Student's t Vector AutoRegression (St-VAR) Model

by Mohammad Mahdi Banasaz[1]

In this manuscript, I introduce Student's t Vector Autoregression (**St-VAR**) model[2] using the method of maximum likelihood for its estimation. I unfold Probabilistic Reduction Approach which is used to derive log likelihood function. I discuss it by giving an example, Stationary St-VAR, and deriving its log likelihood function. Also, I review Multivariate Student's t distributions. I explain St-VAR model and its reparameterization. Stationary St-VAR and Heterogenous St-VAR are two version of St-VAR which are presented here.

1. Probabilistic Reduction Approach

The probabilistic reduction approach starts with a multivariate distribution which collects all the information about a stochastic process. It further aims to reduce the dimensions of that joint distribution by using probabilistic assumptions regarding observational data and give a feasible log likelihood function for maximum likelihood estimation at the end. Assumptions are divided to three categories: Distribution (D), Dependence (I), and Heterogeneity (H). These assumptions help to reduce the joint distribution to a simple (reduced) form. Table 1 provides example of assumptions that are used in the Stationary St-VAR.

Table 1: Probabilistic Assumptions

Distribution (D)	Student's t
Dependence (I)	Markov (p)
Heterogeneity (H)	2^{nd} Order Stationarity

As a showcase of probabilistic reduction approach. Let $\boldsymbol{z} \coloneqq (r_1, r_2, \ldots, r_k)'$ is a k dimensional random vector and $\boldsymbol{Z} \coloneqq (\boldsymbol{z}_T, \boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_1)$ is a stochastic process. $D(\boldsymbol{Z};\ \boldsymbol{\Theta})$ denotes the joint distribution which can be written as a product of conditional distribution,

$$D(\boldsymbol{Z};\ \boldsymbol{\Theta}) = D(\boldsymbol{z}_T, \boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_1;\ \boldsymbol{\Theta}) = D_T(\boldsymbol{z}_T |\ \boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_1;\ \boldsymbol{\Theta}_T)\, D(\boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_1;\ \boldsymbol{\Theta})$$

$$= D_T(\boldsymbol{z}_T |\ \boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_1;\ \boldsymbol{\Theta}_T)\, D_{T-1}(\boldsymbol{z}_{T-2} | \boldsymbol{z}_{T-2}, \ldots, \boldsymbol{z}_1;\ \boldsymbol{\Theta}_{T-1}) D(\ \boldsymbol{z}_{T-2}, \ldots, \boldsymbol{z}_1;\ \boldsymbol{\Theta}) = \ldots,$$

So,

$$D(\boldsymbol{Z};\ \boldsymbol{\Theta}) = D(\boldsymbol{z}_T, \boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_1;\ \boldsymbol{\Theta}) = D_T(\boldsymbol{z}_T |\ \boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_1;\ \boldsymbol{\Theta}_T) \prod_{t=T-1}^{1} D(\boldsymbol{z}_t\ | \boldsymbol{z}_{t-1}, \boldsymbol{z}_{t-2}, \ldots, \boldsymbol{z}_1;\ \boldsymbol{\Theta}\,(t))$$

Markov (p) and 2^{nd} order stationarity implies reduced form:

$$D(\boldsymbol{Z};\ \boldsymbol{\Theta}) = D\big(\boldsymbol{Z}_{\boldsymbol{p}}^{\boldsymbol{1}};\ \boldsymbol{\Theta}_1\big) \prod_{t=p+1}^{T} D(\boldsymbol{Z}_t\ | \boldsymbol{Z}_{t-1}^{t-p};\ \boldsymbol{\Theta}_2)$$

---

[1] PhD candidate in Econometrcis and Quantitative Economics, Department of Economics at Virginia Tech, mbanasaz@vt.edu.

[2] Firstly, introduced in Niraj Poudyal's PhD Thesis; "Confronting Theory with Data: The Case of DSGE Modeling"

2. Multivariate Student's t Distribution

It is a brief review of joint, marginal, and conditional student's t distributions. Although, practitioners use the Normal distribution for regression analysis, the Student's t is a preferable distribution to modeling financial data. It has higher peak and fat tails that can capture high volatilities in data. Figures 1 and 2 shows the density of the bivariate Normal distribution and the bivariate Student's t distribution.

Figure 1: Bivariate Normal Distribution

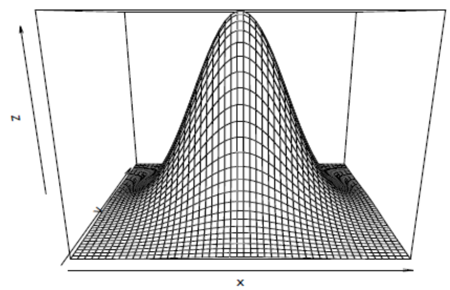
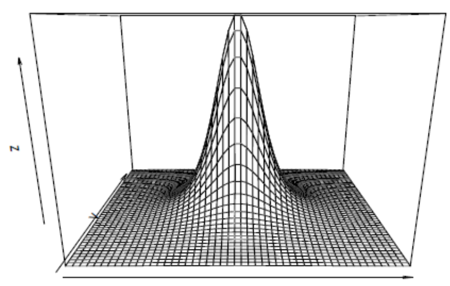


Figure 2: Bivariate Student's t Distribution



2.1. Joint Distribution

Consider $\boldsymbol{z} \coloneqq (r_1, r_2, \dots, r_k)'$ as a k dimensional random vector. The statement "$\boldsymbol{z}$ have the k-variate student's t distribution with the degree of freedom or shape parameter ν, a location vector $\boldsymbol{\mu}$ and a scaling matrix $\boldsymbol{\Sigma}$" can be summarized by $\boldsymbol{z} \sim st_k(\boldsymbol{\mu}, \boldsymbol{\Sigma};\ \nu)$. $f(\boldsymbol{z};\theta)$ is a joint probability density function for $\boldsymbol{z}$ where θ is the parameter space and $\Gamma(.)$ is the gamma function.

$$f(\boldsymbol{z};\theta) = \frac{\Gamma(\frac{\nu+k}{2})}{(\pi\nu)^{\frac{k}{2}}\,\Gamma(\frac{\nu}{2})}\,|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\left[1+\frac{1}{\nu}\,(\boldsymbol{z}-\boldsymbol{\mu})'\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\right]^{-\frac{\nu+k}{2}}, \theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$$

$$\boldsymbol{z} \coloneqq \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{bmatrix}, \boldsymbol{\mu} \coloneqq \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}, \boldsymbol{\Sigma} \coloneqq \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2k} \\ \vdots & & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix}$$

2.2. Marginal Distributions

As $\boldsymbol{z} \sim st_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, one can divide the k dimensional random vector into two k_1 and k_2 dimensional random vectors, $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, such that $k_1 + k_2 = k$. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are partitioned accordingly.

$$\begin{pmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{pmatrix} \sim St\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}; \nu\right]$$

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{z}_1\ (k_1 \times 1) \\ \boldsymbol{z}_2\ (k_2 \times 1) \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1\ (k_1 \times 1) \\ \boldsymbol{\mu}_2\ (k_2 \times 1) \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}(k_1 \times k_1) & \boldsymbol{\Sigma}_{12}(k_1 \times k_2) \\ \boldsymbol{\Sigma}_{21}(k_2 \times k_1) & \boldsymbol{\Sigma}_{22}(k_2 \times k_2) \end{pmatrix}; k_1 + k_2 = k$$

Marginal distributions can be defined by for $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ as $\boldsymbol{z}_1 \sim st_{k_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11};\ \nu)$ and $\boldsymbol{z}_2 \sim st_{k_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22};\ \nu)$. Density functions are presented below.

$$f(\boldsymbol{z}_1; \theta_1) = \frac{\Gamma(\frac{\nu + k_1}{2})}{(\pi\nu)^{\frac{k_1}{2}}\,\Gamma(\frac{\nu}{2})}\,|\boldsymbol{\Sigma}_{11}|^{-\frac{1}{2}}\left[1 + \frac{1}{\nu}\,(\boldsymbol{z}_1 - \boldsymbol{\mu}_1)'\,\boldsymbol{\Sigma}_{11}{}^{-1}(\boldsymbol{z}_1 - \boldsymbol{\mu}_1)\right]^{-\frac{\nu + k_1}{2}}$$

$$f(\boldsymbol{z}_2; \theta_2) = \frac{\Gamma(\frac{\nu + k_2}{2})}{(\pi\nu)^{\frac{k_2}{2}}\,\Gamma(\frac{\nu}{2})}\,|\boldsymbol{\Sigma}_{22}|^{-\frac{1}{2}}\left[1 + \frac{1}{\nu}\,(\boldsymbol{z}_2 - \boldsymbol{\mu}_2)'\,\boldsymbol{\Sigma}_{22}{}^{-1}(\boldsymbol{z}_2 - \boldsymbol{\mu}_2)\right]^{-\frac{\nu + k_2}{2}}$$

2.3. Conditional Distribution

(Spanos, 1994)[3] derived the conditional distribution of $\boldsymbol{z}_1$ on $\boldsymbol{z}_2$ when they are two random vectors with k_1 and k_2 variables/dimensions, respectively.

$$[\boldsymbol{z}_1 | \boldsymbol{z}_2] \sim St\,(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\,\boldsymbol{\Sigma}_{22}^{-1}\,[\boldsymbol{z}_2 - \boldsymbol{\mu}_2], [\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\,\boldsymbol{\Sigma}_{22}^{-1}\,\boldsymbol{\Sigma}_{21}]\,q(\boldsymbol{z}_2); \nu + k_2)$$

Where $\boldsymbol{z}_1 \sim st_{k_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11};\ \nu)$, $\boldsymbol{z}_2 \sim st_{k_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22};\ \nu)$, and

$$q(\boldsymbol{z}_2) = \frac{\nu}{\nu + k_2}\left[1 + \frac{1}{\nu}\,(\boldsymbol{z}_2 - \boldsymbol{\mu}_2)'\,\boldsymbol{\Sigma}_{22}{}^{-1}(\boldsymbol{z}_2 - \boldsymbol{\mu}_2)\right]$$

3. Student’s t Vector AutoRegression (St-VAR) Model

3.1. Join Probability Distribution Function

Consider $\{\boldsymbol{z}_t, t = 1,2, \dots, T\}$ be a vector stochastic process where $\boldsymbol{z} \coloneqq (r_1, r_2, \dots, r_k)'$ is a k dimensional random vector. I define $\boldsymbol{Z}$ as a long vector in which $\boldsymbol{z}_t$, $\boldsymbol{z}_{t-1}$, $\boldsymbol{z}_{t-2}$, …, $\boldsymbol{z}_{t-p}$ are stacked. The statement “$\boldsymbol{Z}$ have the k-variate student’s t distribution with the degree of freedom or shape parameter ν, a location vector $\boldsymbol{\mu}$ and a scaling matrix $\boldsymbol{\Sigma}$” is summarized by $\boldsymbol{Z} \sim St(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ or in extended matrix form

[3] Spanos, Aris. "On modeling heteroskedasticity: The Student's t and elliptical linear regression models." *Econometric Theory* 10.2 (1994): 286-315.

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{z}_t \\ \boldsymbol{z}_{t-1} \\ \boldsymbol{z}_{t-2} \\ \vdots \\ \boldsymbol{z}_{t-p} \end{pmatrix} \sim St\left[\begin{pmatrix} \boldsymbol{\mu}_{z_t} \\ \boldsymbol{\mu}_{z_{t-1}} \\ \boldsymbol{\mu}_{z_{t-2}} \\ \vdots \\ \boldsymbol{\mu}_{z_{t-p}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \boldsymbol{\Sigma}_{02} & \cdots \boldsymbol{\Sigma}_{0p} \\ \boldsymbol{\Sigma}_{10} & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots \boldsymbol{\Sigma}_{1p} \\ \boldsymbol{\Sigma}_{20} & \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \ddots \boldsymbol{\Sigma}_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\Sigma}_{p0} & \boldsymbol{\Sigma}_{p1} & \boldsymbol{\Sigma}_{p2} & \cdots \boldsymbol{\Sigma}_{pp} \end{pmatrix}; \nu\right]$$

where ν is degree of freedom parameter. Also, $Var(\boldsymbol{Z}) = \frac{\nu}{\nu-2}\,\boldsymbol{\Sigma}$, that makes the distribution of estimators depend on ν. $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ are defined as below.

$$\boldsymbol{Z} = [\boldsymbol{z}_t, \boldsymbol{z}_{t-1}, \boldsymbol{z}_{t-2}, \dots, \boldsymbol{z}_{t-p}]' \text{ and } \boldsymbol{\mu} = [\boldsymbol{\mu}_{z_t}, \boldsymbol{\mu}_{z_{t-1}}, \boldsymbol{\mu}_{z_{t-2}}, \dots, \boldsymbol{\mu}_{z_{t-p}}]'$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \boldsymbol{\Sigma}_{02} & \cdots \boldsymbol{\Sigma}_{0p} \\ \boldsymbol{\Sigma}_{10} & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots \boldsymbol{\Sigma}_{1p} \\ \boldsymbol{\Sigma}_{20} & \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \ddots \boldsymbol{\Sigma}_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\Sigma}_{p0} & \boldsymbol{\Sigma}_{p1} & \boldsymbol{\Sigma}_{p2} & \cdots \boldsymbol{\Sigma}_{pp} \end{pmatrix}$$

$$\boldsymbol{z}_t: (k \times 1),\ \boldsymbol{\mu}_{z_t}: (k \times 1),\ \boldsymbol{\mu}: ((p+1)k \times 1),$$
$$\boldsymbol{\Sigma}_{ij}: (k \times k),\ \boldsymbol{\Sigma}: ((p+1)k \times (p+1)k),$$

Probability Density Function (PDF) of $\boldsymbol{Z}$ where $\boldsymbol{\Theta}$ is the parameter space is

$$f(\boldsymbol{Z}; \boldsymbol{\Theta}) = \frac{\Gamma(\frac{\nu+m}{2})}{(\pi\nu)^{\frac{m}{2}}\,\Gamma(\frac{\nu}{2})}\, |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu}\,(\boldsymbol{Z}-\boldsymbol{\mu})'\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{\mu})\right]^{-\frac{\nu+m}{2}}, \boldsymbol{\Theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$$

Also, $Var(\boldsymbol{Z}) = \frac{\nu}{\nu-2}\,\boldsymbol{\Sigma}$, $m = (p+1)k$ denotes the number of random variables in $\boldsymbol{Z}$, k denotes the number of variables in $\boldsymbol{z_t}$ and p denotes number of lags.

3.2. Regression and Skedastic Function

To get the conditional distribution of $\boldsymbol{z}_t$ on $\boldsymbol{Z}^0_{t-p}$, I partitioned Vectors $\boldsymbol{Z}$ and $\boldsymbol{\mu}$, and matrix $\boldsymbol{\Sigma}$ as below where $\boldsymbol{\mu}_{pk}(pk \times 1)$ is a vector of lk $\boldsymbol{\mu}_z$'s.

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_t(k \times 1) \\ \boldsymbol{Z}^0_{t-p}\,(pk \times 1) \end{bmatrix}, \boldsymbol{\mu}_t = \begin{bmatrix} \boldsymbol{\mu}_{z_t}(k \times 1) \\ \boldsymbol{\mu}_{pk}(pk \times 1) \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}\,(k \times k) & \boldsymbol{\Sigma}_{12}\,(k \times pk) \\ \boldsymbol{\Sigma}_{21}\,(pk \times k) & \mathbf{Q}(pk \times pk) \end{bmatrix}$$

Then by implementing the formula of (Spanos, 1994), one write down the conditional distribution of $\boldsymbol{z}_t$ on $\boldsymbol{Z}^0_{t-1}$ as

$$[\boldsymbol{z}_t | \boldsymbol{Z}^0_{t-p}] \sim St\left(\boldsymbol{\mu}_{z_t} + \boldsymbol{\Sigma}_{12}\,\boldsymbol{\Sigma}^{-1}_{22}\,[\boldsymbol{Z}^0_{t-\mathrm{p}} - \boldsymbol{\mu}_{pk}], [\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\mathbf{Q}^{-1}\boldsymbol{\Sigma}_{21}]\, q(\boldsymbol{Z}^0_{t-\mathrm{p}}); \nu + pk\right)$$

Where $\boldsymbol{z}_t \sim st_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11};\, \nu)$, $\boldsymbol{Z}^0_{t-\mathrm{p}} \sim st_{pk}(\boldsymbol{\mu}_{pk}, \mathbf{Q};\, \nu)$, and

$$q\left(\boldsymbol{Z}_{t-p}^{0}\right) = \frac{\nu + pk}{\nu + pk - 2} \left[1 + \frac{1}{\nu} (\boldsymbol{Z}_{t-1}^{0} - \boldsymbol{\mu}_2)' \, \mathbf{Q}^{-1} (\boldsymbol{Z}_{t-1}^{0} - \boldsymbol{\mu}_2)\right]$$

As a result, the regression function, conditional mean; $E\left[\boldsymbol{z}_t \middle| \boldsymbol{Z}_{t-p}^{0}\right]$, is as below.

$$E\left[\boldsymbol{z}_t \middle| \boldsymbol{Z}_{t-p}^{0}\right] = \boldsymbol{\mu}_{\boldsymbol{z}_t} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \left[\boldsymbol{Z}_{t-\mathrm{p}}^{0} - \boldsymbol{\mu}_{pk}\right]$$

Also, the skedastic function, condition variance; $Var\left[\boldsymbol{z}_t \middle| \boldsymbol{Z}_{t-p}^{0}\right]$, is as below.

$$Var\left[\boldsymbol{z}_t \middle| \boldsymbol{Z}_{t-p}^{0}\right] = \left[\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \mathbf{Q}^{-1} \boldsymbol{\Sigma}_{21}\right] \frac{\nu + pk}{\nu + pk - 2} \left[1 + \frac{1}{\nu} (\boldsymbol{Z}_{t-1}^{0} - \boldsymbol{\mu}_2)' \, \mathbf{Q}^{-1} (\boldsymbol{Z}_{t-1}^{0} - \boldsymbol{\mu}_2)\right]$$

3.3. Reparameterization, log likelihood Function and Estimation

Likelihood function (often simply called the likelihood) measures the likelihood of a sample of data for given values of the unknown parameters. It is viewed as a function of parameters and formed from joint probability distribution. To derive the likelihood function of St-VAR model I start from joint distribution that can be written as a multiplication of a conditional and marginal distribution.

$$D\left(\boldsymbol{z}_t, \boldsymbol{Z}_{t-p}^{0}\,;\, \boldsymbol{\Theta}\right) = D\left(\boldsymbol{z}_t \middle| \, \boldsymbol{Z}_{t-p}^{0}\,;\, \boldsymbol{\Theta}_1\right) D\left(\boldsymbol{Z}_{t-p}^{0}\,;\, \boldsymbol{\Theta}_2\right) \sim St(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu), \qquad \boldsymbol{\Theta} = \boldsymbol{\Theta}_1 \cup \boldsymbol{\Theta}_2$$

Such that conditional distribution is a Student's t distribution itself.

$$D\left(\boldsymbol{z}_t \middle| \, \boldsymbol{Z}_{t-p}^{0}\,;\, \boldsymbol{\Theta}_1\right) \sim St\left(\boldsymbol{a}_0 + \boldsymbol{A}' \boldsymbol{Z}_{t-p}^{0}, \boldsymbol{\Omega}\, q\left(\boldsymbol{Z}_{t-p}^{0}\right);\, \nu + pk\right);$$

$$q\left(\boldsymbol{Z}_{t-p}^{0}\right) = \left[1 + \frac{1}{\nu} \left(\boldsymbol{Z}_{t-p}^{0} - \boldsymbol{\mu}_{pk}\right)' \boldsymbol{Q}^{-1} \left(\boldsymbol{Z}_{t-\boldsymbol{p}}^{0} - \boldsymbol{\mu}_{pk}\right)\right], \qquad \boldsymbol{\Theta}_1 = \{\boldsymbol{a}_0, \boldsymbol{A}, \boldsymbol{\Omega}, \boldsymbol{\mu}\}$$

And marginal distribution is presented in below.

$$D\left(\boldsymbol{Z}_{t-p}^{0}\,;\, \boldsymbol{\Theta}_2\right) \sim St(\boldsymbol{\mu}_{pk}, \boldsymbol{Q}\,;\, \nu),\ \boldsymbol{\Theta}_2 = \{\boldsymbol{\mu}, \boldsymbol{Q}\}$$

Parameters of the St-VAR model, as presented above, are $\boldsymbol{a}_0$, $\boldsymbol{A}$, $\boldsymbol{\Omega}$, $\boldsymbol{\mu}$, and $\mathbf{Q}$.

$\boldsymbol{A}' = \boldsymbol{\Sigma}_{12}\ \mathbf{Q}^{-1}$

$\boldsymbol{a}_0 = \boldsymbol{\mu}_z - \boldsymbol{A}' \boldsymbol{\mu}_{pk}$

$\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\, \mathbf{Q}^{-1} \boldsymbol{\Sigma}_{12}'$

$\mathbf{Q}^{-1} = \{q_{ij}\,; i, j = 1, 2, \ldots, pk\}$

$\boldsymbol{\Omega} = \{\omega_{ij}\,; i, j = 1, 2, \ldots, k\}$

Log-likelihood function is formulated as below.

$$ln f(\boldsymbol{\Theta}; \boldsymbol{Z}) = c - \frac{1}{2} \ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\nu + m) \ln\left[1 + \frac{1}{\nu} (\boldsymbol{Z} - \boldsymbol{\mu})' \, \boldsymbol{\Sigma}^{-1} (\boldsymbol{Z} - \boldsymbol{\mu})\right]$$

$$c = \ln \Gamma\left(\frac{\nu + m}{2}\right) - \ln \Gamma\left(\frac{\nu}{2}\right) - \frac{m}{2} \ln \pi\nu$$

If $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \dots, \boldsymbol{Z}_T$ are identically distributed, then log-likelihood function will be as below.

$$l_T(\boldsymbol{\Theta}) = \ lnL(\boldsymbol{\Theta}; \boldsymbol{Z}_1, \boldsymbol{Z}_2, \dots, \boldsymbol{Z}_T)$$

$$\propto Tc - \frac{T}{2}\ \ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\nu + m)\sum_{t=1}^{T} \ln\left[\ 1 + \frac{1}{\nu}(\boldsymbol{Z} - \boldsymbol{\mu})'\ \boldsymbol{\Sigma}^{-\mathbf{1}}\ (\boldsymbol{Z} - \boldsymbol{\mu})\right]$$

To estimate the St-VAR model, I use log-likelihood maximization method. Log-likelihood function is maximized by bringing the numerical optimization methods into play and standard deviation of estimation. (more …)

4. Stationary St-VAR Model

Probabilistic assumptions for Stationary St-VAR statistical model; Distribution (D), Dependence (I), and Heterogeneity (H), are presented in Table 2.

Table 2: Stationary St-VAR Assumptions

Distribution (D)	Student's t
Dependence (I)	Markov (p)
Heterogeneity (H)	2^{nd} Order Stationarity

Consider $\{\boldsymbol{z}_t, t = 1,2, \dots, T\}$ be a vector stochastic process and $\boldsymbol{Z} \sim St(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$. Joint distribution parameters are changed below based on the assumptions in Table 2.

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{z}_t \\ \boldsymbol{z}_{t-1} \\ \boldsymbol{z}_{t-2} \\ \vdots \\ \boldsymbol{z}_{t-p} \end{pmatrix} \sim St\left[\begin{pmatrix} \boldsymbol{\mu}_Z \\ \boldsymbol{\mu}_Z \\ \boldsymbol{\mu}_Z \\ \vdots \\ \boldsymbol{\mu}_Z \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \boldsymbol{\Sigma}_{02} & \cdots & \boldsymbol{\Sigma}_{0p} \\ \boldsymbol{\Sigma}_{01}^T & \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \cdots & \boldsymbol{\Sigma}_{1p-1} \\ \boldsymbol{\Sigma}_{02}^T & \boldsymbol{\Sigma}_{01}^T & \boldsymbol{\Sigma}_{00} & \ddots & \boldsymbol{\Sigma}_{1p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{\Sigma}_{0p}^T & \boldsymbol{\Sigma}_{1p-1}^T & \boldsymbol{\Sigma}_{1p-2}^T & \cdots & \boldsymbol{\Sigma}_{00} \end{pmatrix}; \nu\right]$$

Table 3 summarizes specification of the Stationary $St - VAR(p; \nu)$ statistical model. As a result, conditional mean is a linear function, but conditional variance is a nonlinear quadratic function and heteroskedastic. Note that Table 3 results from the assumptions in Table 2 in a genuine way. It is derived by taking advantage of probabilistic reduction approach.

Table 3: Stationary $(St - VAR(p;v))$ Model

Statistical GM: $\boldsymbol{z}_t = \boldsymbol{a}_0 + \sum_{i=1}^{p} \boldsymbol{A}_i' \boldsymbol{z}_{t-i} + \boldsymbol{u}_t,\ t \in \mathbb{N},$		
1	Student's t	D $(\boldsymbol{z}_t, \boldsymbol{Z}_{t-p}^0;\ \boldsymbol{\Theta})$ is Student's t with v degree of freedom, for $\boldsymbol{Z}_{t-p}^0 := (\boldsymbol{z}_{t-1}, \dots, \boldsymbol{z}_{t-p}),$
2	Linearity	$E(\boldsymbol{z}_t \vert \sigma(\boldsymbol{Z}_{t-1}^{t-0})) = \boldsymbol{a}_0 + \sum_{i=1}^{p} \boldsymbol{A}_i' \boldsymbol{z}_{t-i}$ is linear in $\boldsymbol{Z}_{t-1}^0 := (\boldsymbol{z}_{t-1}, \dots, \boldsymbol{z}_1),$
3	Heteroskedasticity	$Var(\boldsymbol{z}_t \vert \sigma(\boldsymbol{Z}_{t-1}^0) = \frac{v+pk}{v+pk-2}\, \boldsymbol{\Omega}\, q(\boldsymbol{Z}_{t-1}^{t-0}),$ $q(\boldsymbol{Z}_{t-1}^0) = [1 + \frac{1}{v} \left(\boldsymbol{Z}'^{t-p}_{t-1} - \boldsymbol{\mu}_{pk}\right)' \mathbf{Q}^{-1} (\boldsymbol{Z}'^{t-p}_{t-1} - \boldsymbol{\mu}_{pk})],$
4	Markov (p)	$\boldsymbol{Z} = \{\boldsymbol{z}_{t,}\ t \in \mathbb{N}\}$ is a Markov(p) process,
5	t-invariance	$\boldsymbol{\Theta} = \{\boldsymbol{a}_0, \boldsymbol{A}_1, \dots, \boldsymbol{A}_p, \boldsymbol{\Omega}, \mathbf{Q}\}$ are t-invariant for all $t \in \mathbb{N}$.

5. Heterogenous St-VAR Model

A stationary St-VAR model can be extended to a heterogenous one by considering a heterogenous mean for stochastic process, $\boldsymbol{\mu}_z(t)$. As a result, mean heterogeneity spills into conditional variance heterogeneity naturally for the multivariate distribution through the heteroskedasticity. Table 4 presents probabilistic assumptions for Heterogeneous St-VAR statistical model; Distribution (D), Dependence (I), and Heterogeneity (H).

Table 4: Heterogenous St-VAR Assumptions

Distribution (D)	Student's t
Dependence (I)	Markov (p)
Heterogeneity (H)	Mean Heterogeneity

Replacing $\boldsymbol{\mu}_z$ with $\boldsymbol{\mu}_z(t)$, joint distribution can be written as below.

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{z}_t \\ \boldsymbol{z}_{t-1} \\ \boldsymbol{z}_{t-2} \\ \vdots \\ \boldsymbol{z}_{t-p} \end{pmatrix} \sim St \left[\begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{Z}}(t) \\ \boldsymbol{\mu}_{\boldsymbol{Z}}(t-1) \\ \boldsymbol{\mu}_{\boldsymbol{Z}}(t-2) \\ \vdots \\ \boldsymbol{\mu}_{\boldsymbol{Z}}(t-p) \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \boldsymbol{\Sigma}_{02} & \cdots & \boldsymbol{\Sigma}_{0p} \\ \boldsymbol{\Sigma}_{01}^T & \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \cdots & \boldsymbol{\Sigma}_{1p-1} \\ \boldsymbol{\Sigma}_{02}^T & \boldsymbol{\Sigma}_{01}^T & \boldsymbol{\Sigma}_{00} & \ddots & \boldsymbol{\Sigma}_{1p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{\Sigma}_{0p}^T & \boldsymbol{\Sigma}_{1p-1}^T & \boldsymbol{\Sigma}_{1p-2}^T & \cdots & \boldsymbol{\Sigma}_{00} \end{pmatrix}; v \right]$$

For example, if $\boldsymbol{\mu}_z(t) = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\mu}_2 t^2$. Reparameterization for $\boldsymbol{a}_0$ leads to the below.

$$\boldsymbol{a}_0 = \boldsymbol{\mu}_z(t) - \boldsymbol{A}_1^T \boldsymbol{\mu}_z(t-1) - \boldsymbol{A}_2^T \boldsymbol{\mu}_z(t-2) - \boldsymbol{A}_3^T \boldsymbol{\mu}_z(t-3) = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_1 t + \boldsymbol{\delta}_2 t^2.$$

Table 5 summarizes specification of the Heterogenous St-VAR Model ($St - VAR(p; \nu)$).

Table 5: Heterogenous St-VAR Model ($St - VAR(p; \nu)$)

Statistical GM: $\boldsymbol{z}_t = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_1 t + \boldsymbol{\delta}_2 t^2 + \sum_{i=1}^{p} \boldsymbol{A}_i' \boldsymbol{z}_{t-i} + \boldsymbol{u}_t,\ t \in \mathbb{N}$,		
1	Student's t	D ($\boldsymbol{z}_t, \boldsymbol{Z}_{t-1}^0;\ \boldsymbol{\Theta}$) is Student's t with v degree of freedom, for $\boldsymbol{Z}_{t-1}^0 :=$ $(\boldsymbol{z}_{t-1}, \dots, \boldsymbol{z}_1)$,
2	Linearity	$E(\boldsymbol{z}_t \vert \sigma(\boldsymbol{Z}_{t-1}^0)) = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_1 t + \boldsymbol{\delta}_2 t^2 + \sum_{i=1}^{p} \boldsymbol{A}_i' \boldsymbol{z}_{t-i}$ is linear in $\boldsymbol{Z}_{t-1}^0 :=$ $(\boldsymbol{z}_{t-1}, \dots, \boldsymbol{z}_1)$,
3	Heteroskedasticity	$Var(\boldsymbol{z}_t \vert \sigma(\boldsymbol{Z}_{t-1}^0) = \frac{v}{v+pk-2}\, \boldsymbol{\Omega}\, q(\boldsymbol{Z}_{t-1}^0)$, $q(\boldsymbol{Z}_{t-1}^0) = [1 + \frac{1}{v}\left(\boldsymbol{Z'}_{t-1}^{t-p} - \boldsymbol{\mu}_{pk}(t)\right)'\ \mathbf{Q}^{-1}(\boldsymbol{Z'}_{t-1}^{t-p} - \boldsymbol{\mu}_{pk}(t))]$,
4	Markov	$\boldsymbol{Z} = \{\boldsymbol{z}_{t,}\ t \in \mathbb{N}\}$ is a Markov (p) process,
5	t-invariance	$\boldsymbol{\Theta} = \left\{\boldsymbol{\delta}_0,\ \boldsymbol{\delta}_1,\ \boldsymbol{\delta}_2,\ \boldsymbol{A_1}, \dots,\ \boldsymbol{A_p}, \boldsymbol{\Omega}, \mathbf{Q}\right\}$ are t-invariant for all $t \in \mathbb{N}$.