Online appendix for "Estimating high-dimensional Markov-switching VARs"

By Kenwin Maung

1. Proofs

This section contains the proofs for the following results in the main text.

1.1. Proof for Lemma 1

Proof. (A.2) follows from taking the derivative of (A.1), so it suffices to show the latter. First, write $\nabla^1 \log \mathcal{L}_T(\phi_T^*) = \sum_{t=p_T+1}^{T} \nabla^1 \log \mathcal{L}(y_t|\mathcal{I}_{t-p_T}^{t-1};\phi_T^*)$. By the Louis missing information principle (Louis, 1982)[1], we can express the first derivative with respect to the i^{th} element in ϕ_T^* as

$$\begin{aligned}
\nabla_i^1 \log \mathcal{L}(y_t|\mathcal{I}_{t-p_T}^{t-1};\phi_T^*) = \sum_{k=t-p_T}^{t-1} \Bigg\{ & E\left[\nabla_i^1 \log p_{(S_{k-1})(S_k)}(\phi_T^*) + \nabla_i^1 \log g(y_k|\mathcal{I}_{k-p_T}^{k-1};\Phi_{S_k,T}^*)\middle|\mathcal{I}_{t-p_T}^{t}\right] \\
& - E\left[\nabla_i^1 \log p_{(S_{k-1})(S_k)}(\phi_T^*) + \nabla_i^1 \log g(y_k|\mathcal{I}_{k-p_T}^{k-1};\Phi_{S_k,T}^*)\middle|\mathcal{I}_{t-p_T}^{t-1}\right]\Bigg\} \\
& + E\left[\nabla_i^1 \log g(y_t|\mathcal{I}_{t-p_T}^{t-1};\Phi_{S_t,T}^*)\middle|\mathcal{I}_{t-p_T}^{t}\right] \\
& + E\left[\nabla_i^1 \log p_{t-p_T}(\phi_T^*)\middle|\mathcal{I}_{t-p_T}^{t}\right] - E\left[\nabla_i^1 \log p_{t-p_T}(\phi_T^*)\middle|\mathcal{I}_{t-p_T}^{t-1}\right]. \qquad (1)
\end{aligned}$$

Then, using (1), the law of iterated expectation, and the assumption that $g(y_t|\cdot)$ is a valid probability density function, we have that $E[\nabla_i^1 \log \mathcal{L}(y_t|\mathcal{I}_{t-p_T}^{t-1};\phi_T^*)|\mathcal{I}_{t-p_T}^{t-1}] = 0$, and therefore (A.1) follows. $\square$

1.2. Proof for Lemma 2

Proof. Under assumptions (A1)-(A4), we can employ Lemmas 3-6 from Bickel et al. (1998), with some modification to the conditional density, to show that $\nabla_j^1 \mathcal{L}_T(\phi_T^*) \to 0$ in L_2. Likewise, Lemmas 7-9 of Bickel et al. (1998) also imply that $\nabla_{j,k}^2 \mathcal{L}_T(\phi_T^*)$ is bounded in L_1. However, from the proofs of Lemmas 7-9, we can see that this result will extend to L_2.

---

[1]See, for example, equation (5) in Bickel et al. (1998)

Hence, here we show the result for the third derivative. Similar to what we did in Lemma 1, write $\nabla^3 \log \mathcal{L}_T(\phi_T) = \sum_{t=p_T+1}^{T} \nabla^3 \log \mathcal{L}(y_t|\mathcal{I}_{t-p_T}^{t-1}; \phi_T)$. Again, by the Louis missing information principle, we can get

$$\begin{aligned}
&\nabla^3_{j,k,l} \log \mathcal{L}(y_t|\mathcal{I}_{t-p_T}^{t-1}; \phi_T) \\
&= \nabla^3_{j,k,l} \log \mathcal{L}(\underbrace{y_{t-p_T}, \dots, y_t}_{\equiv y_{t-p_T}^t} |\mathcal{X}_{t-p_T}^{t-1}; \phi_T) - \nabla^3_{j,k,l} \log \mathcal{L}(\underbrace{y_{t-p_T}, \dots, y_{t-1}}_{\equiv y_{t-p_T}^{t-1}} |\mathcal{X}_{t-p_T}^{t-1}; \phi_T),
\end{aligned} \tag{2}$$

where $\mathcal{X}_{t-p_T}^{t-1} = (x_{t-q_T}, \dots, x_{t-1})$,

$$\begin{aligned}
&\nabla^3_{j,k,l} \log \mathcal{L}(y_{t-p_T}^t|\mathcal{X}_{t-p_T}^{t-1}; \phi_T) \\
&\quad= E[\nabla^3_{j,k,l} \log \mathcal{L}(y_{t-p_T}^t, S_{t-p_T}^t|\mathcal{X}_{t-p_T}^{t-1}; \phi_T)|\mathcal{I}_{t-p_T}^t] \\
&\quad+ 2E[\nabla^2_{j,k} \log \mathcal{L}(y_{t-p_T}^t, S_{t-p_T}^t|\mathcal{X}_{t-p_T}^{t-1}; \phi_T)|\mathcal{I}_{t-p_T}^t] \\
&\quad- 2E[\nabla^1_{j} \log \mathcal{L}(y_{t-p_T}^t, S_{t-p_T}^t|\mathcal{X}_{t-p_T}^{t-1}; \phi_T)|\mathcal{I}_{t-p_T}^t] \times E[\nabla^2_{j,k} \log \mathcal{L}(y_{t-p_T}^t, S_{t-p_T}^t|\mathcal{X}_{t-p_T}^{t-1}; \phi_T)|\mathcal{I}_{t-p_T}^t],
\end{aligned} \tag{3}$$

and $S_a^b = (S_a \dots, S_b)$. Note that the last two terms are bounded in L_2 because $\nabla^2_{j,k}\mathcal{L}_T(\phi_T)$ can be expressed as a function of the conditional expectations of first and second derivatives similar to (3) (see (14) in Kasahara and Shimotsu, 2019), and it is bounded in L_2. So we only have to show for the first term (label this $E[\nabla^3|\mathcal{I}_{t-p_T}^t]$). Fortunately, this term results in a similar expression to (1). From (2),

$$\begin{aligned}
&E[\nabla^3|\mathcal{I}_{t-p_T}^t] - E[\nabla^3|\mathcal{I}_{t-p_T}^{t-1}] \\
&= \sum_{n=t-p_T}^{t-1} \left\{ E\left[\nabla^3_{j,k,l} \log p_{(S_{n-1})(S_n)}(\phi_T) + \nabla^3_{j,k,l} \log g(y_n|\mathcal{I}_{n-p_T}^{n-1}; \Phi_{S_n,T}) \middle| \mathcal{I}_{t-p_T}^t \right] \right. \\
&\left. - E\left[\nabla^3_{j,k,l} \log p_{(S_{n-1})(S_n)}(\phi_T) + \nabla^3_{j,k,l} \log g(y_n|\mathcal{I}_{n-p_T}^{n-1}; \Phi_{S_n,T}) \middle| \mathcal{I}_{t-p_T}^{t-1} \right] \right\} \\
&+ E\left[\nabla^3_{j,k,l} \log g(y_t|\mathcal{I}_{t-p_T}^{t-1}; \Phi_{S_t,T}) \middle| \mathcal{I}_{t-p_T}^t \right] \\
&+ E\left[\nabla^3_{j,k,l} \log p_{t-p_T}(\phi_T) \middle| \mathcal{I}_{t-p_T}^t \right] - E\left[\nabla^3_{j,k,l} \log p_{t-p_T}(\phi_T) \middle| \mathcal{I}_{t-p_T}^{t-1} \right].
\end{aligned} \tag{4}$$

With this expression, the proof is completed if we can show the following:

(i) $\left\| E[\nabla^3_{j,k,l} \log g(y_n|\mathcal{I}_{n-p_T}^{n-1}; \Phi_{S_n,T})|\mathcal{I}_{t-p_T}^t] - E[\nabla^3_{j,k,l} \log g(y_n|\mathcal{I}_{n-p_T}^{n-1}; \Phi_{S_n,T})|\mathcal{I}_{t-p_T}^{t-1}] \right\|_2 \le K\beta^T$,

(ii) $\left\|E[\nabla^3_{j,k,l}\log p_{(S_{n-1})(S_n)}(\phi_T)|\mathcal{I}^t_{t-p_T}] - E[\nabla^3_{j,k,l}\log p_{(S_{n-1})(S_n)}(\phi_T)|\mathcal{I}^{t-1}_{t-p_T}]\right\|_2 \leq K\beta^T$,

(iii) $\left\|E[\nabla^3_{j,k,l}\log p_{t-p_T}(\phi_T)|\mathcal{I}^t_{t-p_T}] - E[\nabla^3_{j,k,l}\log p_{t-p_T}(\phi_T)|\mathcal{I}^{t-1}_{t-p_T}]\right\|_2 \leq K\beta^T$,

where $K > 0$ is constant, and $\beta \in [0, 1)$.

This can be shown with a similar approach to that of lemma 6 in Bickel et al. (1998). Under (A1) and (A2), there exists a constant $\nu > 0$ such that $\inf\{p_{(s)(s')}(\phi_T)|s, s', \phi_T \in \Theta_T\} \geq \nu$ and $\inf\{p_s(\phi_T)|s, \phi_T \in \Theta_T\} \geq \nu$. Define $\mu(y_t) = [1 + (M-1)\nu^{-2}\rho(y_t)]^{-1}$ where $\rho(\cdot)$ is defined in (A4). We show the result for (i) only, while the rest follow similarly.

$$
\begin{aligned}
&|E[\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{S_n,T})|\mathcal{I}^t_{t-p_T}] - E[\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{S_n,T})|\mathcal{I}^{t-1}_{t-p_T}]| \\
&= \left|\sum_{s=1}^{M}\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{s,T})\{P(S_n = s|\mathcal{I}^t_{t-p_T}) - P(S_n = s|\mathcal{I}^{t-1}_{t-p_T})\}\right| \\
&\leq \max_s|\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{s,T})|K\prod_{i=n+1}^{t-1}\exp(-2\mu(y_i)),
\end{aligned}
$$

for some constant K, and the last inequality follows from lemma 5 in Bickel et al. (1998). Then, observe that

$$
\begin{aligned}
&\|E[\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{S_n,T})|\mathcal{I}^t_{t-p_T}] - E[\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{S_n,T})|\mathcal{I}^{t-1}_{t-p_T}]\|_2^2 \\
&\leq KE\left[\max_s|\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{s,T})|^2\prod_{i=n+1}^{t-1}\exp(-4\mu(y_i))\right] \\
&= KE\left[E\left[\max_s|\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{s,T})|^2\prod_{i=n+1}^{t-1}\exp(-4\mu(y_i))\middle|S^n_{t-p_T+1}\right]\right] \\
&= KE\left[E[\max_s|\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{s,T})|^2|S_n]\prod_{i=n+1}^{t-1}E[\exp(-4\mu(y_i))|S_i]\right] \\
&\leq K\max_{s'}E[\max_s|\nabla^3_{j,k,l}\log g(y_n|\mathcal{I}^{n-1}_{n-p_T};\Phi_{s,T})|^2|S_n = s']\beta^{t-n-1}
\end{aligned}
$$

where we have used the fact that $0 < \mu(\cdot) < 1$ and the property of the exponential to bind it with $\beta \in [0, 1)$. Together with (A3), we get the result in (i). $\square$

1.3. Proof for Corollary 1

Proof. The proof is similar to the proof of Lemma A.1 in Kwon and Kim (2012).

For notational convenience, let $(A)_{ij}$ indicate the element in the i^{th} row and j^{th} column of A, $\nabla^1 \log \mathcal{L}_T(\phi_T^*) \equiv \nabla^1 \mathcal{L}$, and $\nabla^2 \mathcal{L}_T(\phi_T^*) - E[\nabla^2 \mathcal{L}_T(\phi_T^*)] \equiv \nabla^2 \mathcal{L}$.

For (1),

$$E[\|\nabla^1 \mathcal{L}\|^2] = \sum_{i=1}^{K_T^*} E[(\nabla^1 \mathcal{L})_i^2] = O(TK_T^*),$$

where the last equality is an application of Cauchy-Schwarz, and the L_2 bound in Lemma 2. Then, (2) follows from Chebyshev's inequality.

The proof of (2) is almost identical except for the fact that we are dealing with a matrix so that

$$E[\|\nabla^2 \mathcal{L}\|_{\mathbf{1}}^2] = \sum_{i=1}^{K_T^*} \sum_{j=1}^{K_T^*} E[(\nabla^2 \mathcal{L})_{ij}^2] = O(TK_T^{*2}).$$

(3) follows similarly from (2) since it it just the coordinate-wise version. Likewise, the proof for (4) is identical but with a triple sum instead, and (5) follows. $\square$

References

Bickel, P.J., Ritov, Y., Ryden, T., 1998. Asymptotic normality of the maximum-likelihood estimator for general hidden markov models. The Annals of Statistics 26, 1614–1635.

Kasahara, H., Shimotsu, K., 2019. Asymptotic properties of the maximum likelihood estimator in regime switching econometric models. Journal of econometrics 208, 442–467.

Kwon, S., Kim, Y., 2012. Large sample properties of the scad-penalized maximum likelihood estimation on high dimensions. Statistica Sinica , 629–653.

Louis, T.A., 1982. Finding the observed information matrix when using the em algorithm. Journal of the Royal Statistical Society: Series B (Methodological) 44, 226–233.