Macroeconomic Analysis with Machine Learning and Big Data

Lecture 4: Statistical Model in Macroeconomics (I)

July 9

Instructor: Yucheng Yang *Scribe: Yumin Hu, Lu Yang*

1 Vector Autoregressive Model

1.1 Setup of VAR

Covariance stationary process Y_t satisfies vector autoregressive model of order p (denoted as VAR(p)) if

$$Y_t = \nu + \sum_{i=1}^{p} A_i Y_{t-i} + Z_t$$

where

$$\begin{array}{lcl} Y_t & \in & \mathcal{R}^{n\times 1} \\ \nu & \in & \mathcal{R}^{n\times 1} \\ A_i & \in & \mathcal{R}^{n\times n} \quad i = 1, 2, ..., p \\ Z_t & is & White\ Noise \end{array}$$

Covariance Stationary $(weakly\ stationary)$:

- $\mathbb{E}[Y_{it}^2] < \infty,\ \forall i$
- $\mathbb{E}[Y_t] \triangleq \mu_t,\ Cov(Y_t, Y_{t-l}) \triangleq \Gamma_Y(l),\ \forall l \in \mathcal{Z}\ don`t\ depend\ on\ t$

Strictly stationary:

$$(Y_t, Y_{t+1}, \cdots, Y_{t+k}) \stackrel{d}{=} (Y_{t+l}, Y_{t+l+1}, \cdots, Y_{t+k+l})$$

Z_t is White Noise if it satisfies the following conditions:

$$\mu_Z = 0,\ \Gamma_Z(l) = \begin{cases} \Sigma & l = 0 \\ 0 & l \neq 0 \end{cases}$$

Remark:

$$VAR(p) \longrightarrow VMA(\infty)$$

$$\begin{array}{rcl} Y_t & = & \nu + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + Z_t \\ \Longrightarrow\ Y_t & = & \nu + \sum_{i=1}^{\infty} \Psi_i Z_{t-i} \end{array}$$

Notation:

$$\begin{array}{rcl} Lag\ operator & : & \mathcal{L} \quad e.g.\ \mathcal{L}Y_t = Y_{t-1} \\ Lag\ Polynomial & : & A(\mathcal{L}) = \mathbf{I} - \sum_{i=1}^{p} A_i \mathcal{L}^i \end{array}$$

We can rewrite the model as

$$A(\mathcal{L})Y_t = Z_t.$$

Theorem: If $x \longrightarrow det(A(x))$ has all roots outside unit cycle.$\Longrightarrow \exists$ abs. sum. $A(\mathcal{L})^{-1}$, s.t. Y_t has a unique covariance stationary representation $Y_t = A(\mathcal{L})^{-1}Z_t$.
Proof. The proof of this Theorem is on Brockwell&Davis (Thm 11.3.1).

How to check whether the condition holds? For $p = 1$,

$$A(\mathcal{L}) = \mathcal{I} - A_1\mathcal{L}$$

$$det(A(x)) = 0 \Leftrightarrow |\mathbf{I} - \sum_{i=1}^{p} A_i x| = 0$$

easily calculated as the inverse of eigenvalues.

We can write VAR(p) into the companion form VAR(1) as follows:

$$X_t = (Y_t, Y_{t-1}, \cdots, Y_{t-p+1})^T$$

$$X_{t-1} = (Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p})^T$$

$$X_t = \tilde{A}X_{t-1} + \tilde{Z}_t + \tilde{\nu}$$

1.2 Estimation of VAR

For VAR(p) function

$$Y_t = \nu + \sum_{i=1}^{p} A_i Y_{t-i} + Z_t$$

Assume that $Z_t \sim \mathcal{N}(0, \Sigma)$. And all parameters $(\nu, A_1, \cdots)$ is denoted by θ.

$$P(Y_1, Y_2, \cdots, Y_T|\theta) \overset{factorization}{=} P(Y_1|\theta)\Pi_{t=2}^{T} P(Y_t|\mathcal{Y}_{t-1}, \theta)$$

Where

$$\mathcal{Y}_{t-1} = \{y_t\}_{\tau \leq t-1}$$

$$logP(Y_1, Y_2, \cdots, Y_T|\theta) = -\frac{nT}{2}log2\pi - \frac{1}{2}\sum_{t=1}^{T}[log|\hat{V}_t(\theta)| + \hat{e}_t(\theta)\hat{V}_t(\theta)^{-1}\hat{e}_t(\theta)]$$

Where

$$\begin{array}{rcl} \hat{e}_t(\theta) & = & Y_t - \mathbb{E}[Y_t|\mathcal{Y}_{t-1}, \theta] \\ \hat{V}_t(\theta) & = & Var(\hat{e}_t|\theta) \end{array}$$

Write VAR(p) in a more compact way:

$$Y_t = \beta X_t + Z_t$$

then the CMLE estimators are:

$$\begin{aligned} \hat{\beta}_{CMLE} &= (\textstyle\sum_{t=1}^{T} Y_t X_t^T)(\sum_t X_t X_t^T)^{-1} \\ \hat{\Sigma}_{CMLE} &= \tfrac{1}{T}\textstyle\sum_{t=1}^{T}(Y_t - \hat{\beta}X_t)(Y_t - \hat{\beta}X_t)^T \end{aligned}$$

The biggest problem of VAR: small data sample size, large number of parameters - CMLE perform badly. The most common estimation methods are Bayesian methods. Nice conjugate priors for β:

$$prior : vec(\beta) \sim \mathcal{N}(\beta^*, V_\beta) \quad \Longrightarrow \quad posterior : vec(\beta) \sim \mathcal{N}(\bar{\beta}, \bar{V_\beta}).$$

e.g. Minnesota prior: prior centered at random walk.

1.3 Granger "causality"

- $\{x_t\}$ Granger cause $\{y_t\}$ if $Var^*(y_t|\mathcal{Y}_{t-1}) > Var^*(y_t|\mathcal{Y}_{t-1}, \mathcal{X}_{t-1})$
- It is just about whether it helps in prediction. Not a causality relationship.
- Sims (1972): show that money Granger cause income, not opposite.

2 Structural Vector Autoregressive Model

2.1 Setup of SVAR

- **VAR**

$$\begin{aligned} Y_t &= \nu + \textstyle\sum_{i=1}^{p} A_i Y_{t-i} + \epsilon_t,\ \epsilon_t \sim N(0, \Sigma) \\ \epsilon_t &= Y_t - \mathbb{E}^*(Y_t|\mathcal{Y}_{t-1}) \text{ is one-step ahead forecasting error} \end{aligned}$$

- **SVAR**: Shocks are independent. The SVAR assumes that ϵ_t is a linear combination of the unobserved structural shocks η_t, and structural shocks are assumed to be uncorrelated:
  - $\epsilon_t = H\eta_t, \eta_t \sim \mathcal{N}(0, diag\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2\})$
  - VAR can be written as $Y_t = \nu + \sum_{i=1}^{p} A_i Y_{t-i} + H\eta_t$.
  - SVAR is $B_0 Y_t = C + \sum_{i=1}^{p} B_i Y_{t-i} + \eta_t,\ \eta_t \sim N(0, \Sigma_\eta)$
  - If B_0^{-1} exists $\Longrightarrow Y_t = B_0^{-1}C + \sum_{i=1}^{p} B_0^{-1} B_i Y_{t-i} + B_0^{-1}\eta_t,\ B_0^{-1}\eta_t \sim N(0, B_0^{-1}\Sigma_\eta (B_0^{-1})^T)$
  - After estimating VAR, we need to estimate B_0^{-1} to estimate SVAR.

2.2 Identification

- the estimation of B_0^{-1} requires the user to impose additional identifying restrictions on B_0^{-1} that can be motivated based on economic theory, or other external constraints on structural model. Imposing these additional restrictions allows one to decompose the reduced-form errors ϵ_t into manually uncorrelated structural shocks η_t, with an economic interpretation.
- degree of freedom: $\frac{n(n-1)}{2}$.
  - Total # of unknowns in B_0^{-1}: n^2.
  - Since $B_0^{-1}\Sigma_\eta(B_0^{-1})^T = \Sigma$ and Σ_η is diagonal: $\frac{n(n-1)}{2}$ restrictions.
  - Normalization condition ($\Sigma_\eta = \mathbf{I}$): n restrictions.
- Short-run restriction (Sims, 1980):
  - impose $\frac{n(n-1)}{2}$ zero restrictions on B_0^{-1}.
  - for example, B_0^{-1} is lower triangular matrix, then we can directly get it from Cholesky decomposition.

3 State Space Model

Observation/Measurement equation: $Y_t = g(S_t, u_t;\ \theta)$

State/Transition equation: $S_t = h(S_{t-1}, \epsilon_t;\ \theta)$

- Y_t: n-dim observable
- S_t: m-dim state(unobservable/latent)
- $(u_t^T, \epsilon_t^T) \overset{iid}{\sim} F_\theta,\ \{\epsilon_t\}$ and $\{u_t\}$ are orthogonal.

Filtering Problem: given knowledge of θ, learn S_t from data $\mathcal{Y}_t = \{Y_\tau\}_{\tau=1}^t$

- $P(S_{t-1}|\mathcal{Y}_{t-1}, \theta)$
- forecast $S:\ P(S_t|\mathcal{Y}_{t-1}, \theta) = \int P(S_t|S_{t-1}, \theta)P(S_{t-1}|\mathcal{Y}_{t-1}, \theta)dS_{t-1}$
- forecast $Y_t:\ P(Y_t|\mathcal{Y}_{t-1}, \theta) = \int P(Y_t|S_t, \theta)P(S_t|\mathcal{Y}_{t-1}, \theta)dS_t$
- update via Bayes Rule given Y_t:

$$P(S_t|\mathcal{Y}_t, \theta) = P(S_t|Y_t, \mathcal{Y}_{t-1}, \theta) = \frac{P(S_t, Y_t|\mathcal{Y}_{t-1}, \theta)}{P(Y_t|\mathcal{Y}_{t-1}, \theta)} = \frac{P(S_t|\mathcal{Y}_{t-1}, \theta)P(Y_t|S_t, \theta)}{P(Y_t|\mathcal{Y}_{t-1}, \theta)}$$

Smoothing Problem: $S_t|y_T$, given $P(S_{t+1}|y_T, \theta)$

$$
\begin{aligned}
P(S_t|\mathcal{Y}_T,\theta) &= \int P(S_t|S_{t+1},\mathcal{Y}_T,\theta)P(S_{t+1}|\mathcal{Y}_T,\theta)dS_{t+1} \\
&= \int P(S_t|S_{t+1},\mathcal{Y}_t,\theta)P(S_{t+1}|\mathcal{Y}_T,\theta)dS_{t+1} \\
&= P(S_t|\mathcal{Y}_t,\theta)\int \frac{P(S_{t+1}|S_t,\theta)}{P(S_{t+1}|\mathcal{Y}_t,\theta)}P(S_{t+1}|\mathcal{Y}_T,\theta)dS_{t+1}
\end{aligned}
$$

4 Linear Gaussian State Space Model

Special important case: Dynamic Factor Model

- Observable equation:
$$X_t = \Lambda F_t + \epsilon_t,\ \epsilon_t \in N(0,R)$$
- State equation:
$$F_t = \Phi F_{t-1} + u_t,\ u_t \in N(0,Q)$$
- How to compute the likelihood when $\{F_t\}$ is latent?
  - EM-algorithm (Expectation and Maximization)
  - PCA (Principle Component Analysis)

EM-algorithm

- E-step (Kalman filter and Kalman smoother): given $\Phi, \Lambda, Q, R \Rightarrow \{F_t\}_{t=1}^T \Big| \{X_t\}_{t=1}^T$
  - Denote:
$$
\begin{aligned}
F_t|\mathcal{X}_{t-1} &\sim N(F_{t|t-1}, P_{t|t-1}) \\
X_t|\mathcal{X}_{t-1} &\sim N(X_{t|t-1}, V_{t|t-1}) \\
F_t|\mathcal{X}_t &\sim N(F_{t|t}, P_{t|t})
\end{aligned}
$$
  - Kalman Filter: $F_t|\mathcal{X}_{t-1} \to X_t|\mathcal{X}_{t-1} \to F_t|\mathcal{X}_t \to F_{t+1}|\mathcal{X}_t \to ...$
$$
\begin{array}{rcl}
\hat{F}_{t+1|t} & = & \Phi \hat{F}_{t|t} \\
P_{t+1|t} & = & \Phi P_{t|t}\Phi^T + Q \\
\hat{F}_{t+1|t+1} & = & \hat{F}_{t+1|t} + K_{t+1}(X_{t+1} - \Lambda \hat{F}_{t+1|t}) \\
P_{t+1|t+1} & = & P_{t+1|t} - K_{t+1}\Lambda P_{t+1|t}
\end{array}
$$
    where the optimal Kalman Gain $K_{t+1} = P_{t+1|t}\Lambda^T(\Lambda P_{t+1|t}\Lambda^T + R)^{-1}$

- Kalman Smoother: $F_T|\mathcal{X}_T \rightarrow F_{T-1}|\mathcal{X}_T \rightarrow ... \rightarrow F_t|\mathcal{X}_T$

$$\hat{F}_{t|T} = \hat{F}_{t|t} + L_t(\hat{F}_{t+1|T} - \hat{F}_{t+1|t})$$

$$P_{t|T} = P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t})L_t^T$$

  where $L_t = P_{t|t}\Lambda^T P_{t+1|t}^{-1}$

- M-step(update parameter): update Φ, Λ, Q, R such that expected log-likelihood is maximized.

$$\{F_t, X_t\}_{t=1}^T \Rightarrow \Phi, \Lambda, Q, R$$

- Repeat until convergence, which is guaranteed to local optimum

PCA

- $Var(F_t) = I$, columns of Λ are orthogonal
- When the condition above is satisfied, $X_t \stackrel{PCA}{\rightarrow} \Lambda F_t$

5 Reference

Kilian, L. and Lütkepohl, H., 2017. Structural vector autoregressive analysis. Cambridge University Press.
Webpage: `https://sites.google.com/site/lkilian2019/textbook/preliminary-chapt`

Stock, J.H. and Watson, M.W., 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In Handbook of macroeconomics (Vol. 2, pp. 415-525). Elsevier.

Jur van den Berg's notes on EM Algorithm to estimate linear Gaussian state space model: `http://arl.cs.utah.edu/resources/EM\%20Algorithm.pdf`.