(\frac{t}{N}), \ldots, A_p(\frac{t}{N}))$ and $\underline{Y}_{t,N}^T = (\underline{X}_{t,N}^T, \ldots, \underline{X}_{t-p+1,N}^T)$.

1.2 Definitions

In the following section we will study the asymptotic properties of the estimators. We will use the following definitions. Suppose the vector process $\underline{X}_t(u)$ satisfies

$$\begin{aligned}\underline{X}_t(u) &= \sum_{i=1}^{p} A_i(u)\underline{X}_{t-i}(u) + \Sigma(u)\underline{\varepsilon}_t \\ &= \mathcal{A}(u)\underline{Y}_{t-1}(u) + \Sigma(u)\underline{\varepsilon}_t,\end{aligned}$$

where $\underline{Y}_t(u)^T = (\underline{X}_t(u)^T, \ldots, \underline{X}_{t-p+1}(u)^T)$. For fixed u, $\{\underline{X}_t(u)\}_t$ is a stationary process. This process is related to the tv-VAR(p) process (4), that you have considered. Fix t and suppose the process $\{\tilde{\underline{X}}_s(t)\}_s$ satisfies the model

$$\tilde{\underline{X}}_s(t) = \sum_{i=1}^{p} A_{t,i}\tilde{\underline{X}}_{s-i}(t)$$

Recall that t is fixed, so the parameters are not changing over the time. This implies that the process $\{\tilde{\underline{X}}_s(t)\}_s$ is a stationary process. If $u = t/N$ we have that $\tilde{\underline{X}}_s(t) = \underline{X}_s(u)$.

We now define the stationary covariance at location u. We will use this as the reference point in our asymptotic results later on. Let

$$R(u) = \mathrm{var}\{\underline{Y}_t(u)\}.$$

To get some feeling about $R(u)$, we can show if $u = t/N$, then

$$R(u) = \mathrm{var}\{\underline{Y}_t(u)\} \approx \mathrm{var}\{\underline{Y}_{t,N}\}.$$

1.3 An online estimator of the parameters

In this section we define an online estimator of the tv-VAR parameters and study their asymptotic properties.

We define the exponentially weighted least squares criterion:

$$\mathcal{L}_{t,N}(\mathcal{A}) = \sum_{j=0}^{t-1} \lambda(1-\lambda)^j \big\{\underline{X}_{t-j,N} - \mathcal{A}\underline{Y}_{t-1-j,N}\big\}^T \big\{\underline{X}_{t-j,N} - \mathcal{A}\underline{Y}_{t-1-j,N}\big\}. \tag{5}$$

We use $\hat{\mathcal{A}}_{t,N}$ as an estimator $\mathcal{A}(\frac{t}{N})$, where

$$\hat{\mathcal{A}}_{t,N} = \arg\min_{\mathcal{A}} \mathcal{L}_{t,N}(\mathcal{A}) \tag{6}$$

and $\hat{\mathcal{A}}_{t,N}$ is a $q \times pq$ matrix.

Equivalently $\hat{\mathcal{A}}_{t,N}$ can be obtained by using the recursive algorithm

$$\hat{\mathcal{A}}_{t,N} = \hat{\mathcal{A}}_{t-1,N} + \lambda\big\{\underline{X}_{t,N} - \hat{\mathcal{A}}_{t-1,N}\underline{Y}_{t-1,N}\big\}\underline{Y}_{t-1,N}^T \hat{R}_{t-1,N}^{-1} \tag{7}$$

where $\hat{R}_{t-1,N}$ is a $pq \times pq$-dimensional matrix which estimates the Hessian and is defined by

$$\hat{R}_{t-1,N} = \sum_{k=0}^{t-2} \lambda(1-\lambda)^k \underline{Y}_{t-k-1,N}\underline{Y}_{t-k-1,N}^T.$$

Therefore (6) and (7) are the same estimators.

Remark 1.1 *Equivalently we can use the algorithm*

$$\hat{\mathcal{A}}_t = \hat{\mathcal{A}}_{t-1} + \big\{\underline{X}_t - \hat{\mathcal{A}}_{t-1}\underline{Y}_{t-1}\big\}\underline{Y}_{t-1}^T \hat{R}_{t-1}^{-1}, \tag{8}$$

where

$$\hat{R}_{t-1} = \sum_{k=0}^{t-2} (1-\lambda)^k \underline{Y}_{t-k-1}\underline{Y}_{t-k-1}^T.$$

However, the algorithms (7) and (8) are the same. The differences are that (i) equation (7) includes the rescaling term N, but we have shown in the sections above that this is only there for the asymptotics and need not be there, (ii) equation (7) contains a λ whereas (8) does not contain λ - however it is straightforward to show that the λs cancel. Therefore the algorithms (7) and (8) are the same.

1.4 The asymptotic properties

In this section we study the asymptotic properties of the recursive online estimator (7). Since (6) and (7) are the same, we can do this by considering the exponential least squares estimator (6). We can show that the estimator is weakly consistent, but here we shall concentrate on showing asymptotic normality of the estimator. We will give an outline of the proof. be the asymptotic normality result.

First let us consider the estimator $\hat{\mathcal{A}}_{t,N}$. We said above that it was an estimator of $A(\frac{t}{N})$ or equivalently an estimator of A_t (if we do not want to worry about N). However, the exponential kernel is not symmetric, therefore $\hat{\mathcal{A}}_{t,N}$ is a better estimator of $A(u-\frac{1}{N\lambda})$ or equivalently $A_{t-(1/\lambda)}$.

To prove asymptotic normality we need to consider the gradient of the likelihood function at the true parameter $(A(u-\frac{1}{N\lambda}))$. Suppose $f:\mathbb{R}^p\to\mathbb{R}$ and let $\nabla f^T=(\frac{\partial f}{\partial x_1},\ldots,\frac{\partial f}{\partial x_p})$.

We now consider the gradient of $\mathcal{L}_{t,N}(\mathcal{A})$ at $\mathcal{A}(u-\frac{1}{N\lambda})$ (where $u=t/N$). Partitioning $\nabla\mathcal{L}_{t,N}\{\mathcal{A}(u-\frac{1}{N\lambda})$ we have that

$$\nabla\mathcal{L}_{t,N}\{\mathcal{A}(u-\frac{1}{N\lambda})\}=\mathcal{S}_{t,N}+\mathcal{B}_{t,N}, \tag{9}$$

where

$$\begin{aligned}\mathcal{S}_{t,N} &= -2\sum_{k=0}^{t-1}\lambda(1-\lambda)^k\{\underline{X}_{t-k,N}-\mathcal{A}(\frac{t-k}{N})\underline{Y}_{t-k-1,N})\}\underline{Y}_{t-k-1,N}^T\\ &= -2\sum_{k=0}^{t-1}\lambda(1-\lambda)^k\Sigma(\frac{t}{N})\underline{\varepsilon}_t\underline{Y}_{t-k-1,N}^T\end{aligned}$$

and

$$\mathcal{B}_{t,N}=2\sum_{k=0}^{t-1}\lambda(1-\lambda)^k\{\mathcal{A}(u-\frac{1}{N\lambda})-\mathcal{A}(\frac{t-k}{N})\}\underline{Y}_{t-k-1,N}\underline{Y}_{t-k-1,N}^T.$$

Let

$$\Omega(u)=\mathrm{var}\{\Sigma(u)\varepsilon_t\}=\Sigma(u)\Delta\Sigma(u)^T.$$

If $u=t/N$, we can show that

$$\lambda^{-1/2}\mathcal{S}_{t,N}\xrightarrow{\mathcal{D}}MVN_{q\times pq}\Big(0_{q\times pq},\Omega(u)\otimes R(u)\Big), \tag{10}$$

where $N\to\infty$, $1/(\lambda N)\to 0$ as $N\to\infty$.

Let $\mathcal{A}'(u)$ denote the derivative of every element of $\mathcal{A}$ with respect to u. Therefore,

$$\mathcal{A}'(u-\frac{1}{N\lambda})_{ij}=\frac{\partial\mathcal{A}(x)_{ij}}{\partial x}\rfloor_{x=(u-\frac{1}{N\lambda})}.$$

By considering the bias $\mathcal{B}_{t,N}$ we can show that

$$\mathcal{B}_{t,N} = \left[2\mathcal{A}'(u - \frac{1}{\lambda N})R'(u - \frac{1}{\lambda N}) - \mathcal{A}''(u - \frac{1}{\lambda N})\right]\frac{1}{(\lambda N)^2} + o_p(\frac{1}{(\lambda N)}), \tag{11}$$

where $R(\cdot)$ is the variance-covariance matrix defined in (5).

Remark 1.2 *Please check my calculations:*

$$\sum_{k=0}^{t-1} \lambda(1-\lambda)^k(\frac{k}{N} - \frac{1}{N\lambda})^2$$
$$= \int_0^\infty \frac{1}{b}\exp(-x/b)\{x^2 - 2bx + b^2\}dx + O(\frac{1}{N}) = b^2 + O(\frac{1}{N}) = \frac{1}{(\lambda N)^2} + O(\frac{1}{N}),$$

where $b = 1/(\lambda N)$.

We can also show that

$$\nabla^2\mathcal{L}_{t,N} \xrightarrow{\mathcal{P}} R(u). \tag{12}$$

In other words $\nabla^2\mathcal{L}_{t,N} \approx \text{var}\{\underline{Y}_{t,N}\} = \text{var}\{\underline{Y}_t\}$.

Therefore since $\nabla\mathcal{L}_{t,N}(\hat{\mathcal{A}}_{t,N}) = 0$ (it is the minimum of the least squares criterion $\mathcal{L}_{t,N}$) by using a Taylor series expansion or by simply expanding $\nabla\mathcal{L}_{t,N}$ we have that

$$\begin{aligned}\nabla\mathcal{L}_{t,N}\{\hat{\mathcal{A}}_{t,N}\} &= \nabla\mathcal{L}_{t,N}\{\mathcal{A}(\frac{t}{N})\} + \{\hat{\mathcal{A}}_{t,N} - \mathcal{A}(\frac{t}{N})\}\nabla^2\mathcal{L}_{t,N} \\ \{\hat{\mathcal{A}}_{t,N} - \mathcal{A}(\frac{t}{N})\}\nabla^2\mathcal{L}_{t,N} &= \nabla\mathcal{L}_{t,N}\{\mathcal{A}(\frac{t}{N})\}.\end{aligned}$$

By substituting (9) into the above and using (10), (11) and (12) we have that

$$\lambda^{-1/2}\{\hat{\mathcal{A}}_{t,N} - \mathcal{A}(\frac{t}{N})\} - \lambda^{-1/2}\mathcal{B}(u)R(u)^{-1} \xrightarrow{\mathcal{D}} MVN_{q\times pq}\left(0_{q\times pq}, \Omega(u)\otimes R(u)^{-1}\right), \tag{13}$$

where

$$\mathcal{B}(u) = \left[2\mathcal{A}'(u - \frac{1}{\lambda N})R'(u - \frac{1}{\lambda N}) - \mathcal{A}''(u - \frac{1}{\lambda N})\right]\frac{1}{(\lambda N)^2}.$$

We now summarise (without the use of the rescaling factor N) (14) in the theorem below. Strictly speaking it is not a theorem - but simply a means of using (14) in practice.

Theorem 1 *Suppose $\hat{\mathcal{A}}_t$ is defined as in (8). Then we have*

$$\lambda^{-1/2}\{\hat{\mathcal{A}}_t - \mathcal{A}_t\} - \lambda^{-1/2}\mathcal{B}_t R_t^{-1} \xrightarrow{\mathcal{D}} MVN_{q\times pq}\left(0_{q\times pq}, \Omega_t \otimes R_t^{-1}\right), \tag{14}$$

where

$$\mathcal{B}_t = \left[2\mathcal{A}'_{t-\frac{1}{\lambda}}R'_{t-\frac{1}{\lambda}} - \mathcal{A}''_{t-\frac{1}{\lambda}}\right]\frac{1}{(\lambda N)^2},$$

$R_t = var\{\underline{Y}_t\}$ *and*

$$\Omega_t = var\{Y_t\}\Delta var\{Y_t\}.$$

Remark 1.3 *(i) The theorem above involves $R'_{t-\frac{1}{\lambda}}$ and $\mathcal{A}''_{t-\frac{1}{\lambda}}$. Of course a formal derivative of these quantities do not exist - since $\mathcal{A}_t$ and R_t are only over discrete time. However, these derivative can be defined as the first and second differences:*

$$\begin{aligned} R'_{t-\frac{1}{\lambda}} &= R_{t-\frac{1}{\lambda}} - R_{t-\frac{1}{\lambda}-1} \\ \mathcal{A}'_{t-\frac{1}{\lambda}} &= \mathcal{A}_{t-\frac{1}{\lambda}} - \mathcal{A}_{t-\frac{1}{\lambda}-1} \\ \mathcal{A}''_{t-\frac{1}{\lambda}} &= \mathcal{A}_{t-\frac{1}{\lambda}+1} - 2\mathcal{A}_{t-\frac{1}{\lambda}} + \mathcal{A}_{t-\frac{1}{\lambda}-1}, \end{aligned}$$

which can be estimated from the observations.

(ii) We can use $\hat{R}_t$ as an estimator of R_t, where

$$\hat{R}_t = \sum_{j=0}^{t-1} \lambda(1-\lambda)\underline{Y}_{t-j}\underline{Y}^T_{t-j}.$$

We can use $\hat{\mathcal{A}}_{t-\frac{1}{N\lambda}}$, defined in (7), as an estimator of $\mathcal{A}_{t-\frac{1}{N\lambda}}$. And

$$\begin{aligned} \hat{R}'_{t-\frac{1}{\lambda}} &= \hat{R}_{t-\frac{1}{\lambda}} - \hat{R}_{t-\frac{1}{\lambda}-1} \\ \hat{\mathcal{A}}'_{t-\frac{1}{\lambda}} &= \hat{\mathcal{A}}_{t-\frac{1}{\lambda}} - \hat{\mathcal{A}}_{t-\frac{1}{\lambda}-1} \\ \hat{\mathcal{A}}''_{t-\frac{1}{\lambda}} &= \hat{\mathcal{A}}_{t-\frac{1}{\lambda}+1} - 2\hat{\mathcal{A}}_{t-\frac{1}{\lambda}} + \hat{\mathcal{A}}_{t-\frac{1}{\lambda}-1}, \end{aligned}$$

as estimators of the derivatives.

(iii) The theorem above can be used to test for the significance of a parameter.