607
Covariance stationarity and related analytics

Jon Faust

http://e105.org/e607

September 6, 2016

- **Cov. stat. processes**
  - Cov. stat: first two moments finite and time invariant

    Second central moments include variances and (auto)covariacnes
  - Few macro series probably satisfy this restriction on invariance of moments

    already our guest presentations have provided ample evidence of this
  - But the cov. stat. paradigm is still very important
  - Especially for anyone doing dynamic econmomic theory/game theory

    much of this work involves cov. stat. processes
  - And cov. stat. paradigm gives us key insights for applied work
- **Remember**
  - If Y $(T \times 1)$ is cov. stat. then the vcov matrix of Y is s.t. each diagonal has only one value

    j^{th} diagonal has $\sigma(j)$.
- **Variance-covariance matrix of cov. stat.**
  - Write $Y = (y_1, \ldots, y_T)'$ as a vector.
  - variance-covariance is, $E(Y - i\mu)(Y - i\mu)' = \text{vcov}(y) \equiv V =$

$$\begin{bmatrix} \sigma_0 & . & . & . & . \\ \sigma_1 & \sigma_0 & . & . & . \\ \sigma_2 & \sigma_1 & \ddots & . & . \\ \vdots & \ddots & \ddots & \ddots & . \\ \sigma_{T-1} & \ddots & \sigma_2 & \sigma_1 & \sigma_0 \end{bmatrix}$$

- Aside: This kind of matrix called Toeplitz and has many special properties.

  Can think of much of cov. stat. ecmet as the study of ℓ^2 or alternatively as the study of Toeplitz matrices if you are so inclined

- Lovely book:

  Toeplitz Forms and Their Application, by Ulf Grenander & Gábor Szegö, California monographs in mathematical sciences, 1958

▶ **Remember:Wold decomp. thm.**

- Wold decomp. thm. tells us,

  every cov.stat. process can be represented as

$$y_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$$

  where...

- $E\varepsilon_t = 0$, $E\varepsilon_t \varepsilon_s = 0 \; t \neq s$, $E\varepsilon_t^2 = \sigma^2$
- Details: here we are ignoring deterministic bits such as a nonzero mean

▶ **Next**

- Define AR, MA, ARMA processes
- And give the relations among them
- And some math tools for manipulating them

▶ MA **representation**

- The gen. lin. model also is an infinite-order moving average $MA(\infty)$ representation
- y_t is a moving average of the driving variables or shocks–the εs.

▶ $MA(p)$

- If there are p terms in the sum, we have a p^{th} order moving average:
- $MA(p)$:

$$y_t = \sum_{j=0}^{p} a_j \varepsilon_{t-j}$$

▶ **Autoregressive model:** $AR(p)$

- In the $AR(q)$ model, we have

$$y_t = \mu + \sum_{j=1}^{q} \rho_j y_{t-j} + \varepsilon_t$$

- **ARMA(p,q)**
  - ARMA(p,q)

$$y_t = \alpha + \sum_{j=1}^{p} \rho_j y_{t-j} + \sum_{i=0}^{q-1} \theta_i \varepsilon_{t-i}$$

- **Useful facts**
  - Every cov. stat. process has an MA representation

    Wold decomp. tells us this

  - And an AR representation

    We'll explore below

  - Hence, trivially, an ARMA rep.
  - Note: these claims require us to consider infinite order processes

- **What mean by 'representation'**
  - There are some deep issues re: what we might mean every covariance stationary process has a given 'representation'
  - Suppose $z_t, t = 1, 2, \ldots, T$ is cov. stat. with joint dist, F_T.
  - I can write an MA process:

$$y_t = \sum \theta_j \varepsilon_{t-j}$$

    such that $y_t, t = 1, 2, \ldots, T$ has joint dist F_T where ε is as in the Wold decomp. thm.

  - Similarly, I can write an AR process

$$y_t = \sum \rho_j y_{t-j} + \varepsilon_t$$

    satisfying the same conditions

  - Fairly generally in this class, we will focus on the first two moments (means and variances) of the observable variable.
  - Of course, the conditions for cov. stat. are only about the first two moments
  - Thus, the Wold decomp. thm. obviously can't say ε is indep. or iid.

    Or say anyting about the higher moments.

  - Thus, in our statements about representations, the key piece is that that we can match the first and second moment properties of any cov. stat. with an AR, MA, and ARMA
  - Of course, if the εs are Gaussian, this means we match all properties…

Two moments fully characterize Gaussians

- **So far**
  - Defined cov. stat.
  - And said any cov. stat. process has AR, MA, and ARMA rep.
  - As in most math, the key is notation
  - See a lovely lecture by Don Knuth (who invented TeX in order to revise one of his books) on the subject of notation

    go

    http://stanford-online.stanford.edu/seminars/knuth/031017-knuth-100.asx

- **Lag Operator Notation**
  - $Ly_t = y_{t-1}$
    $L^p y_t = LL \times \ldots \times Ly_t = y_{t-p}$
  - MA(q) in lag op. notation:

$$y_t = \theta(L)\varepsilon_t,$$

$$\theta(L) = \sum_{j=0}^{q} L^j \theta_j, \quad \theta_0 = 1$$

  - Throughout ε_t is not serially correlated. Ok to think of it as iid unless stated otherwise.

- **AR**
  - AR(p):

$$\begin{aligned} \rho(L)y_t &= \varepsilon_t \\ \rho(L) &= \sum_{j=0}^{p} L^i \rho_i \end{aligned}$$

    $\rho_0 = 1$

- **Danger Danger Danger**
  - In this notation, the ρ parameters are the negative of when we state the model in std. OLS notation:

$$y_t = \sum_{i=1}^{p} \rho_i y_{t-i} + \varepsilon_t$$

- **ARMA**

- ARMA(p,q):

$$\rho(L)y_t = \theta(L)\varepsilon_t$$

$$\rho_0 = 1, \theta_0 = 1$$

▶ **A preview**

- You can use standard algebraic operations on lag polynomials.
- Thus, can relate AR to MA by noting the AR

$$\rho(L)y_t = \varepsilon_t$$

has the same second moment propertis as the MA:

$$y_t = \rho(L)^{-1}\varepsilon_t$$

- Understanding lag polynomials and their inverses reveals much about time series.

▶ **A brief excursion into inversion.**

- ... and complex numbers, etc.

▶ **Treat L like any variable in algebra**

- We have a polynomial,

$$A(z) = 1 + a_1 z + \ldots + a_p z^p$$

and we are interested in the case of $z = L$.

- We can borrow classic results from algebra to manipulate this polynomial
- Fundamental thm. of algebra: every p^{th} order poly. is a product of p first order polys.

$$\begin{aligned} A(z) &= 1 + a_1 z + \ldots + a_p z^p \\ A(z) &= \Pi_{k=1}^{p}(1 - (1/\lambda_k)z) \end{aligned}$$

- The λ_ks are the 'roots' of the polynomial

Roots also called 'zeros'; for each of the k roots, we have $A(\lambda_k) = 0$.

▶ **Why is this factoring impt?**

- Since every polynomial is the product of a bunch of first order polynomials, all we need to understand is the first-order case

and how to multiply

- **Aside: Complex numbers**
  - Fund. Thm. holds in the complex domain.

    Even if the coefficients of A are real, the roots may be complex.
  - Complex number can be represented ordered pair of reals $\{a, b\}$:

$$c = a + bi$$

    where $i \equiv \sqrt{-1}$
  - Defn: Complex conjugate is $c^* = a - bi$.

    just flip the sign of the imaginary part
  - Defn: Modulus is $\sqrt{cc^*} = \sqrt{a^2 + b^2}$, so a positive real.
  - Modulus often written $|c|$: it is the complex generalization of abs. val.
- **Aside^2: Modulus**
  - If we treat the complex number $c = (a, b)'$ as a vector of two reals, the modulus is its Euclidean length.

    in this sense, modulus is a measure of how big c is.
  - If plot ray from origin to (a, b) in Euclidean space, we see a length and a direction

    Thus giving rise to the polar notation for complex numbers.
  - Modulus is the length, a direction-free measure of size of c.
- **End of asides, back ARs and MAs**
  - We can relate ARs to MAs by inverting the lag polynomial
  - That is, if $\rho(L)y_t = \varepsilon_t$ is an AR, then the equivalent MA is $y_t = \rho(L)^{-1}\varepsilon_t$.
  - We just have to figure out what this type of division is.
- **Polynomial algebra: division**
  - Suppose,

$$A(z) = \Pi_k(1 - \lambda_k^{-1} z)$$

  - And suppose divsion works in the natural way so that $(xy)^{-1} = x^{-1} \times y^{-1}$.
  - Thus, the polynomial inverse is:,

$$A(z)^{-1} = \Pi_k \left((1 - \lambda_k^{-1} z)^{-1}\right)$$

- Inverse of a poly. is product of inverses of its first-order factors
- Thus, all we must understand is inverse of first order polys.

  illustrates a general point we will see several versions of: in cov. stat. time series, all you really need to do is understand the first order case really well.

▶ **Inverse of first order poly**

- Let's try guess and verify to get the inverse
- Suppose $\rho(z) = (1 - \rho z)$ and assume $(1 - \rho z)^{-1}$ exists and itself is a (potentially infinite order) polynomial, $\theta(z) = (1 + \theta_1 z + \theta_2 z^2 + \ldots)$
- If the inverse means the usual thing then,

$$\rho(z)\theta(z) = 1$$

  Or expanded:

$$(1 - \rho z)(1 + \theta_1 z + \theta_2 z^2 + \ldots) = 1$$

- We can multiply this out and solve for the θs.

  In the product, all terms in z must cancel

▶ **Question**

- What is θ_1?

  ρ

- And θ_2

  ρ^2

- and so forth ...

▶ **The inverse**

- Thus,

$$(1 - \rho L)^{-1} \equiv \theta(L) = \sum_{j=0}^{\infty} \rho^j L^j$$

  $\theta_0 = 1$.

- This is not a proof because we need some sort of reason to believe we can take this to the infinite limit.

  Hint: we'll need $|\rho| < 1$ or the coefficients in the sum are exploding.

▶ **Implications for AR(1)**

- The AR(1) has an representation as an MA(∞) with coefs. $\theta_j = \rho^j$.
- You can easily verify that this AR and MA have the same autocov. function.
- Can do many things with this relation

▶ **Variance of AR(1)**

- Using the MA representation,

$$y_t = \sum \rho^j \varepsilon_{t-j}$$

  what is the variance?

$$\sigma_\varepsilon^2 \sum (\rho^j)^2 = \frac{\sigma^2}{1-\rho^2}$$

- What did we assume?

  that the sum converges

- And what does this require?

  $|\rho| < 1$

▶ **Alterantive route to variance of AR(1)**

- Using stationarity.
- What is the mean of

$$y_t = \rho y_{t-1} + \varepsilon_t$$

  (using our usual assumptions on ε)?

  Take expectations of both sides:

$$Ey = \rho Ey + 0$$

  or $Ey = 0$ so long as $\rho \neq 1$ (ooh, the unit root case!).

- Thus, the variance of y_t is simply Ey_t^2:

$$\begin{aligned} \sigma_y^2 &= Ey_t^2 \\ Ey_t^2 &= \rho^2 Ey_{t-1}^2 + E\varepsilon_t^2 + \text{zero} \end{aligned}$$

  in the last expression, the expectation of the cross product term is zero.

- And under cov. stationarity:

$$Ey_t^2 = Ey_{t-1}^2 = \sigma_y^2$$

▶ **Thus,**

- $$\sigma_y^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

▶ **Notice**

- Using the MA rep. we found that we needed that $|\rho| < 1$ for the sum to converge.
- How can we see that same issue when deriving the variance using the formula for the variance?

  For $|\rho| = 1$ we are dividing by zero; for $|\rho| > 1$, our expression for the variance is negative.

▶ **Bottom line on stationarity of AR(1)**

- By defn. covariance stationarity requires finite variance
- In the AR(1) case, $|\rho| < 1$ is necess. and suffic

▶ **Aside: Language**

- If $\rho(L) = (1 - \rho L)$, then the root of this polynomial is $1/\rho$.
- Thus, we want the absolute value of the root to be greater than 1.
- And, in general, we allow complex roots . . .

  so we can call the relevant absolute value 'the modulus.'

- Thus, for stationarity of the AR(1) we say we want the root to have modulus greater than one.

▶ **And remember polar notation**

- When we plot the complex number as a vector, modulus greater than 1 amounts to having length greater than 1

  Implying the length takes us outside the unit circle.

- Thus, the most common language is: stationarity of the AR(1) requires that the root is outside the unit circle

▶ **Aside:: Ouch, ouch, ouch**

- Sometimes (especially when there is only 1 root involved) folks are sloppy and refer to ρ as the root

  rather than $1/\rho$.

▶ **Stationarity in the general AR(p)**

- $\rho(L)y_t = \varepsilon_t$, p^{th} order

- Given that a p^{th} order poly. is the product of p first order polys.
- If each root outside unit circle, then for each term if we invert we'll get something of the form,

$$\sum_{j=0}^{\infty}((1/\lambda)^j)$$

- And the full inverse will be of the form,

$$\rho(L)^{-1} = \pi_{j=1}^{p}\left(\sum_{k=0}^{\infty}(1/\lambda_j)^k\right)$$

where the λs are the roots of $\rho(L)$.

- Thus, the inverse will generally be an infinite order polynomial
- And the coefficients will be square summable

  Each of the infinite sums has square summable coefs, so the poly. implied by the product will as well. You should be able to show this

▶ **Thus,**

- The AR(p) is cov. stationary iff all roots are outside the unit circle.

▶ **Aside:: Usage note**

- Roots on the unit circle are called 'unit roots' bn As you know, AR processes with a unit root are nonstationary and do not have finite variance
- The AR(1) unit root process is called a ...

  random walk. If there is a constant in the AR(1) it is a random walk with drift

- AR roots strictly inside unit circle are often called unstable or explosive

▶ **The MA case**

- In some ways the MA case is simpler
- The MA(q):

$$y_t = \theta(L)\varepsilon_t = \sum_{j=0}^{q}\theta_j\varepsilon_{t-i}$$

- What is the variance?

  $\sigma_\varepsilon^2\sum\theta^2$

- Does this require parameter restrictions?

No. Well, not so long as we take finite q, θ_js, and σ_ε^2

- **Stationarity in the MA(q)**
  - The MA(q) for finite q is stationary in all cases

    requiring only finite parameters

- **Inverse of the MA(q)**
  - We'd like to write the MA as an AR:

$$\begin{aligned} y_t &= \theta(L)\varepsilon_t \\ \theta(L)^{-1} y_t &= \varepsilon_t \end{aligned}$$

  - We know that doing this requires...

    that the roots of the poly. are outside unit circle

- **Inverse of the MA(q)**
  - The finite-order MA(q) is stationary in all cases

    Requiring no restriction on roots

  - The MA(q) is 'invertible' giving an equivalent AR iff all roots are outside the unit circle.

- **Aside:: Noninvertible**
  - If the MA is not invertible, the process can still be matched arbitrarily well by some AR
  - How? Well, every MA(p) $p > 0$ has multiple equivalent MA representations.
  - And one and only one of these is invertible

- **Aside::**
  - Thus, to find the AR 'that matches' a noninvertible MA, first find the invertible MA equivalent to the original MA, and then invert it!
  - We'll talk a bit more about noninvertible MAs later.

    For now, just a hint.

- **Two unit root problems**
  - Unit roots in the AR case give rise to nonstationarity, infinite variance, etc.

- When you hear folks talking about things like 'unit root econometrics' they generally mean 'unit AR roots'

▶ **Unit MA roots**

- Noninvertible MA roots present their own problems
- In practical estimation this is also part of the pileup problem we will be discussing.

  In finite samples, the likelihood has a 'false' maximum at exactly a unit MA root for a set of sampls with pos. probability

▶ **Bonus topic: Impulse responses**

- Take an MA rep:

$$y_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$$

- This expresses observable y in terms of the current and past of assumed exogenous variable, ε.
- The as (viewed as a function of the integers) are also called the 'impulse response function'

  a_j gives how y_{t+j} is affected by a unit impulse, $\varepsilon = 1$, at time t

▶ **Question**

- What is the impulse resp. function of an AR(1)?

  $\rho^0, \rho^1, \ldots$

- An MA(1)?

  $1, \theta, 0, \ldots$

▶ **Impulse responses and dynamic forecast**

- Take AR(1). What is $E_t y_{t+1}$ given $y_t = 0$ and information $\varepsilon_{t+1} = e$?

  e

- And $E_t y_{t+k}$?

$$\rho^{k-1} e$$

▶ **General statement**

- ARMA, AR, or MA:

$$\rho(L) y_t = \theta(L) \varepsilon_t$$

- The impulse response to any shock can be calculated by 'forecasting the model' from time t with
  - all ys dated t or before set to zero
  - All ε set to zero EXCEPT
  - $\varepsilon_{t+1} = 1$
- Note: By 'forecast' we mean solving the model forward recursively
- If you have an AR model, the associated MA rep. is the impulse resp. function.
- You can compute it by inverting the AR lag polynomial, or
- By forecasting
- Many computer programs invert lag polynomials using this forecasting approach.

▶ **Economic interp. of MA rep**

- The impulse response at horizon k can be seen as the change in the expectation of y at $t+k$ due to a unit change in the expectation of ε_{t+1}.
- Thus, can see the impulse response function as the reaction of agents' expectations to 'news'

  How would my expectation of GDP or interest rates or inflation in a years time change if I received news that there will be a productivity shock tomorrow?

▶ **Summary for 3 reps. of cov. stat.**

- Autocov. function, AR and MA reps. are each a potentially infinite sequence of numbers characterizing a cov. stat. process.

▶ **The autocov. func.**

- Autocov. fully summarizes the observable second moment properties

▶ **The MA rep.:**

- Gives the future of y in terms of history of exogenous (but unobservable) variable (ε)
- Stationarity is pretty trivial viewed in terms of the MA rep.

  If finite order, done. If infinite order, need square summable as
- This rep. makes calculating variance and covariances easy
- Invertibility is a bit of a pain

▶ **The AR rep.**

- Summarizes expectation of future ys as a function of history of observables only.

- Stationarity is a bit of a pain.

- **An issue left dangling**
    - Every MA has multiple MA representations. What should we do about this?

      Worry about it later.