Topics in Time Series Econometrics

Cesar E. Tamayo
Department of Economics, Rutgers University
ctamayo@econ.rutgers.edu

September 29, 2012

Contents

Chapter 1

Martingales, Markov Chains and the AR(1)

1.1 Martingales, Supermartingales and MDS

A discrete-time stochastic process $\{x_t\}$ is said to be a *martingale* if, for any t:

$$E(|x_t|) < \infty \quad \text{and} \quad E(x_{t+1}|x_t, x_{t-1}, ...) = x_t$$

that is, if the expectation of x_t conditional on all past values is equal to the last observation.

Remark 1 *A random walk (RW) is a martingale. To see why consider the RW:*

$$x_{t+1} = x_t + \varepsilon_t \quad \text{where } \varepsilon_t \text{ iid}\left(0, \sigma^2\right)$$

then notice:

$$\begin{aligned} E(x_{t+1}|x_t, x_{t-1}, ...) &= E_t(x_t + \varepsilon_t) \\ &= x_t \end{aligned}$$

A stochastic process $\{x_t\}$ is a supermartingale if

$$E(x_{t+1}|x_t, x_{t-1}, ...) \leq x_t$$

Proposition 2 *A concave (convex) function of a martingale is a supermartingale (submartingale)*

Proof. Let $\{x_t\}$ be a martingale and $f(\cdot)$ a concave function. Then notice that $E_t(x_{t+1}) = x_t$ and by Jensen's inequality f concave implies:

$$\begin{aligned} E_t(f(x_{t+1})) &\leq f(E_t(x_{t+1})) \\ &= f(x_t) \end{aligned}$$

■

A stochastich process $\{x_t\}$ is called a *martingale difference sequence* (MDS) if its expectation conditional on past values of another stochastic series is zero, that is:

$$E(x_{t+1}|y_t, y_{t-1}, ...) = 0$$

Remark 3 *If $\{x_t\}$ is a martingale, then the stochastic process $\{y_t\}$ defined as $y_t \equiv x_{t+1} - x_t$ for every t is a MDS. To see this note:*

$$\begin{aligned} E_t(y_t) &= E_t(x_{t+1} - x_t) \\ &= E_t(x_{t+1}) - x_t \\ &= 0 \end{aligned}$$

hence its name MDS.

A generalization of the Lingeberg-Levy CLT for can be obtained by replacing "iid" in e.g. for MDS:

Theorem 4 (CLT for MDS) *If $\{x_t\}$ is a martingale difference process with finite second moment $E(x_t x_t') = \Sigma$ then:*

$$\sqrt{T}\bar{x} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \xrightarrow{d} N(0, \Sigma)$$

1.2 Markov chains

A stochastic process $\{x_t\}$ has the Markov property if $\Pr(x_{t+1}|x_t, x_{t-1}, ...) = \Pr(x_{t+1}|x_t)$. A specific process satisfying the Markov property is a **Markov chain** which in turn is a triplet of objects $\left\{[e_i]_{i=1}^{S}, \mathbb{P}, \pi_0\right\}$ where:

- $e_i = (0, 0, ...0, 1, 0, ...0)'$ is a $n \times 1$ vector with 1 as its *ith* element and zeros elsewhere. That is, e_i says that "the state of nature is i".
- $\mathbb{P}$ is a transition probability matrix with typical element $P_{ij} = \Pr(x_{t+1} = e_j | x_t = e_i)$ so that $\sum_{j=1}^{S} P_{ij} = 1$. Notice that this implies $x_{t+1} \in \mathbf{e} \subseteq \mathbb{R}^S$ where $\mathbf{e} = (e_1, e_2, ..., e_S)$.
- π_0 is the probability of being at state i in period 0.

We easily make k-step forecasts of a Markov chain. For instance, if $k = 1$:

$$\Pr(x_{t+2} = e_j | x_t = e_i) = \sum_{h=1}^{S} P_{ih} P_{hj} = P_{ij}^{(2)}$$

ahere $P_{ij}^{(2)}$ is the ij-th element of $\mathbb{P}^2$. Similarly, the k-step ahead forecast can be found as $P_{ij}^{(k)}$. Notice that if, for some ij we have $P_{ij} = 1$ then state i is an absorbing sate.

Definition 5 *A Markov chain is said to be **irreducible** if it is possible to get to any state from any state*

Definition 6 *A state i is **aperiodic** if $\exists$ n such that $\forall$ $n' \geq n$: $\Pr(x_{n'} = e_i \mid x_0 = e_i) > 0$, that is, if returns to state i can occur at irregular times.*

Definition 7 *A Markov chain is said to be aperiodic if every state is aperiodic.*

Definition 8 *State i is **positive recurrent** if the mean recurrence time at state i is finite, i.e. if*

$$M_i = \mathbb{E}[T_i = \inf\{n \geq 1 \ : \ x_n = e_i \mid x_0 = e_i\}] < \infty$$

Definition 9 *A state i is said to be **ergodic** if it is aperiodic and positive recurrent.*

Definition 10 *An irreducible Markov chain is ergodic if every state is ergodic.*

1.2.1 Unconditional probabilities and stationary distribution

The unconditional probability of a Markov process are determined by:

$$\pi_t = \Pr(x_t) = \pi_0' \mathbb{P}^t \Rightarrow \pi_{t+1} = \pi_t' \mathbb{P}$$

since $\pi_t' \mathbb{P} = (\pi_0' \mathbb{P}^t)\, \mathbb{P} = \pi_0' \mathbb{P}^{t+1}$. An unconditional distribution is said to be time-invariant or stationary if

$$\begin{aligned} \pi &= \pi' \mathbb{P} \\ \pi' (I - \mathbb{P}) &= 0 \\ (I - \mathbb{P}') \pi &= 0 \end{aligned}$$

that is, the stationary distribution π can be found as the eigenvector (normalized to satisfy $\sum_{j=1}^{S} P_{ij} = 1$) associated with the unit eigenvalue of $\mathbb{P}'$. Notice that $\mathbb{P}$ stochastic $\Rightarrow \exists$ at least one unit eigenvalue. Furthermore, the stationary distribution may not be unique because $\mathbb{P}$ may have a repeated unit eigenvalue.

1.2.2 Asymptotic stationarity of Markov chains

When do unconditional distributions π_t approach a stationary distribution? That is, does the following condition hold:

$$\lim_{t \to \infty} \pi_t = \pi_\infty$$

where $(I - \mathbb{P}')\, \pi_\infty = 0$? And if it does hold, does this depend upon the initial distribution π_0?. If the condition holds regardless of the initial distribution then the process is *asymptotically stationary with a unique invariant distribution.* Markov chains whose matrix $\mathbb{P}$ has all nonzero elements satisfy this condition (Theorem 1 LS, pp33).

1.3 The AR(1): Representation and properties

A specific example of a Markov process is the AR(1). Let $\varepsilon_t \sim N(0,1)$ *iid* shock. Then z_t follows an AR(1) process if we can write it as:

$$z_t = (1 - \varphi)\theta + \varphi z_{t-1} + \sigma \varepsilon_t$$

this is the *recursive* formulation of the AR(1) process because it *recurs* in the same form at eact t. To go from the recursive formulation, to the infinite order MA formulation, first replace z_{t-1} in the expression for z_t :

$$z_t = (1 - \varphi)\theta + \varphi \left[(1 - \varphi)\theta + \varphi z_{t-2} + \sigma \varepsilon_{t-1}\right] + \sigma \varepsilon_t$$

next, repeat this recursive replacing and after $k + 1$ times one obtains:

$$z_t = (1 - \varphi)\theta \sum_{j=0}^{k-1} \varphi^j + \varphi^k z_{t-k} + \sigma \sum_{j=0}^{k-1} \varphi^j \varepsilon_{t-j}$$

which can be rearranged as follows:

$$\begin{aligned} z_t &= \theta \sum_{j=0}^{k-1} \varphi^j - \varphi\theta \sum_{j=0}^{k-1} \varphi^j + \varphi^k z_{t-k} + \sigma \sum_{j=0}^{k-1} \varphi^j \varepsilon_{t-j} \\ &= \theta \left[\varphi^0 + \sum_{j=0}^{k-1} \varphi^{j+1} \right] - \varphi\theta \sum_{j=0}^{k-1} \varphi^j + \varphi^k z_{t-k} + \sigma \sum_{j=0}^{k-1} \varphi^j \varepsilon_{t-j} \\ &= \theta + \theta \sum_{j=0}^{k-1} \varphi^{j+1} - \theta \sum_{j=0}^{k-1} \varphi^{j+1} + \varphi^k z_{t-k} + \sigma \sum_{j=0}^{k-1} \varphi^j \varepsilon_{t-j} \\ &= \theta + \varphi^k z_{t-k} + \sigma \sum_{j=0}^{k-1} \varphi^j \varepsilon_{t-j} \end{aligned}$$

and if we let $k \to \infty$ one gets:

$$z_t = \theta + \sigma \sum_{j=0}^{\infty} \varphi^j \varepsilon_{t-j}$$

that is, the infinite-order MA representation of the AR(1) process, saying that the AR(1) process can be written as an infinite sum of past shocks. If $|\varphi| = 1$ we have a unit root or say that z_t has infinite memory.

1.4 Conditional Distribution

The distribution of z_t conditional on knowing z_{t-1}. Recall that a linear function of a normal RV is itself a normal RV. Since at t the quantity z_{t-1} is known, it can be treated as a constant and therefore z_t, conditional on z_{t-1} is just a normal RV with its mean shifted by $(1-\varphi)\theta+\varphi z_{t-1}$.To obtain the conditional mean and variance of z_t first note that the variance remains unchanged as σ^2 while the mean:

$$\begin{aligned} \mathbb{E}_{t-1}[z_t] &= \mathbb{E}_{t-1}[(1-\varphi)\theta + \varphi z_{t-1} + \sigma\varepsilon_t] \\ &= \mathbb{E}_{t-1}[(1-\varphi)\theta + \varphi z_{t-1}] + \mathbb{E}_{t-1}[\sigma\varepsilon_t] \\ &= (1-\varphi)\theta + \varphi z_{t-1} \end{aligned}$$

so the conditional (on $t-1$) distribution of z_t :

$$z_t \sim_{t-1} N((1-\varphi)\theta + \varphi z_{t-1}, \sigma^2)$$

1.5 Unconditional Distribution

The distribution of z_t presuming no knowledge of z_{t-1}, z_{t-2}...This is equivalent to the distribution of z_t conditional on knowing z_{t-k} for a very large k, that is, the distribution of z_{t+k} for a very large k with information on t. This is why the inconditional distribution is also called the long-run distribution. To obtain this, we use the infinite order MA representation:

$$\begin{aligned} \mathbb{E}z_t &= \mathbb{E}\left[\theta + \sigma \sum_{j=0}^{\infty} \varphi^j \varepsilon_{t-j}\right] \\ &= \theta + \mathbb{E}\left[\sigma \sum_{j=0}^{\infty} \varphi^j \varepsilon_{t-j}\right] \\ &= \theta \end{aligned}$$

since $|\varphi| = 1$ and each $\varepsilon_{t-j} \sim N(0,1)$, while the unconditional variance is:

$$\begin{aligned} Var\left[z_t\right] &= Var\left[\theta + \sigma \sum_{j=0}^{\infty} \varphi^j \varepsilon_{t-j}\right] \\ &= Var\left[\sigma \sum_{j=0}^{\infty} \varphi^j \varepsilon_{t-j}\right] \\ &= \left(\sigma^2 \sum_{j=0}^{\infty} \varphi^{2j}\right) Var\left[\varepsilon_{t-j}\right] \\ &= \frac{\sigma^2}{1-\varphi^2} \end{aligned}$$

note that in the last step the following expansion is used:

$$\sum_{j=0}^{\infty} \varphi^{2j} = \begin{cases} \frac{1}{\varphi^2 - 1}\left(\varphi^{\infty} - 1\right) & \text{if} \quad \neg\varphi \in \{-1, 1\} \\ \infty & \text{if} \quad \varphi \in \{-1, 1\} \end{cases}$$

so that the unconditional distribution of z_t is:

$$z_t \sim N\left(\theta, \frac{\sigma^2}{1-\varphi^2}\right)$$

Ntaurally, as long as $0 < |\varphi| < 1$ the unconditional variance is greater than the conditional variance.

Chapter 2

Vector autregressions

2.1 The VAR(p) as a VAR(1)

The first task is to show that any general *VAR(p)* process can be represented as a *VAR(1)*. Consider the unrestricted *VAR(p)*:

$$\mathbf{x}_t = \boldsymbol{\delta} + \Phi_1 \mathbf{x}_{t-1} + \Phi_2 \mathbf{x}_{t-2} + ... + \Phi_p \mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t$$

where $\mathbf{x}$ is a $(n \times 1)$ vector of endogenous variables in the model and Φ_j is a $(n \times n)$ matrix of coefficients. That is, the first equation of this system is of the form:

$$\begin{aligned} x_{1t} &= \delta_{11} + \phi_{11}^{(1)} x_{1,t-1} + ... + \phi_{11}^{(p)} x_{1,t-p} + ... + \phi_{12}^{(1)} x_{2,t-1} + ... \\ &\quad + \phi_{12}^{(p)} x_{2,t-p} + ... + \phi_{1n}^{(1)} x_{n,t-1} + ... + \phi_{1n}^{(p)} x_{n,t-p} + \varepsilon_{1t} \end{aligned}$$

Next, define:

$$\mathbf{X}_t \equiv \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{bmatrix}, \quad \mathcal{F} \equiv \begin{bmatrix} \Phi_1 & \Phi_2 & ... & \Phi_{p-1} & \Phi_p \\ \mathbf{I}_n & \mathbf{0} & ... & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & ... & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & ... & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & ... & \mathbf{I}_n & \mathbf{0} \end{bmatrix}, \quad \mathbf{v}_t \equiv \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \boldsymbol{\mu} \equiv \begin{bmatrix} \boldsymbol{\delta} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

and notice that this system can be written as:

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\delta} + \Phi_1 \mathbf{x}_{t-1} + \Phi_2 \mathbf{x}_{t-2} + ... + \Phi_p \mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t \\ \mathbf{x}_{t-1} &= \mathbf{x}_{t-1} \\ &\vdots \\ \mathbf{x}_{t-p+1} &= \mathbf{x}_{t-p+1} \end{aligned}$$

or simply:

$$\mathbf{X}_t = \boldsymbol{\mu} + \mathcal{F} \mathbf{X}_{t-1} + \mathbf{v}_t$$

which is the so-called *companion form* of the *VAR(p)*. That is, the *VAR(p)* expressed as a *VAR(1)*. The usual stability conditions required for estimating the *VAR(p)* now apply to the "companion" matrix $\mathcal{F}$. That is, stationarity requires that all the eigenvalues of the $\mathcal{F}$ matrix lie inside the unit circle. Because any *VAR(p)* can be expressed as a *VAR(1)* in what follows only a VAR(1) representation is considered.

2.2 The VAR(1): from a structural model to VMA representations

This is done in two steps. First, from the structural model obtain the SVAR and reduced form VAR(1) representations. Then, from the SVAR and the reduced form VAR(1) obtain the respective VMA representations. Consider the "structural" model:

$$\begin{aligned} y_{1t} &= \lambda_1 + \phi_{12}^{(0)} y_{2t} + \phi_{11}^{(1)} y_{1,t-1} + \phi_{12}^{(1)} y_{2,t-1} + \zeta_{1t} \\ y_{2t} &= \lambda_2 + \phi_{21}^{(0)} y_{1t} + \phi_{21}^{(1)} y_{1,t-1} + \phi_{22}^{(1)} y_{2,t-1} + \zeta_{2t} \end{aligned} \tag{2.1}$$

2.2.1 From the structural model to the VAR(1)

The structural model in (2.1) can be solved for and rearranged to yield y_{1t} and y_{2t} only as functions of lagged endogenous variables. The easiest way to do this is by using matrix algebra. First, re-write (2.1) as:

$$\begin{bmatrix} 1 & -\phi_{12}^{(0)} \\ -\phi_{21}^{(0)} & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} + \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix}$$

or, defining $\mathbf{X}_t = \begin{bmatrix} y_{1t} & y_{2t} \end{bmatrix}'$ simply as:

$$\Phi_0 \mathbf{X}_t = \boldsymbol{\lambda} + \Phi_1 \mathbf{X}_{t-1} + \boldsymbol{\zeta}_t \tag{2.2}$$

Notice that from (2.2) the so-called *SVAR* representation can be obtained:

$$\Phi(L)\,\mathbf{X}_t = \boldsymbol{\lambda} + \boldsymbol{\zeta}_t \tag{2.3}$$

where $\Phi(L) = \Phi_0 - \Phi_1 L$. Now, if Φ_0 is non-singular, Φ_0^{-1} exists and the *reduced form* VAR(1) can be obtained from (2.2) as:

$$\mathbf{X}_t = \boldsymbol{\delta} + \Theta_1 \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t \tag{2.4}$$

with:

$$\begin{aligned} \boldsymbol{\delta} &= \boldsymbol{\Phi}_0^{-1}\boldsymbol{\lambda} = \begin{bmatrix} \frac{1}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} & \frac{\phi_{12}^{(0)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} \\ \frac{\phi_{21}^{(0)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} & \frac{1}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}, \\ \Theta_1 &= \Phi_0^{-1}\Phi_1 = \begin{bmatrix} \frac{\phi_{11}^{(1)}+\phi_{12}^{(0)}\phi_{21}^{(1)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} & \frac{\phi_{12}^{(1)}+\phi_{12}^{(0)}\phi_{22}^{(1)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} \\ \frac{\phi_{21}^{(1)}+\phi_{21}^{(0)}\phi_{11}^{(1)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} & \frac{\phi_{22}^{(1)}+\phi_{21}^{(0)}\phi_{21}^{(1)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \\ \text{and} \quad : \quad \boldsymbol{\varepsilon}_t &= \boldsymbol{\Phi}_0^{-1}\boldsymbol{\zeta}_t = \begin{bmatrix} \frac{1}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} & \frac{\phi_{12}^{(0)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} \\ \frac{\phi_{21}^{(0)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} & \frac{1}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} \end{bmatrix} \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \end{aligned}$$

2.2.2 From the VAR(1) to the VMA

Now, take the reduced form VAR(1):

$$\mathbf{X}_t = \boldsymbol{\delta} + \Theta_1 \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$$

and notice that:

$$\begin{aligned} \mathbf{X}_t - \Theta_1 \mathbf{X}_{t-1} &= \boldsymbol{\delta} + \boldsymbol{\varepsilon}_t \\ \Theta(L)\,\mathbf{X}_t &= \boldsymbol{\delta} + \boldsymbol{\varepsilon}_t \end{aligned}$$

where in this case $\Theta(L) = I - \Theta_1 L$. If all the eigenvalues of Θ lie inside the unit circle, the following can be defined:

$$\begin{aligned} \boldsymbol{\mu} &= [\Theta(L)]^{-1}\boldsymbol{\delta} \\ C(L) &= [\Theta(L)]^{-1} \end{aligned}$$

and we arrive at the VMA representation for the reduced form VAR:

$$\mathbf{X}_t = \boldsymbol{\mu} + C(L)\boldsymbol{\varepsilon}_t \tag{2.5}$$

Naturally, a VMA representation also exists for the SVAR model (2.3) as:

$$\mathbf{X}_t = \boldsymbol{\mu} + \Gamma(L)\boldsymbol{\zeta}_t$$

where $\Gamma(L) = [\Phi(L)]^{-1}$. At this point the following can be established:

$$\boldsymbol{\varepsilon}_t = \Phi_0^{-1}\boldsymbol{\zeta}_t = \Gamma_0\boldsymbol{\zeta}_t$$

with:

$$\Gamma_0 = \Phi_0^{-1} = \begin{bmatrix} \frac{1}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} & \frac{\phi_{12}^{(0)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} \\ \frac{\phi_{21}^{(0)}}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} & \frac{1}{1-\phi_{12}^{(0)}\phi_{21}^{(0)}} \end{bmatrix} = \begin{bmatrix} \gamma_{11}^{(0)} & \gamma_{12}^{(0)} \\ \gamma_{21}^{(0)} & \gamma_{22}^{(0)} \end{bmatrix}$$

and:

$$C(L) = [\Theta(L)]^{-1} = \left[\Phi_0^{-1}\Phi_1(L)\right]^{-1} = \Phi_1^{-1}(L)\Phi_0 = \boldsymbol{\Gamma}(L)\boldsymbol{\Gamma}_0^{-1} \tag{2.6}$$

Notice also that the mean of both VMA representations is the same, $\boldsymbol{\mu}$, since:

$$\begin{aligned} [\Theta(L)]^{-1}\boldsymbol{\delta} &= \left(I - \Phi_0^{-1}\Phi_1 L\right)^{-1}\boldsymbol{\delta} \\ &= \left(I - \Phi_0^{-1}\Phi_1 L\right)^{-1}\Phi_0^{-1}\boldsymbol{\lambda} \\ &= \left(\Phi_0 - \Phi_1 L\right)^{-1}\boldsymbol{\lambda} \\ &= [\Phi(L)]^{-1}\boldsymbol{\lambda} \end{aligned}$$

2.3 Identification

Recall that from (2.4):

$$\boldsymbol{\varepsilon}_t = \Gamma_0\boldsymbol{\zeta}_t \tag{2.7}$$

which gives rise to the system of equations:

$$\begin{aligned} \varepsilon_{1t} &= \gamma_{12}^{(0)}\zeta_{2t} + \gamma_{11}^{(0)}\zeta_{1t} \\ \varepsilon_{2t} &= \gamma_{22}^{(0)}\zeta_{2t} + \gamma_{21}^{(0)}\zeta_{1t} \end{aligned}$$

Next suppose that we want to study the effects of $\zeta_{1t} = 1$ and $\zeta_{2t} = 0$. Can we carry out such an experiment? this is equivalent to asking: does $\boldsymbol{\zeta}_t = [1 \quad 0]'$ result in a unique $\boldsymbol{\varepsilon}_t$ vector? Try and solve the system above. From the first equation we have that $\varepsilon_{1t} = \gamma_{11}^{(0)}$, while from the second equation we have that $\varepsilon_{2t} = \gamma_{21}^{(0)}$. But $\gamma_{21}^{(0)} = \phi_{21}^{(0)}\gamma_{11}^{(0)}$ which in turn implies that $\phi_{21}^{(0)}\varepsilon_{1t} = \varepsilon_{2t}$. We then have two unknowns but only one linearly independent equation; the system is underidentified. In fact, the number of restrictions that we need to identify the structural shocks can be obtained as follows. From (2.7) we know that:

$$\mathbb{V}(\boldsymbol{\varepsilon}_t) = \Sigma_\varepsilon = \Gamma_0\Sigma_{\boldsymbol{\zeta}}\Gamma_0'$$

Now, we have estimates for Σ_ε which, in our current example is a (2×2) matrix. This gives us the RHS but not Γ_0 or $\mathbb{V}(\boldsymbol{\zeta}_t)$ separately. Even if we assume that $\mathbb{V}(\boldsymbol{\zeta}_t) = \Sigma_{\boldsymbol{\zeta}} = I$, so that the structural shocks are orthogonal to each other, $\Sigma_\varepsilon = \Gamma_0 \Gamma_0'$ represents only $(n \times (n+1))/2$ linearly independent restrictions because of the symmetry of Σ_ε. We need exactly $(n \times (n-1))/2$ additional restrictions, in this case, $(n \times (n-1))/2 = 1$. Two popular ways around this problem are to impose either short run or long run restrictions.

2.3.1 Short run identifying restrictions

Take the first approach. If we argue from the structural model that shocks ζ_{2t} contemporaneously affect y_{2t} only, but shocks ζ_{1t} contemporaneously affect both y_{1t} and y_{2t} then we are saying that $\phi_{12}^{(0)} = 0$. Thus, we are imposing zero-restrictions directly on the Γ_0 matrix:

$$\Gamma_0 = \begin{bmatrix} 1 & 0 \\ \gamma_{21}^{(0)} & 1 \end{bmatrix}$$

and now from the same observation of $\zeta_{1t} = 1$ and $\zeta_{2t} = 0$ we can solve for$\varepsilon_{1t} = 1$ and $\varepsilon_{2t} = \gamma_{21}^{(0)}$. This is the so-called triangular identification strategy originally proposed by Sims(1980) which obviously implies a particular ordering for the system. With this identification at hand, we can now study the effects of structural shocks $\boldsymbol{\zeta}_t$ using the VMA representation (2.5)

$$\begin{aligned} \mathbf{X}_t &= \boldsymbol{\mu} + C(L)\,\boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\mu} + C(L)\;\Gamma_0 \boldsymbol{\zeta}_t \end{aligned}$$

so that:

$$\begin{aligned} \mathbf{X}_t &= \boldsymbol{\mu} + C(L) \begin{bmatrix} 1 & 0 \\ \gamma_{20} & 1 \end{bmatrix} \boldsymbol{\zeta}_t \\ &= \boldsymbol{\mu} + \begin{bmatrix} \sum_{j=0}^{\infty} c_{11}^{(j)} L^j & 0 \\ \gamma_{20} \left(\sum_{j=0}^{\infty} c_{21}^{(j)} L^j \right) & \sum_{j=0}^{\infty} c_{22}^{(j)} L^j \end{bmatrix} \boldsymbol{\zeta}_t \end{aligned}$$

where $c_{ik}^{(0)} = 1 \; \forall i, k$. Since $c_{ik}^{(j)}$ are known from "inverting" the estimated reduced form VAR, we can now study the effects of a structural shock.

2.3.2 Long run identifying restrictions

The second popular approach is to impose long-run restrictions on the system. Notice that the matrix of long run multipliers is given by $\Gamma(L)$ evaluated at $L = 1$, that is:

$$\begin{aligned} \Gamma(1) &= \Gamma_0 + \Gamma_1 + \Gamma_2 + ... \\ &= \begin{bmatrix} \gamma_{11}^{(0)} & \gamma_{12}^{(0)} \\ \gamma_{21}^{(0)} & \gamma_{22}^{(0)} \end{bmatrix} + \begin{bmatrix} \gamma_{11}^{(1)} & \gamma_{12}^{(1)} \\ \gamma_{21}^{(1)} & \gamma_{22}^{(1)} \end{bmatrix} + \begin{bmatrix} \gamma_{11}^{(2)} & \gamma_{12}^{(2)} \\ \gamma_{21}^{(2)} & \gamma_{22}^{(2)} \end{bmatrix} + ... \\ &= \begin{bmatrix} \sum_{j=0}^{\infty} \gamma_{11}^{(j)} L^j & \sum_{j=0}^{\infty} \gamma_{12}^{(j)} L^j \\ \sum_{j=0}^{\infty} \gamma_{21}^{(j)} L^j & \sum_{j=1}^{\infty} \gamma_{22}^{(j)} L^j \end{bmatrix} \end{aligned}$$

Suppose, for instance, that shocks ζ_{2t} are deemed to have only temporary or short run effects, while ζ_{1t} are thought to have permanent or long run effects. That is:

$$\Gamma(1) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

then we can use the equivalence established in (2.6) to write:

$$\begin{aligned} \Gamma(1) &= C(1)\ \Gamma_0 \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} &= \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \gamma_{11}^{(0)} & \gamma_{12}^{(0)} \\ \gamma_{21}^{(0)} & \gamma_{22}^{(0)} \end{bmatrix} \end{aligned}$$

where $C_{ik} = \sum_{j=0}^{\infty} c_{ik}^{(j)}$. The above produces, in principle, a system of four equations on four unknowns because the $c_{ik}^{(j)}$ are known from estimation. However, this results in only one linearly independent restriction:

$$\gamma_{11}^{(0)} = \frac{C_{22}}{C_{11}C_{22} - C_{21}C_{12}} = -\frac{C_{12}}{C_{22}}\gamma_{12}^{(0)} = \gamma_{21}^{(0)} = -\frac{C_{12}}{C_{22}}\gamma_{22}^{(0)}$$

that is, the long-run neutrality assumption of ζ_{2t} indirectly restricts the impact multiplier matrix to be:

$$\Gamma_0 = \begin{bmatrix} \frac{C_{22}}{C_{11}C_{22}-C_{21}C_{12}} & \frac{C_{12}}{C_{21}C_{12}-C_{11}C_{22}} \\ \frac{C_{22}}{C_{11}C_{22}-C_{21}C_{12}} & \frac{C_{12}}{C_{21}C_{12}-C_{11}C_{22}} \end{bmatrix}$$

Chapter 3

State Space Models

3.1 State Space Representation

The idea is to express a dynamical system in a state space form. Suppose we have a state vector $\boldsymbol{\alpha}$. We cannot observe $\boldsymbol{\alpha}$ but we know can be expressed as:

$$\underset{(r\times 1)}{\boldsymbol{\alpha}_t} = \underset{(r\times 1)}{\mu_t} + \underset{(r\times r)}{\mathbb{F}_t}\,\underset{(r\times 1)}{\boldsymbol{\alpha}_{t-1}} + \underset{(r\times r)}{\mathbb{Q}_t}\,\underset{(r\times 1)}{\boldsymbol{\upsilon}_t} \tag{SS1}$$

this is sometimes called the state transition equation. On the other hand, we do observe a vector $\mathbf{y}_t$ which we know is related to the unobserved state in the form:

$$\underset{(n\times 1)}{\mathbf{y}_t} = \underset{(n\times r)(r\times 1)}{\mathbb{H}_t}\,\underset{}{\boldsymbol{\alpha}_t} + \underset{(n\times k)(k\times 1)}{\mathbb{A}_t}\,\underset{}{\mathbf{z}_t} + \underset{(n\times n)(n\times 1)}{\mathbb{R}_t}\,\underset{}{\boldsymbol{\varepsilon}_t} \tag{SS2}$$

this is called the measurement equation and (SS1)-(SS2) is called the state space (SS) representation of the system under study. In this general form, the matrices $\mathbb{F}_t, \mathbb{H}_t, \mathbb{A}_t, \mathbb{Q}_t, \mathbb{R}_t$ may be functions of the data, structural parameters, or constants. The system above has an inherent identification issue: if we replace $\boldsymbol{\alpha}_t$ in the measurement equation, we obtain a mixture of errors $\boldsymbol{\upsilon}_t$ and $\boldsymbol{\varepsilon}_t$ which of course we cannot identify separately. The following assumptions are required to attempt identification:

A1. The number of shocks in the system is $\geq$ the number of observables.

A2. $\boldsymbol{\upsilon}_t, \boldsymbol{\varepsilon}_t \sim N(0, \mathbb{I})$ and $\mathbb{E}[\boldsymbol{\upsilon}_s \boldsymbol{\varepsilon}_t] = 0 \;\forall\; s, t$.

Example 11 *Consider the following unobserved component model. Suppose that two firms produce certain good and we cannot observe their individual output but only the total industry output:*

$$\mathbf{x}_t = x_{1t} + x_{2t}$$

Furthermore suppose that:

$$\begin{aligned} x_{1t} &= \delta + x_{1,t-1} + \varepsilon_{1t} \text{ with } \varepsilon_{1t} \sim N\left(0, \sigma_1^2\right) \\ x_{2t} &= \phi_1 x_{2,t-1} + \phi_2 x_{2,t-2} + \varepsilon_{2t} \text{ with } \varepsilon_{2t} \sim N\left(0, \sigma_2^2\right) \\ \text{and:} &\quad \mathbb{E}[\varepsilon_{1t}, \varepsilon_{2t}] = 0 \;\forall\; t \end{aligned}$$

then $\mathbf{x}_t$ is made up of a unit root with drift component and a (AR(2)) stationary component. To espress this system as (SS1)-(SS2) let $\mathbf{y}_t = \mathbf{x}_t$ and $\boldsymbol{\alpha}_t = [x_{1,t} \;\; x_{2,t} \;\; x_{2,t-1}]'$. Then the measurement equation becomes:

$$\mathbf{y}_t = [1 \;\; 1 \;\; 0]\,\boldsymbol{\alpha}_t$$

while the state transition equation is:

$$\boldsymbol{\alpha}_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{2,t-1} \end{bmatrix} = \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \\ x_{2,t-2} \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

3.2 The Kalman filter

The Kalman filter is a recursion that allows us to obtain the best forecast for the latent variable $\boldsymbol{\alpha}_t$ given the observation of $\mathbf{y}_t$. Let:

$$\begin{aligned} \Psi_t &\equiv \text{ information set at } t \\ \boldsymbol{\alpha}_{t|t-1} &\equiv \mathbb{E}\left[\boldsymbol{\alpha}_t | \Psi_{t-1}\right] \\ \boldsymbol{\alpha}_{t|t} &\equiv \mathbb{E}\left[\boldsymbol{\alpha}_t | \Psi_t\right] \\ P_{t|t-1} &\equiv \mathbb{E}\left[\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}\right)\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}\right)'\right] \\ P_{t|t} &\equiv \mathbb{E}\left[\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}\right)\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}\right)'\right] \end{aligned}$$

where $P_{t|t-1}$ and $P_{t|t}$ are covariance matrices of $\boldsymbol{\alpha}_{t|t-1}$ and $\boldsymbol{\alpha}_{t|t}$ respectively. Next define the observables:

$$\begin{aligned} \mathbf{y}_{t|t-1} &\equiv \mathbb{E}\left[\mathbf{y}_t | \Psi_{t-1}\right] \quad \text{("best forecast")} \\ \boldsymbol{\eta}_{t|t-1} &\equiv \mathbf{y}_t - \mathbf{y}_{t|t-1} \quad \text{(forecast error)} \\ \mathcal{F}_{t|t-1} &\equiv \mathbb{E}\left[\boldsymbol{\eta}_{t|t-1}^2\right] \end{aligned}$$

Finally define $\boldsymbol{\alpha}_{t|t}$ as the "filetered" estimate of $\boldsymbol{\alpha}_t$ which uses information up to and including t and $\boldsymbol{\alpha}_{t|T}$ as the "smoothed" estimate which includes information in the entire sample (i.e., including information available only in periods $s > t$). If the elements of the matrices $\mathbb{F}, \mathbb{H}, \mathbb{A}, \mathbb{Q}, \mathbb{R}$ are known, the Kalman filter recursion is all that is needed to estimate the state vector. Otherwise they can be estimated as described in the next section. The recursion comprises the following steps:

1. Initialize the filter with an initial value for $\boldsymbol{\alpha}_{1|0}$ and $P_{1|0}$.

2. Prediction step: use the prediction equation:

$$\begin{aligned} \boldsymbol{\alpha}_{t|t-1} &= \mu + \mathbb{F}\boldsymbol{\alpha}_{t-1|t-1} \\ P_{t|t-1} &= \mathbb{F}P_{t-1|t-1}\mathbb{F}' + \mathbb{Q}\mathbb{Q}' \\ \mathbf{y}_{t|t-1} &= \mathbb{H}\boldsymbol{\alpha}_{t|t-1} + \mathbb{A}\mathbf{z}_t \end{aligned}$$

3. Observe $\mathbf{y}_t$ and obtain the forecast errors and covariance matrix:

$$\begin{aligned} \boldsymbol{\eta}_{t|t-1} &= \mathbf{y}_t - \mathbf{y}_{t|t-1} = \mathbf{y}_t - \mathbb{H}\boldsymbol{\alpha}_{t|t-1} - \mathbb{A}_t\mathbf{z}_t \\ \mathcal{F}_{t|t-1} &= \mathbb{E}\left[\boldsymbol{\eta}_{t|t-1}^2\right] = \mathbb{H}P_{t|t-1}\mathbb{H}' + \mathbb{R}\mathbb{R}' \end{aligned}$$

4. Update step: use the updating equations:

$$\begin{aligned} \boldsymbol{\alpha}_{t|t} &= \boldsymbol{\alpha}_{t|t-1} + \overbrace{P_{t|t-1}\mathbb{H}'\mathcal{F}_{t|t-1}^{-1}}^{Kalman\ gain}\boldsymbol{\eta}_{t|t-1} = \boldsymbol{\alpha}_{t|t-1} + K_t\boldsymbol{\eta}_{t|t-1} \\ P_{t|t} &= P_{t|t-1} - K_t\mathbb{H}P_{t|t-1} \end{aligned}$$

5. Repeat steps 2-5.

Step 1 is crucial for the recurssion; if the eigenvalues of the matrix $\mathbb{F}$ are known to be all inside the unit circle, then $\boldsymbol{\alpha}_t$ is a stationary Markov process and as such has an ergodic distribution with constant mean and variance. These unconditional first moments (based on the assumptions about $\boldsymbol{\upsilon}_t$) can be used as starting values:

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{\alpha}_{t+1}\right] &= \mu + \mathbb{F}\mathbb{E}\left[\boldsymbol{\alpha}_t\right] \\ &\Rightarrow \mathbb{E}\left[\boldsymbol{\alpha}_t\right] = \left(I - \mathbb{F}\right)^{-1}\mu \end{aligned}$$

since $\boldsymbol{\alpha}_t$ is covariance stationary and therefore $\mathbb{E}\left[\boldsymbol{\alpha}_{t+1}\right] = \mathbb{E}\left[\boldsymbol{\alpha}_t\right]$. Likewise:

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{\alpha}_{t+1}\boldsymbol{\alpha}_{t+1}'\right] &= \mathbb{F}\mathbb{E}\left[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t'\right]\mathbb{F}' + \mathbb{E}\left[\boldsymbol{\upsilon}_t\boldsymbol{\upsilon}_t'\right] \\ &\Rightarrow \Sigma = \mathbb{F}\Sigma\mathbb{F}' + \mathbb{Q}\mathbb{Q}' \end{aligned}$$

a discrete Lyapunov equation which can be solved for Σ. If, on the other hand, some of the eigenvalues of the matrix $\mathbb{F}$ are known to be on the unit circle, then $\boldsymbol{\alpha}_t$ is non-stationary and several options are available (see section 3 in this essay). In step four the Kalman gain is obtained as the optimal stepsize adjustment to be made in $\boldsymbol{\alpha}_{t|t}$ once $\mathbf{y}_t$ is available and we realize the magnitude of our mistake $\boldsymbol{\eta}_{t|t-1}$.

An additional step is to obtain the "smoothed" estimates of the state vector $\boldsymbol{\alpha}_{t|T}, P_{t|T}$. After performing the recursion above, we obtain and store the sequences $\left\{\boldsymbol{\alpha}_{t|t}, P_{t|t}\right\}_{t=1}^{T}$ and $\left\{\boldsymbol{\alpha}_{t+1|t}, P_{t+1|t}\right\}_{t=1}^{T-1}$. The smoothed estimate for the final date is just the last element of $\left\{\boldsymbol{\alpha}_{t|t}, P_{t|t}\right\}_{t=1}^{T}$. Next, generate and store $\left\{J_t = P_{t|t}\mathbb{F}'P_{t+1|t}^{-1}\right\}_{t=1}^{T}$ and then use the stored sequences to calculate:

$$\begin{aligned} \boldsymbol{\alpha}_{t|T} &= \boldsymbol{\alpha}_{t|t} + J_t\left(\boldsymbol{\alpha}_{t+1|T} - \mathbb{F}\boldsymbol{\alpha}_{t|t} - \mu\right) \\ P_{t|T} &= P_{t|t} + J_t\left(P_{t+1|T} - P_{t+1|t}\right)J_t' \end{aligned}$$

for $t = T-1, T-2, ...$

3.3 Predictive decomposition of the likelihood

At this point we assume that

$$\mathbf{y}_t|\Psi_{t-1} \sim N\left(\mathbf{y}_{t|t-1}, \mathcal{F}_{t|t-1}\right)$$

so that we can write the joint density of observing $(y_1, ..., y_T)$ as:

$$\Pr\left(\mathbf{y}_1, ..., \mathbf{y}_T\right) = \prod_{i=1}^{T}\Pr\left(\mathbf{y}_t|\Psi_{t-1}\right)\Pr\left(\mathbf{y}_0\right)$$

and then the log-likelihood may be written:

$$\mathcal{L} = -\frac{1}{2}\sum_{t=1}^{T}\log\left(2\pi|\mathcal{F}_{t|t-1}|\right) - \frac{1}{2}\sum_{t=1}^{T}\left[\boldsymbol{\eta}_{t|t-1}'\mathcal{F}_{t|t-1}^{-1}\boldsymbol{\eta}_{t|t-1}\right] \tag{3.1}$$

so the Kalman filter is going to give us $\boldsymbol{\eta}_{t|t-1}$ and $\mathcal{F}_{t|t-1}$. Naturally, maximization of (3.1) is by numerical methods and therefore we need initial guesses for the elements of the matrices $\mathbb{F}, \mathbb{H}, \mathbb{A}, \mathbb{Q}, \mathbb{R}$.