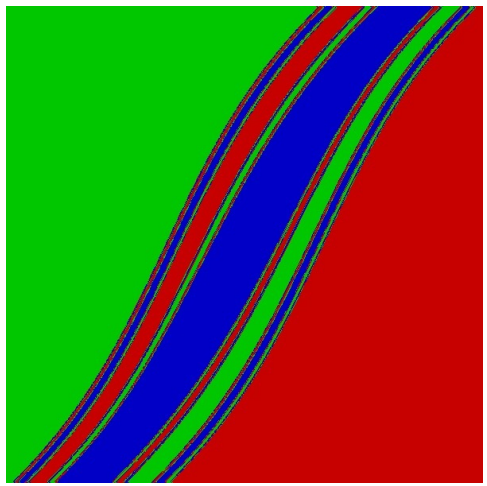
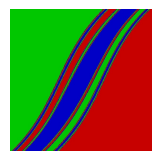
Time Series Econometrics
A Concise Course

Francis X. Diebold
University of Pennsylvania

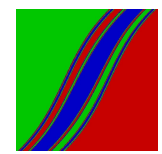


February 17, 2020

The graphic is by Peter Mills and was obtained from Wikimedia Commons.



Time Domain: The General Linear Process and its Approximation

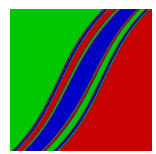


The Environment

Time series Y_t (doubly infinite)

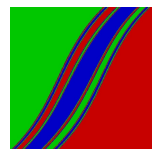
Realization y_t (again doubly infinite)

Sample path $y_t,\ t\ =\ 1,\ ...,\ T$



Strict Stationarity

Joint cdfs for sets of observations
depend only on displacement, not time.



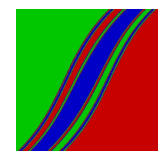
Weak Stationarity

(Covariance stationarity, second-order stationarity, ...)

$$Ey_t = \mu, \ \forall t$$

$$\gamma(t, \tau) = E(y_t - Ey_t)(y_{t-\tau} - Ey_{t-\tau}) = \gamma(\tau), \ \forall t$$

$$0 < \gamma(0) < \infty$$



Autocovariance Function

(a) symmetric

$$\gamma(\tau) = \gamma(-\tau), \ \forall \tau$$

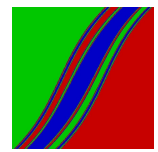
(b) nonnegative definite

$$a'\Sigma a \geq 0, \ \forall a$$

where Toeplitz matrix Σ has ij-th element $\gamma(i-j)$

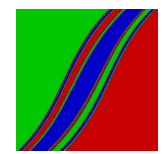
(c) bounded by the variance

$$\gamma(0) \geq |\gamma(\tau)|, \ \forall \tau$$



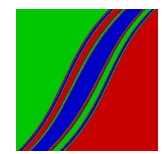
Autocovariance Generating Function

$$g(z) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\, z^{\tau}$$



Autocorrelation Function

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$



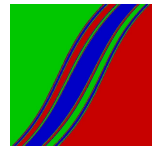
White Noise

White noise: $\varepsilon_t \sim WN(\mu, \sigma^2)$ (*serially uncorrelated*)

Zero-mean white noise: $\varepsilon_t \sim WN(0, \sigma^2)$

Independent (strong) white noise: $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$

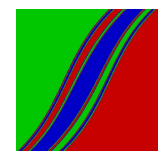
Gaussian white noise: $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$



Unconditional Moment Structure of Strong White Noise

$$E(\varepsilon_t) = 0$$

$$var(\varepsilon_t) = \sigma^2$$



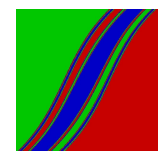
Conditional Moment Structure of Strong White Noise

$$E(\varepsilon_t|\Omega_{t-1}) \;=\; 0$$

$$var(\varepsilon_t|\Omega_{t-1}) \;=\; E[(\varepsilon_t \;-\; E(\varepsilon_t|\Omega_{t-1}))^2|\Omega_{t-1}] \;=\; \sigma^2$$

where

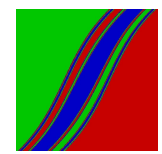
$$\Omega_{t-1} \;=\; \varepsilon_{t-1},\; \varepsilon_{t-2},\; \ldots$$



Autocorrelation Structure of Strong White Noise

$$\gamma(\tau) = \begin{cases} \sigma^2, & \tau = 0 \\ 0, & \tau \geq 1 \end{cases}$$

$$\rho(\tau) = \begin{cases} 1, & \tau = 0 \\ 0, & \tau \geq 1 \end{cases}$$



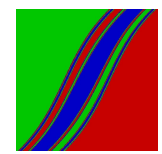
An Aside on Treatment of the Mean

In theoretical work we assume a zero mean, $\mu = 0$.

This reduces notational clutter and is without loss of generality.

(Think of y_t as having been centered around its mean, μ,
and note that $y_t - \mu$ has zero mean by construction.)

(In empirical work we allow explicitly for a non-zero mean,
either by centering the data around the sample mean
or by including an intercept.)



The Wold Decomposition

Under regularity conditions,
every covariance-stationary process $\{y_t\}$ can be written as:

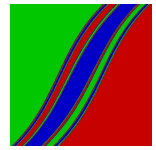
$$y_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

where:

$$b_0 = 1$$

$$\sum_{i=0}^{\infty} b_i^2 < \infty$$

$$\varepsilon_t = [y_t - P(y_t|y_{t-1}, y_{t-2}, ...)] \sim WN(0, \sigma^2)$$



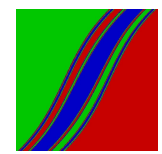
The General Linear Process

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

$$b_0 = 1$$

$$\sum_{i=0}^{\infty} b_i^2 < \infty$$

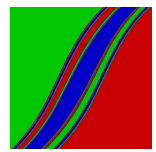


Unconditional Moment Structure (Assuming Strong WN Innovations)

$$E(y_t) = E\left(\sum_{i=0}^{\infty} b_i \varepsilon_{t-i}\right) = \sum_{i=0}^{\infty} b_i E\varepsilon_{t-i} = \sum_{i=0}^{\infty} b_i \cdot 0 = 0$$

$$var(y_t) = var\left(\sum_{i=0}^{\infty} b_i \varepsilon_{t-i}\right) = \sum_{i=0}^{\infty} b_i^2 var(\varepsilon_{t-i}) = \sigma^2 \sum_{i=0}^{\infty} b_i^2$$

(Do these calculations use the strong WN assumption? If so, how?)



Conditional Moment Structure (Assuming Strong WN Innovations)

$$E(y_t|\Omega_{t-1}) = E(\varepsilon_t|\Omega_{t-1}) + b_1 E(\varepsilon_{t-1}|\Omega_{t-1}) + b_2 E(\varepsilon_{t-2}|\Omega_{t-1}) + \ldots$$

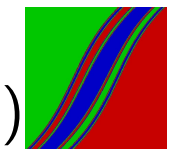
$$(\Omega_{t-1} = \varepsilon_{t-1},\ \varepsilon_{t-2},\ \ldots)$$

$$= 0 + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \ldots = \sum_{i=1}^{\infty} b_i\varepsilon_{t-i}$$

$$var(y_t|\Omega_{t-1}) = E[(y_t - E(y_t|\Omega_{t-1}))^2|\Omega_{t-1}]$$

$$= E(\varepsilon_t^2|\Omega_{t-1}) = E(\varepsilon_t^2) = \sigma^2$$

(Do these calculations use the strong WN assumption? If so, how?)



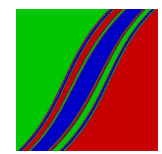
Autocovariance Structure

$$\gamma(\tau) \;=\; E\left[\left(\sum_{i=-\infty}^{\infty} b_i \varepsilon_{t-i}\right)\;\left(\sum_{h=-\infty}^{\infty} b_h \varepsilon_{t-\tau-h}\right)\right]$$

$$= \sigma^2 \sum_{i=-\infty}^{\infty} b_i b_{i-\tau}$$

(where $b_i \equiv 0$ if $i < 0$)

$$g(z) \;=\; \sigma^2\; B(z)\; B(z^{-1})$$

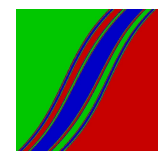


Approximating the Wold Representation, I: Finite-Ordered Autoregressions

$AR(p)$ process

(Stochastic difference equation)

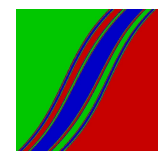
We now study $AR(1)$ (heavy detail)



Approximating the Wold Representation II: Fintite-Ordered Moving Average Processes

$MA(q)$ process
(Obvious truncation)

We now study $MA(1)$ (light detail)



Wiener-Kolmogorov Prediction

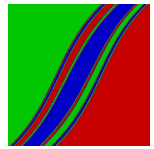
$$y_t = \varepsilon_t + b_1 \varepsilon_{t-1} + ...$$

$$y_{T+h} = \varepsilon_{T+h} + b_1 \varepsilon_{T+h-1} + ... + b_h \varepsilon_T + b_{h+1} \varepsilon_{T-1} + ...$$

Project on $\Omega_T = \{\varepsilon_T, \varepsilon_{T-1}, ...\}$ to get:

$$y_{T+h,T} = b_h \varepsilon_T + b_{h+1} \varepsilon_{T-1} + ...$$

Note that the projection is on the *infinite* past



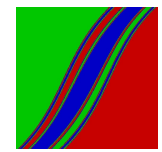
Wiener-Kolmogorov Prediction Error

$$e_{T+h,T} = y_{T+h} - y_{T+h,T} = \sum_{i=0}^{h-1} b_i \varepsilon_{T+h-i}$$

(An $MA(h-1)$ process!)

$$E(e_{T+h,T}) = 0$$

$$var(e_{T+h,T}) = \sigma^2 \sum_{i=0}^{h-1} b_i^2$$



Wold's Chain Rule for Autoregressions

Consider an AR(1) process:

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

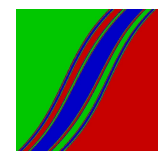
History:

$$\{y_t\}_{t=1}^{T}$$

Immediately,

$$y_{T+1,T} = \phi y_T$$
$$y_{T+2,T} = \phi y_{T+1,T} = \phi^2 y_T$$
$$\vdots$$
$$y_{T+h,T} = \phi y_{T+h-1,T} = \phi^h y_T$$

Extension to $AR(p)$ and $AR(\infty)$ is immediate.



Multivariate (Bivariate)

Define $y_t = (y_{1t},\ y_{2t})'$ and $\mu_t = (\mu_{1t},\ \mu_{2t})'$

y_t is covariance stationary if:

$$E(y_t) = \mu_t = \mu,\ \forall\ t$$

where $\mu = (\mu_1,\ \mu_2)'$ is a vector of constants

"mean does not depend on time"

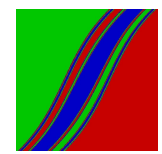
$$\Gamma_{y_1 y_2}(t,\ \tau) = E(y_t - \mu_t)(y_{t+\tau} - \mu_{t+\tau})'$$

$$= \begin{pmatrix} \gamma_{11}(\tau) & \gamma_{12}(\tau) \\ \gamma_{21}(\tau) & \gamma_{22}(\tau) \end{pmatrix},\ \tau = 0,\ 1,\ 2,\ ...$$

"autocovariance depends only on displacement, not on time"

$$var(y_1) < \infty,\ var(y_2) < \infty$$

"finite variance"



Cross Covariances

Cross covariances not symmetric in τ:

$$\gamma_{12}(\tau) \neq \gamma_{12}(-\tau)$$

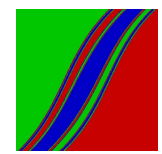
Instead:

$$\gamma_{12}(\tau) = \gamma_{21}(-\tau)$$

$$\Gamma_{y_1 y_2}(\tau) = \Gamma'_{y_1 y_2}(-\tau), \quad \tau = 0,\ 1,\ 2,\ \ldots$$

Covariance-generating function:

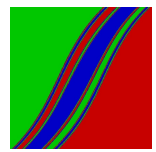
$$G_{y_1 y_2}(z) = \sum_{\tau=-\infty}^{\infty} \Gamma_{y_1 y_2}(\tau) z^{\tau}$$



Cross Correlations

$$R_{y_1 y_2}(\tau) = D^{-1}_{y_1 y_2}\ \Gamma_{y_1 y_2}(\tau)\ D^{-1}_{y_1 y_2},\ \tau = 0,\ 1,\ ,2,\ ...$$

$$D = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$$



The Multivariate General Linear Process

$$y_t = B(L)\,\varepsilon_t = \sum_{i=0}^{\infty} B_i\,\varepsilon_{t-i}$$

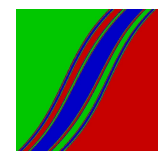
$$E(\varepsilon_t\,\varepsilon_s') = \begin{cases} \Sigma & \text{if } t = s \\ 0 & \text{otherwise} \end{cases}$$

$$B_0 = I$$

$$\sum_{i=0}^{\infty} \|B_i\|^2 < \infty$$

Bivariate case:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

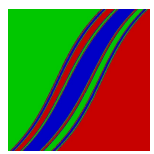


Autocovariance Structure

$$\Gamma_{y_1 y_2}(\tau) = \sum_{i=-\infty}^{\infty} B_i \; \Sigma \; B'_{i-\tau}$$

(where $B_i \equiv 0$ if $i < 0$)

$$G_y(z) = B(z) \; \Sigma \; B'(z^{-1})$$



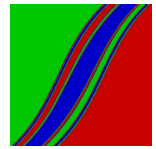
Wiener-Kolmogorov Prediction

$$y_t = \varepsilon_t + B_1\varepsilon_{t-1} + B_2\varepsilon_{t-2} + ...$$

$$y_{T+h} = \varepsilon_{T+h} + B_1\varepsilon_{T+h-1} + B_2\varepsilon_{T+h-2} + ...$$

Project on $\Omega_t = \{\varepsilon_T, \varepsilon_{T-1}, ...\}$ to get:

$$y_{t+h,T} = B_h\varepsilon_T + B_{h+1}\varepsilon_{T-1} + ...$$

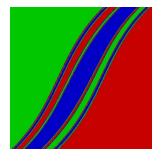


Wiener-Kolmogorov Prediction Error

$$\varepsilon_{T+h,T} = y_{T+h} - y_{T+h,T} = \sum_{i=0}^{h-1} B_i \varepsilon_{T+h-i}$$

$$E[\varepsilon_{T+h,T}] = 0$$

$$E[\varepsilon_{T+h,T}\varepsilon'_{T+h,T}] = \sum_{i=0}^{h-1} B_i \Sigma B_i'$$



Vector Autoregressions (VAR's)

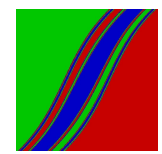
N-variable VAR of order p:

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

$$\varepsilon_t \sim WN(0, \Sigma)$$

where:

$$\Phi(L) = I - \Phi_1 L - ... - \Phi_p L^p$$

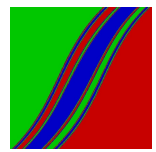


Bivariate VAR(1) in "Long Form"

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim WN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right)$$

– Two sources of cross-variable interaction.



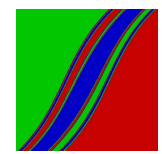
MA Representation of a VAR

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

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

where:

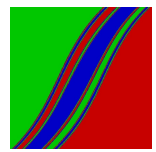
$$\Theta(L) = I + \Theta_1 L + \Theta_2 L^2 + ...$$



MA Representation of Bivariate VAR(1) in Long Form

$$\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} L \right) \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \begin{pmatrix} \theta^1_{11} & \theta^1_{12} \\ \theta^1_{21} & \theta^1_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix} + \ldots$$



Understanding VAR's: Granger-Sims Causality

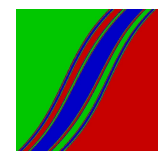
Is the history of y_j useful for predicting y_i,
over and above the history of y_i?

– Granger non-causality corresponds to exclusion restrictions

– In the simple 2-Variable VAR(1) example,

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

y_2 does not Granger cause y_1 iff $\phi_{12} = 0$



Understanding VAR's: Impulse Response Functions (IRF's)

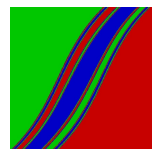
$$(I - \Phi_1 L - ... - \Phi_p L^p) y_t = \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \Sigma)$$

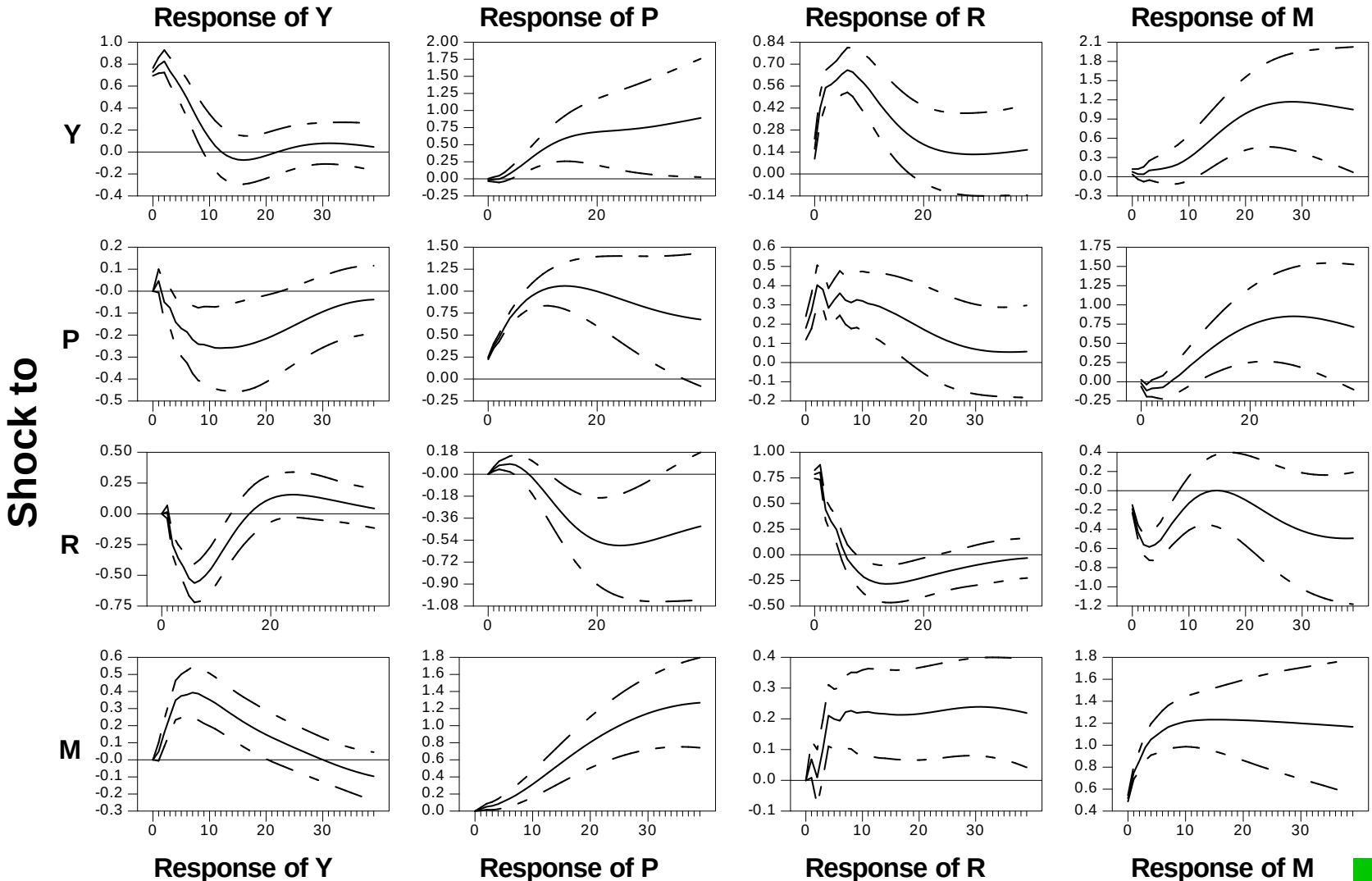
The impulse-response question:
How is y_{it} dynamically affected by a shock to y_{jt} (alone)?

($N \times N$ matrix of IRF *graphs* (over steps ahead))

Problem:
Σ generally not diagonal, so how to shock j alone?



Graphic: IRF Matrix for 4-Variable U.S. Macro VAR



Understanding VAR's: Variance Decompositions (VD's)

$$(I - \Phi_1 L - ... - \Phi_p L^p) y_t = \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \Sigma)$$

The variance decomposition question:
How much of the h-step ahead (optimal) prediction-error variance of y_i is due to shocks to variable j?

($N \times N$ matrix of VD *graphs* (over h).
Or pick an h and examine the $N \times N$ matrix of VD numbers.)

Problem:
Σ generally not diagonal, which makes things tricky, as the variance of a sum of innovations is not the sum of the variances in that case.

Orthogonalizing VAR's by Cholesky Factorization (A Classic Identification Scheme)

Original:

$$(I - \Phi_1 L - ... - \Phi_p L^p) y_t = \varepsilon_t, \quad \varepsilon_t \sim WN(0, \Sigma)$$

Equivalently:

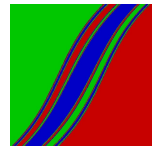
$$(I - \Phi_1 L - ... - \Phi_p L^p) y_t = P v_t, \quad v_t \sim WN(0, I)$$

or

$$(P^{-1} - [P^{-1}\Phi_1] L - ... - [P^{-1}\Phi_p] L^p) y_t = v_t, \quad v_t \sim WN(0, I)$$

where $\Sigma = PP'$, for lower-triangular P
(Cholesky factorization)

Now we can proceed with IRF's and VD's.



IRF's and VD's from the Orthogonalized VAR

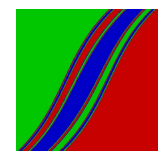
IRF comes from the
orthogonalized moving-average representation:

$$y_t = (I + \Theta_1 L + \Theta_2 L^2 + ...)\ P\, v_t$$

$$= (P + \Theta_1 P\ L + \Theta_2 P\ L^2 + ...)\ v_t$$

$\mathrm{IRF_{ij}}$ is $\{P_{ij},\ (\Theta_1 P)_{ij},\ (\Theta_2 P)_{ij}, ...\}$

$\mathrm{VD_{ij}}$ comes similarly from the
orthogonalized moving-average representation.



Bivariate IRF Example (e.g., IRF_{12})

$$y_t = Pv_t + \Theta_1 Pv_{t-1} + \Theta_2 Pv_{t-2} + ...$$

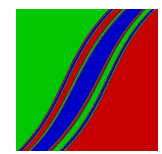
$$v_t \sim \ WN(0, I)$$

$$y_t = C_0 v_t + C_1 v_{t-1} + C_2 v_{t-2} + ...$$

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_{11}^0 & c_{12}^0 \\ c_{21}^0 & c_{22}^0 \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} c_{11}^1 & c_{12}^1 \\ c_{21}^1 & c_{22}^1 \end{pmatrix} \begin{pmatrix} v_{1t-1} \\ v_{2t-1} \end{pmatrix} + ...$$

$$\mathrm{IRF}_{12} = \mathrm{c}_{12}^0, \mathrm{c}_{12}^1, \mathrm{c}_{12}^2, ...$$

$$(\mathrm{Q}: \ \text{What is } c_{12}^0?)$$



Bivariate VD Example (e.g., VD_{12} for $h = 2$)

$$\epsilon_{t+2,t} = C_0 v_{t+2} + C_1 v_{t+1}$$

$$v_t \sim WN(0, I)$$

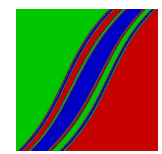
$$\begin{pmatrix} \epsilon^1_{t+2,t} \\ \epsilon^2_{t+2,t} \end{pmatrix} = \begin{pmatrix} c^0_{11} & c^0_{12} \\ c^0_{21} & c^0_{22} \end{pmatrix} \begin{pmatrix} v_{1t+2} \\ v_{2t+2} \end{pmatrix} + \begin{pmatrix} c^1_{11} & c^1_{12} \\ c^1_{21} & c^1_{22} \end{pmatrix} \begin{pmatrix} v_{1t+1} \\ v_{2t+1} \end{pmatrix}$$

$$\epsilon^1_{t+2,t} = c^0_{11} v_{1t+2} + c^0_{12} v_{2t+2} + c^1_{11} v_{1t+1} + c^1_{12} v_{2t+1}$$

$$var(\epsilon^1_{t+2,t}) = (c^0_{11})^2 + (c^0_{12})^2 + (c^1_{11})^2 + (c^1_{12})^2$$

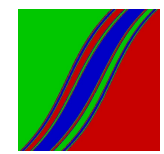
Part coming from v_2: $(c^0_{12})^2 + (c^1_{12})^2$

$$VD_{12}(2) = \frac{(c^0_{12})^2 + (c^1_{12})^2}{(c^0_{11})^2 + (c^0_{12})^2 + (c^1_{11})^2 + (c^1_{12})^2}$$



Frequency Domain

“Spectral Analysis”



Remember...

$$z = a + bi \quad (rectangular)$$

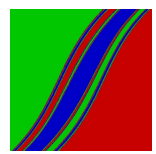
$$z = re^{i\omega} = r(\cos\omega + i\sin\omega) \quad (polar)$$

$$cos(\omega) = \frac{e^{i\omega} + e^{-i\omega}}{2}$$

$$sin(\omega) = \frac{e^{i\omega} - e^{-i\omega}}{2}$$

$$P = \frac{2\pi}{\omega}$$

$$z\bar{z} = |z|^2$$



Recall the General Linear Process

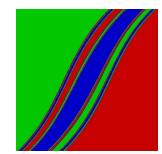
$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

Autocovariance generating function:

$$g(z) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\ z^{\tau}$$

$$= \sigma^2 B(z) B(z^{-1})$$

$\gamma(\tau)$ and $g(z)$ are a z-transform pair



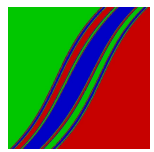
Spectrum

Evaluate g(z) on the unit circle, $z = e^{-i\omega}$:

$$g(e^{-i\omega}) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\, e^{-i\omega\tau}, \quad -\pi < \omega < \pi$$

$$= \sigma^2 B(e^{i\omega})\ B(e^{-i\omega})$$

$$= \sigma^2 \mid B(e^{i\omega}) \mid^2$$



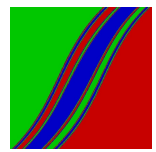
Spectrum

Trigonometric form:

$$g(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-i\omega\tau}$$

$$= \gamma(0) + \sum_{\tau=1}^{\infty} \gamma(\tau) \left(e^{i\omega\tau} + e^{-i\omega\tau} \right)$$

$$= \gamma(0) + 2 \sum_{\tau=1}^{\infty} \gamma(\tau) \ \cos(\omega\tau)$$



Spectral Density Function

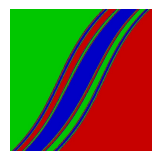
$$f(\omega) = \frac{1}{2\pi} g(\omega)$$

$$f(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-i\omega\tau} \quad (-\pi < \omega < \pi)$$

$$= \frac{1}{2\pi}\gamma(0) + \frac{1}{\pi} \sum_{\tau=1}^{\infty} \gamma(\tau) \cos(\omega\tau)$$

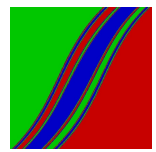
$$= \frac{\sigma^2}{2\pi} B\left(e^{i\omega}\right) B\left(e^{-i\omega}\right)$$

$$= \frac{\sigma^2}{2\pi} \mid B\left(e^{i\omega}\right) \mid^2$$



Properties of Spectrum and Spectral Density

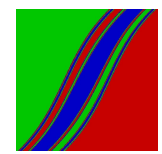
1. symmetric around $\omega = 0$
2. real-valued
3. 2π-periodic
4. nonnegative



A Fourier Transform Pair

$$g(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-i\omega\tau}$$

$$\gamma(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\omega) e^{i\omega\tau} d\omega$$



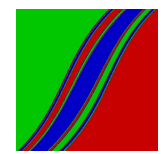
A Variance Decomposition by Frequency

$$\gamma(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\omega) e^{i\omega\tau} d\omega$$

$$= \int_{-\pi}^{\pi} f(\omega) e^{i\omega\tau} d\omega$$

Hence

$$\gamma(0) = \int_{-\pi}^{\pi} f(\omega) d\omega$$



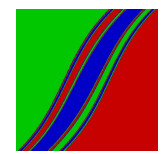
White Noise Spectral Density

$$y_t = \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

$$f(\omega) = \frac{\sigma^2}{2\pi} B\left(e^{i\omega}\right) B\left(e^{-i\omega}\right)$$

$$= \frac{\sigma^2}{2\pi}$$



AR(1) Spectral Density

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

$$\varepsilon_t \sim WN(0, \sigma^2)$$

$$f(\omega) = \frac{\sigma^2}{2\pi} B(e^{i\omega})B(e^{-i\omega})$$

$$= \frac{\sigma^2}{2\pi} \frac{1}{(1-\phi e^{i\omega})(1-\phi e^{-i\omega})}$$

$$= \frac{\sigma^2}{2\pi} \frac{1}{1-2\phi\cos(\omega)+\phi^2}$$

How does shape depend on ϕ? Where are the peaks?

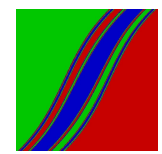
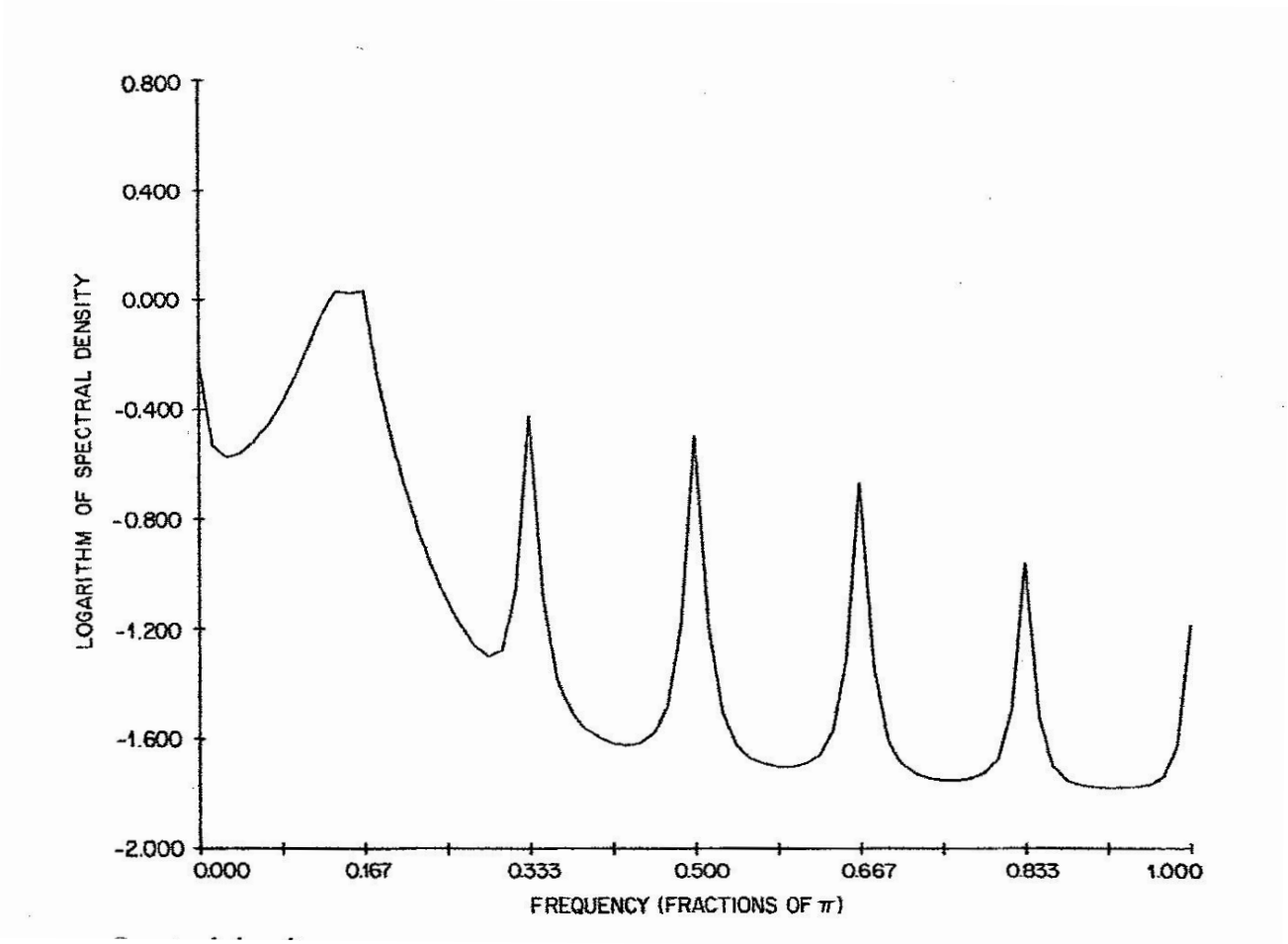
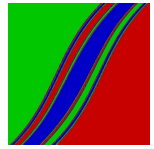


Figure: Granger's Typical Spectral Shape of an Economic Variable





Robust Variance Estimation

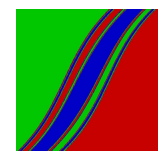
$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

$$var(\bar{y}) = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma(t-s)$$

("Add row sums")

$$= \frac{1}{T}\sum_{\tau=-(T-1)}^{T-1}\left(1-\frac{|\tau|}{T}\right)\gamma(\tau)$$

("Add diagonal sums," using change of variable $\tau = t - s$)

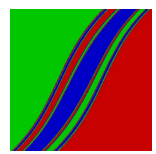


Robust Variance Estimation, Continued

$$\Rightarrow \quad \sqrt{T}(\bar{y} - \mu) \sim \left(0, \sum_{\tau=-(T-1)}^{T-1} \left(1 - \frac{|\tau|}{T}\right) \gamma(\tau)\right)$$

$$\sqrt{T}(\bar{y} - \mu) \overset{d}{\rightarrow} N\left(0, \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\right)$$

$$\sqrt{T}(\bar{y} - \mu) \overset{d}{\rightarrow} N\left(0, g(0)\right)$$



Estimation: Sample Spectral Density

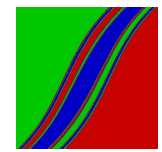
$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \hat{\gamma}(\tau) e^{-i\omega\tau},$$

where

$$\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})(y_{t-\tau} - \bar{y})$$

(Use frequencies $\omega_j = \frac{2\pi j}{T}$, $j = 0, 1, 2, ..., \frac{T}{2}$)

– Inconsistent, unfortunately



Historical and Computational Note: The FFT and Periodogram

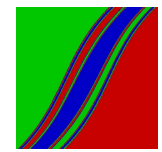
$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \hat{\gamma}(\tau) e^{-i\omega\tau}$$

$$= \left(\frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} y_t e^{-i\omega t} \right) \left(\frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} y_t e^{i\omega t} \right)$$

$$(\mathrm{FFT}) \cdot (\overline{\mathrm{FFT}})$$

$$= \frac{1}{4\pi} I(\omega) \quad (I \text{ is the "periodogram"})$$

– Again: Inconsistent, unfortunately



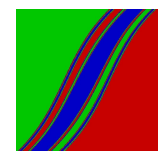
Properties of the Sample Spectral Density

Under conditions:

- $\hat{f}(\omega_j)$ asymptotically unbiased
- But $var(\hat{f}(\omega_j))$ does not converge to 0 (d.f. don't accumulate)
- Hence $\hat{f}(\omega_j)$ is inconsistent. What's true is that:

$$\frac{2\hat{f}(\omega_j)}{f(\omega_j)} \xrightarrow{d} \chi^2_2,$$

where the χ^2_2 random variables are uncorrelated across frequencies $\omega_j = \frac{2\pi j}{T}$, $j = 0, 1, 2, ..., \frac{T}{2}$



Consistent ("Lag Window") Spectral Estimation

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \hat{\gamma}(\tau) e^{-i\omega\tau}$$

$$f^*(\omega) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \lambda(\tau)\hat{\gamma}(\tau) e^{-i\omega\tau}$$

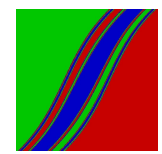
(Weight the sample autocovariances.)

Common "lag windows" with "truncation lag" M_T:

$\lambda(\tau) = 1,\ |\tau| \leq M_T$ and 0 otherwise (rectangular, or uniform)

$\lambda(\tau) = 1 - \frac{|\tau|}{M_T}, \tau \leq M_T$ and 0 otherwise
(triangular, or Bartlett, or Newey-West)

Consistency: $M_T \to \infty$ and $\frac{M_T}{T} \to 0$ as $T \to \infty$



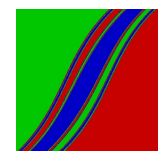
Consistent ("Autoregressive") Spectral Estimation

"Model-based estimation"

Fit $AR(p_T)$ model (using AIC, say)

Calculate spectrum of the fitted model at the fitted parameters

Consistency: $p_T \to \infty$ and $\frac{p_T}{T} \to 0$ as $T \to \infty$



Multivariate Frequency Domain

Covariance-generating function:

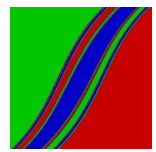
$$G_{y_1 y_2}(z) = \sum_{\tau=-\infty}^{\infty} \Gamma_{y_1 y_2}(\tau) z^{\tau}$$

Spectral density function:

$$F_{y_1 y_2}(\omega) = \frac{1}{2\pi} G_{y_1 y_2}(e^{-i\omega})$$

$$= \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \Gamma_{y_1 y_2}(\tau)\ e^{-i\omega\tau}, \quad -\pi < \omega < \pi$$

(Complex-valued)



Consistent Multivariate Spectral Estimation

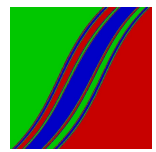
Spectral density matrix:

$$F_{y_1 y_2}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \Gamma_{y_1 y_2}(\tau) e^{-i\omega\tau}, \quad -\pi < \omega < \pi$$

Consistent (lag window) estimator:

$$F^*_{y_1 y_2}(\omega) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{(T-1)} \lambda(\tau) \hat{\Gamma}_{y_1 y_2}(\tau)\ e^{-i\omega\tau}, \quad -\pi < \omega < \pi$$

Or do autoregressive (VAR) spectral estimation.

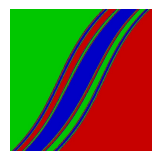


Co-Spectrum and Quadrature Spectrum

$$F_{y_1 y_2}(\omega) = C_{y_1 y_2}(\omega) + iQ_{y_1 y_2}(\omega)$$

$$C_{y_1 y_2}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \Gamma_{y_1 y_2}(\tau) \cos(\omega\tau)$$

$$Q_{y_1 y_2}(\omega) = \frac{-1}{2\pi} \sum_{\tau=-\infty}^{\infty} \Gamma_{y_1 y_2}(\tau) \sin(\omega\tau)$$



Cross Spectrum

$$f_{y_1y_2}(\omega) = ga_{y_1y_2}(\omega)exp(i\ ph_{y_1y_2}(\omega)) \text{ (generic cross spectrum)}$$

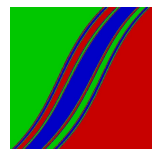
$$ga_{y_1y_2}(\omega) = [c^2_{y_1y_2}(\omega) + q^2_{y_1y_2}(\omega)]^{\frac{1}{2}} \text{ (gain)}$$

$$ph_{y_1y_2}(\omega) = arctan\left(\frac{q_{y_1y_2}(\omega)}{c_{y_1y_2}(\omega)}\right) \text{ (phase shift in radians)}$$

$$\text{(Phase shift in time units is } \frac{ph(\omega)}{\omega}\text{)}$$

$$coh_{y_1y_2}(\omega) = \frac{|f_{y_1y_2}(\omega)|^2}{f_{y_1y_1}(\omega)f_{y_2y_2}(\omega)} \quad \text{(coherence)}$$

Squared correlation decomposed by frequency



Useful Spectral Results for Filter Analysis

If $y_{1t} = B(L)\, y_{2t}$, then:

$$f_{y_1 y_1}(\omega) = |B(e^{-i\omega})|^2 f_{y_2 y_2}(\omega)$$
$$f_{y_1 y_2}(\omega) = B(e^{-i\omega})\, f_{y_2 y_2}(\omega)$$

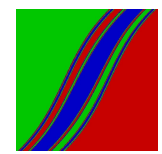
$B(e^{-i\omega})$ is the filter's *frequency response function*

If $y_{1t} = A(L)B(L)y_{2t}$, then:

$$f_{y_1 y_1}(\omega) = |A(e^{-i\omega})|^2 |B(e^{-i\omega})|^2 f_{y_2 y_2}(\omega)$$
$$f_{y_1 y_2}(\omega) = A(e^{-i\omega})B(e^{-i\omega}) f_{y_2 y_2}(\omega)$$

If $y = \sum_{i=1}^{N} y_i$, and the y_i are independent, then:

$$f_y(\omega) = \sum_{i=1}^{N} f_{y_i}(\omega)$$



Nuances...

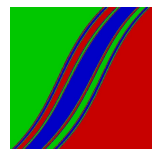
Note that

$$B(e^{-i\omega}) = \frac{f_{y_1y_2}(\omega)}{f_{y_2y_2}(\omega)}$$

$$\implies B(e^{-i\omega}) = \frac{ga_{y_1y_2}(\omega)e^{i\ ph_{y_1y_2}(\omega)}}{f_{y_2y_2}(\omega)} = \left(\frac{ga_{y_1y_2}(\omega)}{f_{y_2y_2}(\omega)}\right) e^{i\ ph_{y_1y_2}(\omega)}$$

Gains of $f_{y_1y_2}(\omega)$ and $B(e^{-i\omega})$ are closely related.

Phases are the same.



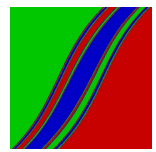
Filter Analysis Example I:
A Simple (but Very Important) High-Pass Filter

$$y_{1t} = (1 - L)y_{2t} = y_{2t} - y_{2,t-1}$$

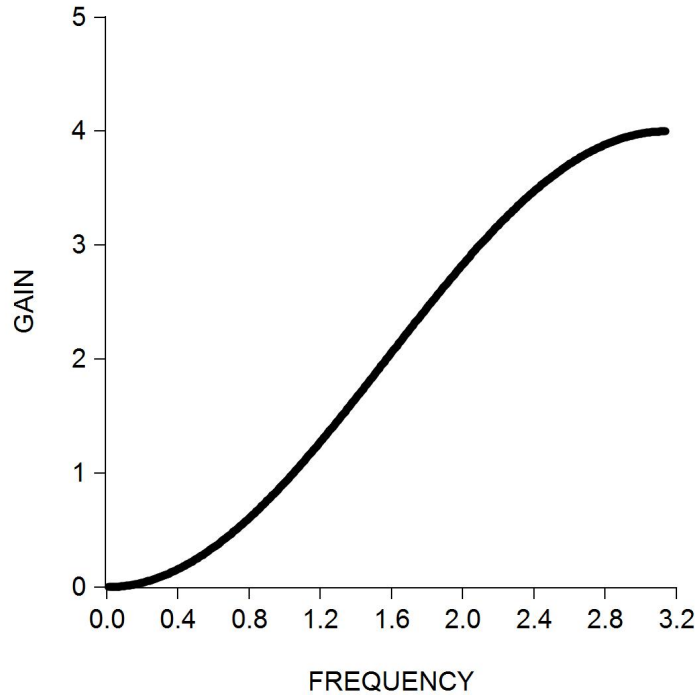
$$B(L) = (1 - L) \implies B(e^{-i\omega}) = 1 - e^{-i\omega}$$

Hence the filter gain is:

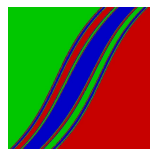
$$\left|B(e^{-i\omega})\right| = \left|1 - e^{-i\omega}\right| = 2(1 - cos(\omega))$$



Gain of the First-Difference Filter, $B(L) = 1 - L$



How would the gain look if the filter were $B(L) = 1 + L$ rather than $B(L) = 1 - L$?



Filter Analysis Example II

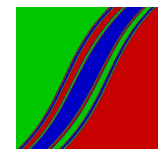
$$y_{2t} = .9y_{2,t-1} + \eta_t$$

$$\eta_t \sim WN(0, 1)$$

$$y_{1t} = .5y_{2,t-1} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, 1)$$

where η_t and ε_t are orthogonal at all leads and lags



Spectral Density of y_2

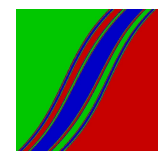
$$y_{2t} = \frac{1}{1 - .9L}\eta_t$$

$$\implies f_{y_2 y_2}(\omega) = \frac{1}{2\pi}\left(\frac{1}{1 - .9e^{-i\omega}}\right)\left(\frac{1}{1 - .9e^{i\omega}}\right)$$

$$= \frac{1}{2\pi}\,\frac{1}{1 - 2(.9)\cos(\omega) + (.9)^2}$$

$$= \frac{1}{11.37 - 11.30\cos(\omega)}$$

Shape?



Spectral Density of y_1

$$y_{1t} = 0.5Ly_{2t} + \varepsilon_t$$

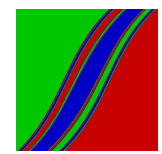
(So $B(L) = .5L$ and $B(e^{-i\omega}) = 0.5e^{-i\omega}$)

$$\implies f_{y_1y_1}(\omega) = \mid 0.5e^{-i\omega} \mid^2 f_{y_2y_2}(\omega) + \frac{1}{2\pi}$$

$$= 0.25 f_{y_2y_2}(\omega) + \frac{1}{2\pi}$$

$$= \frac{0.25}{11.37 - 11.30\cos(\omega)} + \frac{1}{2\pi}$$

Shape?



Cross Spectrum Gain and Phase

$$B(L) = .5L$$

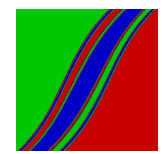
$$B(e^{-i\omega}) = .5e^{-i\omega}$$

$$\begin{aligned} f_{y_1y_2}(\omega) &= B(e^{-i\omega})\ f_{y_2y_2}(\omega) \\ &= .5e^{-i\omega}\ f_{y_2y_2}(\omega) \\ &= (.5f_{y_2y_2}(\omega))\, e^{-i\omega} \end{aligned}$$

$$g_{y_1y_2}(\omega) = .5f_{y_2y_2}(\omega) = \frac{.5}{11.37 - 11.30\cos(\omega)}$$

$$Ph_{y_1y_2}(\omega) = -\omega$$

(In time units, $Ph_{y_1y_2}(\omega) = -1$, so y_1 leads y_2 by -1)



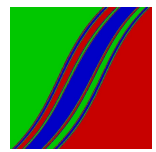
Cross Spectrum Coherence

$$Coh_{y_1y_2}(\omega) = \frac{\mid f_{y_1Y_2}(\omega) \mid^2}{f_{y_2y_2}(\omega) f_{y_1y_1}(\omega)} = \frac{.25 f^2_{y_2y_2}(\omega)}{f_{y_2y_2}(\omega) f_{y_1y_1}(\omega)} = \frac{.25 f_{y_2y_2}(\omega)}{f_{y_1y_1}(\omega)}$$

$$= \frac{.25 \ \frac{1}{2\pi} \ \frac{1}{1-2(.9)\cos(\omega)+.9^2}}{.25 \ \frac{1}{2\pi} \ \frac{1}{1-2(.9)\cos(\omega)+.9^2} + \frac{1}{2\pi}}$$

$$= \frac{1}{8.24 + 7.20\cos(\omega)}$$

Shape?



Filter Analysis Example III: Kuznets' Filters

Low-frequency fluctuations in aggregate real output growth.
Low-frequency "Kuznets cycle": 20-year period

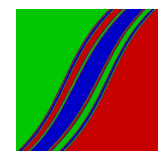
Filter 1 (moving average):

$$y_{1t} = \frac{1}{5}\sum_{j=-2}^{2} y_{2,t-j}$$

$$\Longrightarrow \; B_1(e^{-i\omega}) = \frac{1}{5}\sum_{j=-2}^{2} e^{-i\omega j} = \frac{sin(5\omega/2)}{5sin(\omega/2)}$$

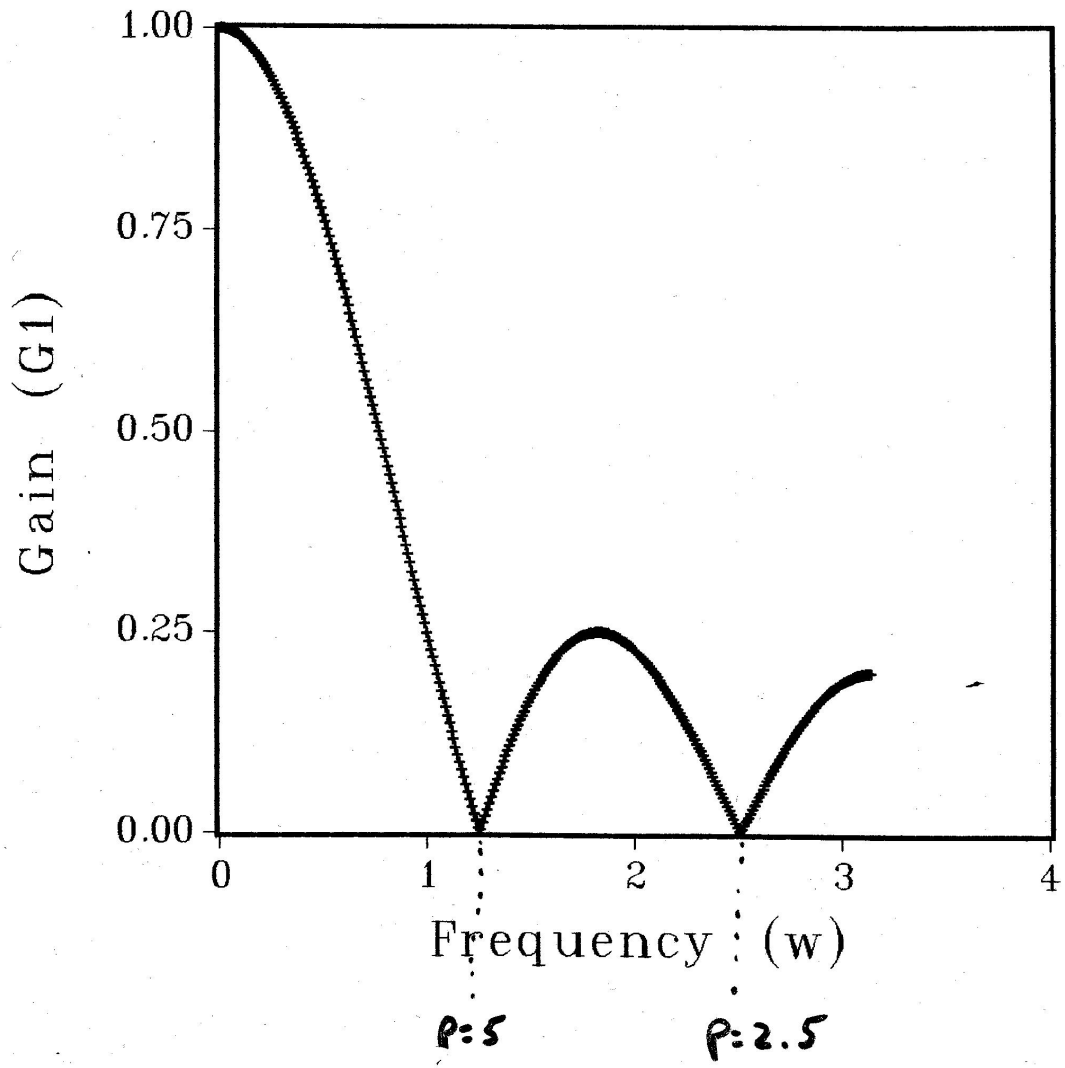
Hence the filter gain is:

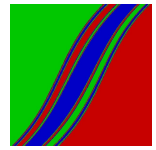
$$\left|B_1(e^{-i\omega})\right| = \left|\frac{sin(5\omega/2)}{5sin(\omega/2)}\right|$$



Kuznets' Filters, Continued

Figure: Gain of Kuznets' Filter 1





Kuznets' Filters, Continued

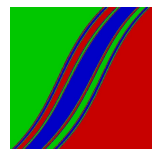
Filter 2 (fancy difference):

$$y_{3t} = y_{2,t+5} - y_{2,t-5}$$

$$\implies B_2(e^{-i\omega}) = e^{i5\omega} - e^{-i5\omega} = 2sin(5\omega)$$

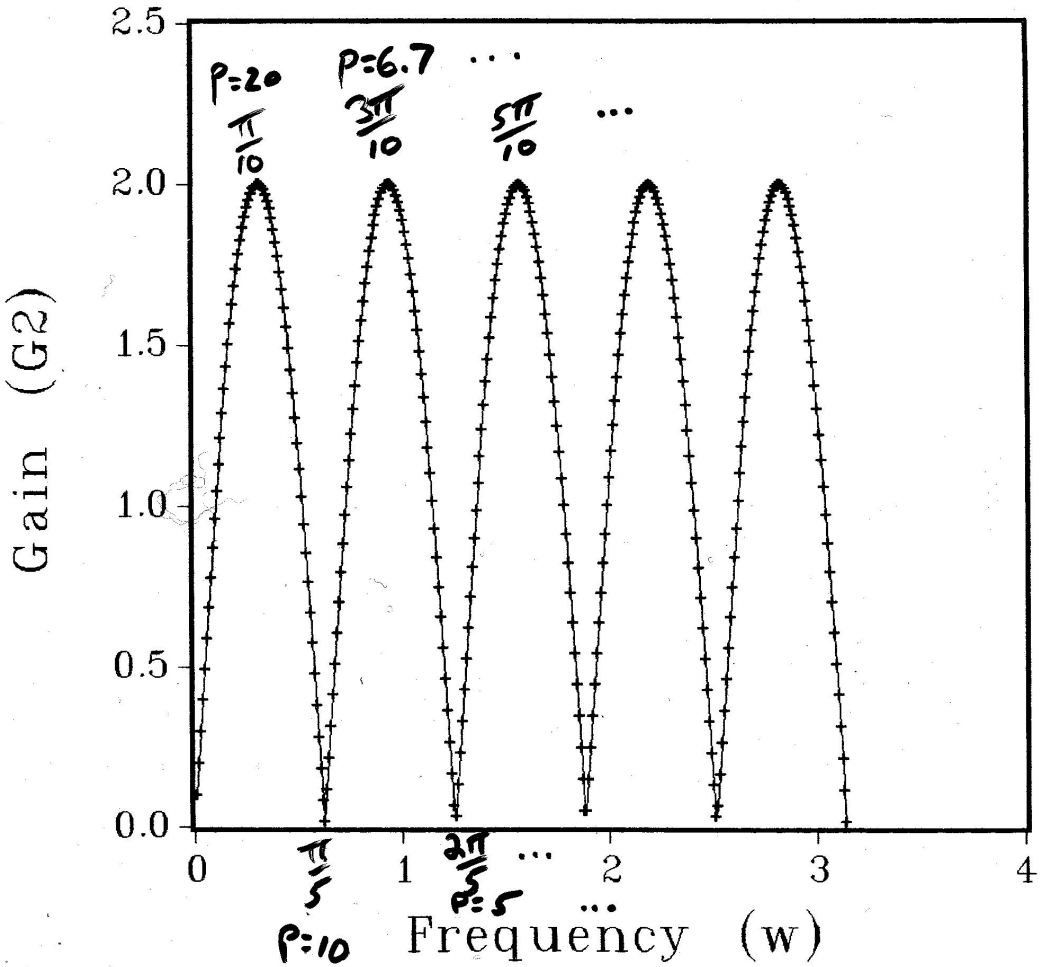
Hence the filter gain is:

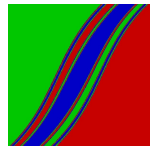
$$\left|B_2(e^{-i\omega})\right| = |2sin(5\omega)|$$



Kuznets' Filters, Continued

Figure: Gain of Kuznets' Filter 2

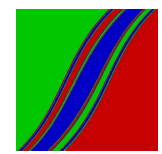




Kuznets' Filters, Continued

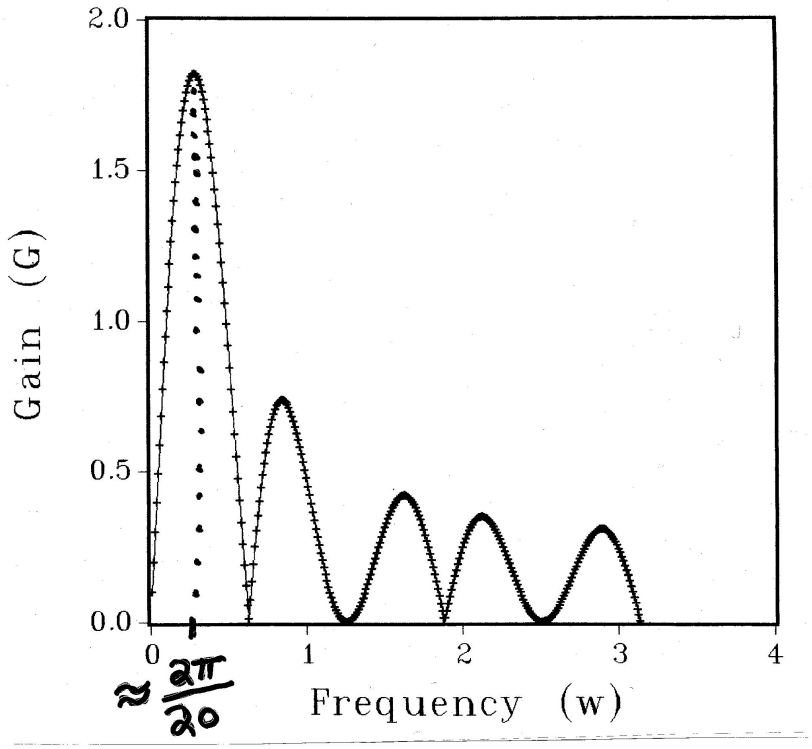
Composite gain:

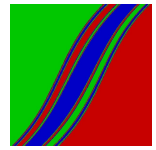
$$\left|B_1(e^{-i\omega})B_2(e^{-i\omega})\right| = \left|\frac{sin(5\omega/2)}{5sin(\omega/2)}\right| |2sin(5\omega)|$$



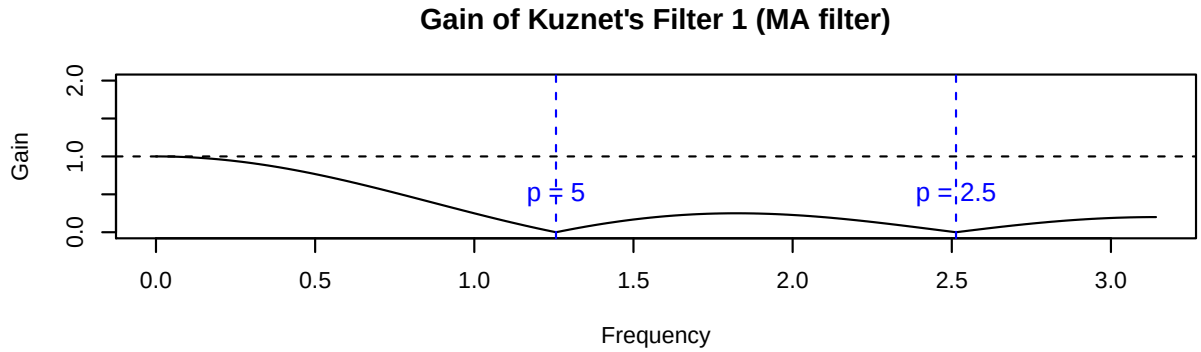
Kuznets' Filters, Continued

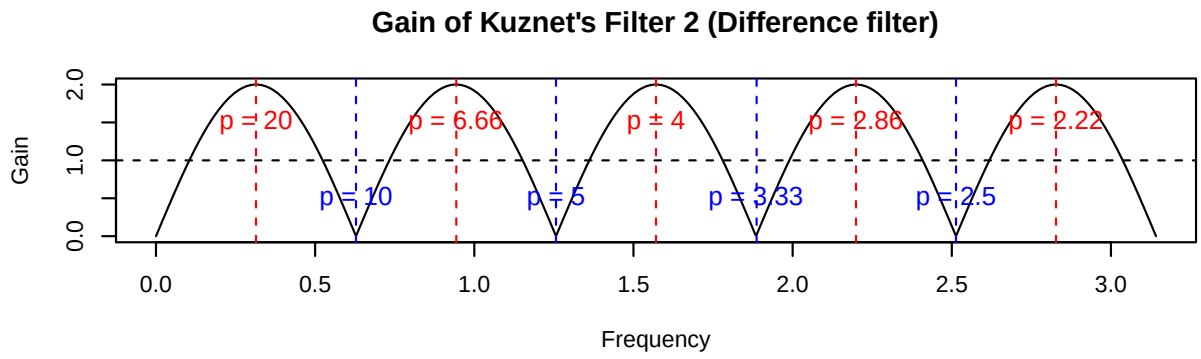
Figure: Composite Gain of Kuznets' two Filters

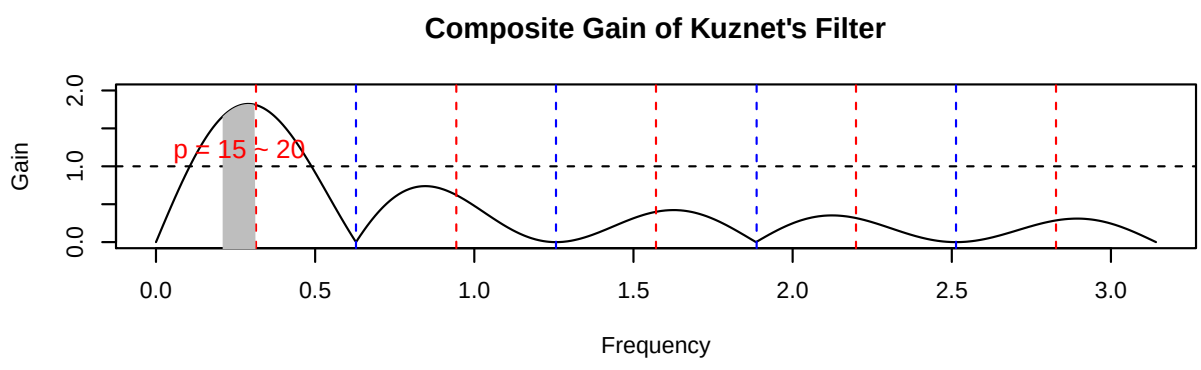


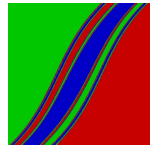


Gains All Together for Comparison









So The Kuznets Procedure may Induce a Spurious "Long-Wave" Cycle

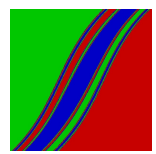
Let's apply the Kuznets filter to WN:

$$\varepsilon_t \sim WN(0,1)$$

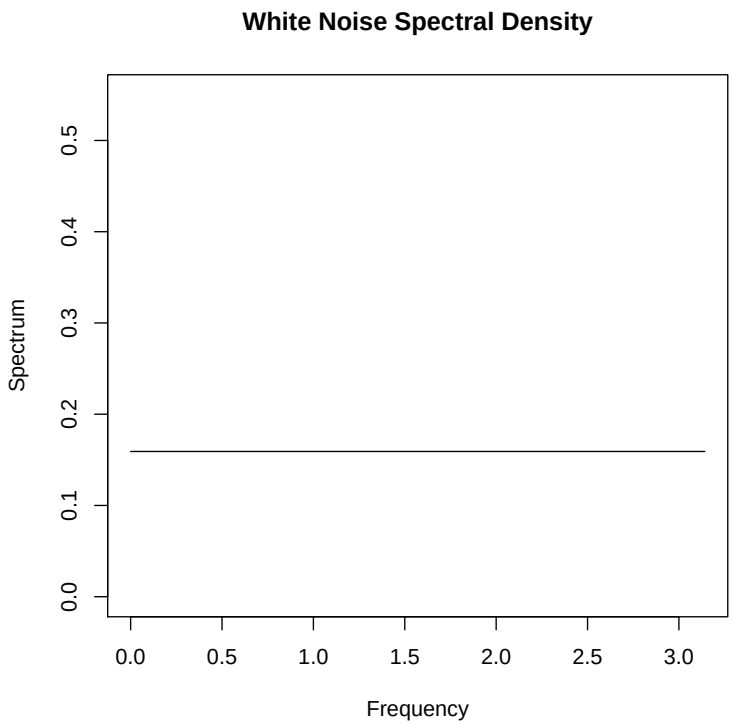
$$f_\varepsilon(w) = \frac{1}{2\pi}$$

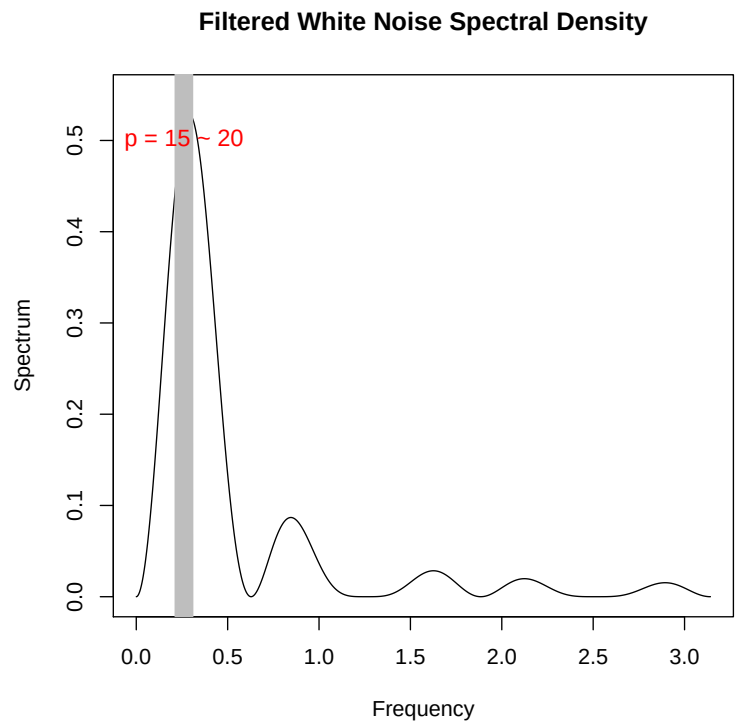
$$\tilde{\varepsilon}_t = B_{Kuznets}(L)\varepsilon_t,$$

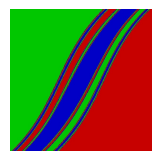
$$f_{\tilde{\varepsilon}_t}(w) = |B_{Kuznets}(e^{iw})|^2 f_\varepsilon(w)$$



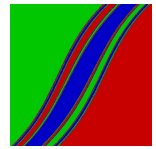
Compare the Two Spectra!



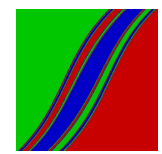




Markovian Structure, State Space, and the Kalman Filter



Part I: Markov Processes



Discrete-State, Discrete-Time Stochastic Process

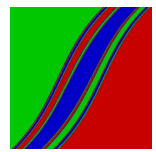
$$\{y_t\}, t = 0, 1, 2, \ldots$$

Possible values ("states") of y_t: $1, 2, 3, \ldots$

First-order homogeneous Markov process:

$$Prob(y_{t+1} = j | y_t = i, y_{t-1} = i_{t-1}, \ldots, y_0 = i_0)$$

$$= Prob(y_{t+1} = j | y_t = i) = p_{ij}$$

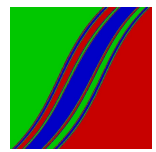


Transition Probability Matrix P

1-step transition probabilities:

$$P \equiv \begin{array}{c} \\ [time\ t] \end{array} \begin{array}{c} [time\,(t+1)] \\ \begin{pmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \end{pmatrix} \end{array}$$

$$p_{ij} \geq 0, \quad \sum_{j=1}^{\infty} p_{ij} = 1$$



Chapman-Kolmogorov

m-step transition probabilities:

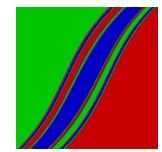
$$p_{ij}^{(m)} = Prob(y_{t+m} = j \mid \ y_t = i)$$

$$\text{Let } P^{(m)} \equiv \left(p_{ij}^{(m)}\right).$$

Chapman-Kolmogorov theorem:

$$P^{(m+n)} = P^{(m)} P^{(n)}$$

Corollary: $P^{(m)} = P^m$



Lots of Definitions...

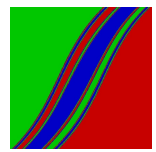
State j is accessible from state i if $p_{ij}^{(n)} > 0$, for some n.

Two states i and j communicate (or are in the same class)
if each is accessible from the other.
We write $i \leftrightarrow j$.

A Markov process is irreducible if there exists only one class
(i.e., all states communicate).

State i has period d if $p_{ii}^{(n)} = 0 \ \forall n$ such that $n/d \notin Z$,
and d is the greatest integer with that property.
(That is, a return to state i can only occur in multiples of d steps.)
A state with period 1 is called an aperiodic state.

A Markov process all of whose states are aperiodic
is called an aperiodic Markov process.



...And More Definitions

The first-transition probability is the probability that, starting in i, the first transition to j occurs after n transitions:

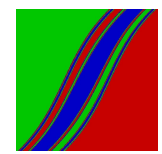
$$f_{ij}^{(n)} = Prob(y_n = j, y_k \neq j, k = 1, ..., (n-1) | y_0 = i)$$

Denote the eventual transition probability from i to j by f_{ij} $(= \sum_{n=1}^{\infty} f_{ij}^{(n)})$.

State j is recurrent if $f_{jj} = 1$ and transient otherwise.

Denote the expected number of transitions needed to return to recurrent state j by $\mu_{jj} (= \sum_{n=1}^{\infty} n f_{jj}^{(n)})$.

A recurrent state j is:
positive recurrent if $\mu_{jj} < \infty$
null recurrent if $\mu_{jj} = \infty$.

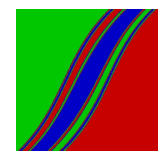


And One More Definition

The row vector π is called the stationary distribution for P if:

$$\pi P = \pi.$$

The stationary distribution is also called the steady-state distribution.



Theorem (Finally!)

Theorem: Consider an irreducible, aperiodic Markov process.

Then either:

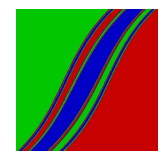
(1) All states are transient or all states are null recurrent

$p_{ij}^{(n)} \to 0$ as $n \to \infty$ $\forall i, j$. No stationary distribution.

or

(2) All states are positive recurrent.

$p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$ $\forall i, j$.
$\{\pi_j, j = 1, 2, 3, ...\}$ is the unique stationary distribution.
π is any row of $\lim_{n\to\infty} P^n$.



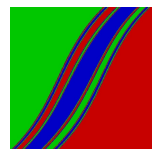
Example

Consider a Markov process with transition probability matrix:

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Call the states 1 and 2.

We will verify many of our claims,
and we will calculate the steady-state distribution.



Example, Continued

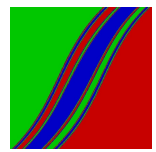
(a) Valid Transition Probability Matrix

$$p_{ij} \geq 0 \quad \forall i, j$$

$$\sum_{j=1}^{2} p_{1j} = 1, \quad \sum_{j=1}^{2} p_{2j} = 1$$

(b) Chapman-Kolmogorov Theorem (for $P^{(2)}$)

$$P^{(2)} = P \cdot P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



Example, Continued

(c) Communication and Reducibility

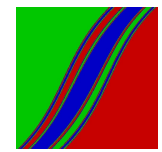
Clearly, $1 \leftrightarrow 2$, so P is irreducible.

(d) Periodicity

$$\text{State 1:} \quad d(1) = 2$$
$$\text{State 2:} \quad d(2) = 2$$

(e) First and Eventual Transition Probabilities

$$f_{12}^{(1)} = 1, \; f_{12}^{(n)} = 0 \; \forall \, n > 1 \quad \Rightarrow \quad f_{12} = 1$$
$$f_{21}^{(1)} = 1, \; f_{21}^{(n)} = 0 \; \forall \, n > 1 \quad \Rightarrow \quad f_{21} = 1$$



Example, Continued

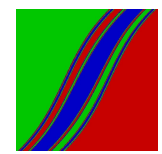
(f) Recurrence

Because $f_{21} = f_{12} = 1$, both states 1 and 2 are recurrent.

Moreover,

$$\mu_{11} = \sum_{n=1}^{\infty} n f_{11}^{(n)} = 2 < \infty \quad \text{(and similarly } \mu_{22} = 2 < \infty\text{)}$$

Hence states 1 and 2 are positive recurrent.



Example, Continued

(g) Stationary Distribution

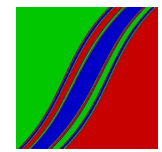
We will guess and verify.

Let $\pi_1 = .5$, $\pi_2 = .5$ and check $\pi P = \pi$:

$$(.5, .5) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (.5, .5).$$

Hence the stationary probabilities are 0.5 and 0.5.

Note that in this example we can *not* get the stationary probabilities by taking $\lim_{n \to \infty} P^n$. Why?



Illustrations/Variations/Extensions: Regime-Switching Models

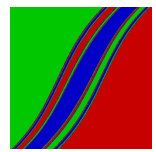
$$P = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix}$$

$$s_t \sim P$$

$$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim iid\ N(0, \sigma^2_{s_t})$$

"Markov switching," or "hidden Markov," model



Illustrations/Variations/Extensions: Heterogeneous Markov Processes

$$P_t = \begin{pmatrix} p_{11,t} & p_{12,t} & \cdots \\ p_{21,t} & p_{22,t} & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \end{pmatrix}.$$

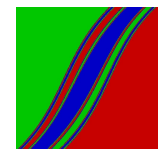
e.g., Regime switching with time-varying transition probabilities:

$$s_t \sim P_t$$

$$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim iid\ N(0, \sigma^2_{s_t})$$

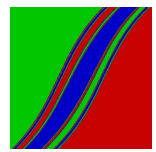
Business cycle duration dependence: $p_{ij,t} = g_{ij}(t)$
Credit migration over the cycle: $p_{ij,t} = g_{ij}(cycle_t)$
General covariates: $p_{ij,t} = g_{ij}(x_t)$



Illustrations/Variations/Extensions: Constructing Markov Processes with Useful Stationary Distributions

- Markov Chain Monte Carlo (e.g., Gibbs sampling)
  – Construct a Markov process from whose steady-state distribution we want to sample.

- Global Optimization (e.g., simulated annealing)
  – Construct a Markov process the support of whose steady-state distribution is the set of global optima of a function we want to maximize.

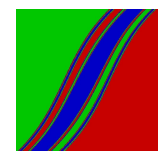


Illustrations/Variations/Extensions: Continuous-State Markov Processes

$AR(1)$

$$\alpha_t = T\alpha_{t-1} + \eta_t$$

$$\eta_t \sim WN$$



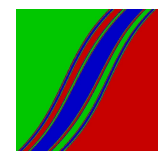
Illustrations/Variations/Extensions: Continuous-State Markov Processes State-Space System

$$\alpha_t = T\alpha_{t-1} + \eta_t$$

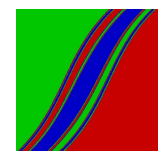
$$y_t = Z\alpha_t + \zeta_t$$

$$\eta_t \sim WN,\ \zeta_t \sim WN$$

We will now proceed to study state-space systems in significant depth.



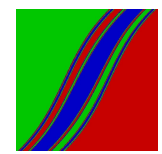
Part II: State Space



Transition Equation

$$\underset{mx1}{\alpha_t} = \underset{mxm}{T}\ \underset{mx1}{\alpha_{t-1}} + \underset{mxg}{R}\ \underset{gx1}{\eta_t}$$

$$t = 1,\ 2,\ ...,\ T$$

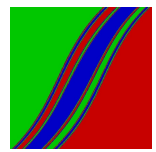


Measurement Equation

$$\underset{1x1}{y_t} = \underset{1xm}{Z}\ \underset{mx1}{\alpha_t} + \underset{1xL}{\Gamma}\ \underset{Lx1}{w_t} + \underset{1x1}{\zeta_t}$$

(This is for univariate y. We'll do multivariate shortly.)

$$t = 1,\ 2,\ ...,\ T$$

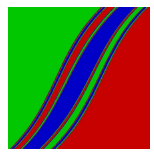


(Important) Details

$$\begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix} \sim WN\left(\underline{0}, \text{diag}(\underbrace{Q}_{g\times g}, \underbrace{h}_{1\times 1})\right)$$

$$E(\alpha_0 \ \eta_t \ ') \ = \ 0_{mxg}$$

$$E(\alpha_0 \ \zeta_t) \ = \ 0_{mx1}$$



All Together Now

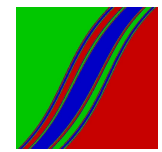
$$\underset{m\times 1}{\alpha_t} = \underset{m\times m}{T}\ \underset{m\times 1}{\alpha_{t-1}} + \underset{m\times g}{R}\ \underset{g\times 1}{\eta_t}$$

$$\underset{1\times 1}{y_t} = \underset{1\times m}{Z}\ \underset{m\times 1}{\alpha_t} + \underset{1\times L}{\Gamma}\ \underset{L\times 1}{w_t} + \underset{1\times 1}{\zeta_t}$$

$$\begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix} \sim WN\left(\mathbf{0}, \text{diag}(\underbrace{Q}_{g\times g}, \underbrace{h}_{1\times 1})\right)$$

$$E(\alpha_0\ \zeta_t) = 0_{m\times 1} \qquad E(\alpha_0\ \eta_t\ ') = 0_{m\times g}$$

(Covariance stationary case: All eigenvalues of T inside $|z| = 1$)



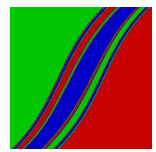
Our Assumptions Balance Generality vs. Tedium

– Could allow time-varying system matrices

– Could allow exogenous variables in measurement equation

– Could allow correlated measurement and transition disturbances

– Could allow for non-linear structure



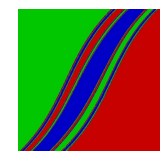
State Space Representations Are Not Unique

Transform by the nonsingular matrix B.

The original system is:

$$\underset{m\text{x}1}{\alpha_t} = \underset{m\text{x}m}{T}\ \underset{m\text{x}1}{\alpha_{t-1}} + \underset{m\text{x}g}{R}\ \underset{g\text{x}1}{\eta_t}$$

$$\underset{1\text{x}1}{y_t} = \underset{1\text{x}m}{Z}\ \underset{m\text{x}1}{\alpha_t} + \underset{1\text{x}1}{\zeta_t}$$



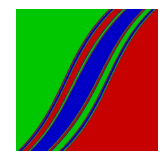
State Space Representations Are Not Unique: Step 1

Rewrite the system in two steps

First, write it as:

$$\underset{m\text{x}1}{\alpha_t} = \underset{m\text{x}m}{T}\ \underset{m\text{x}m}{B^{-1}}\ \underset{m\text{x}m}{B}\ \underset{m\text{x}1}{\alpha_{t-1}} + \underset{m\text{x}g}{R}\ \underset{g\text{x}1}{\eta_t}$$

$$\underset{1\text{x}1}{y_t} = \underset{1\text{x}m}{Z}\ \underset{m\text{x}m}{B^{-1}}\ \underset{m\text{x}m}{B}\ \underset{m\text{x}1}{\alpha_t} + \underset{1\text{x}1}{\zeta_t}$$



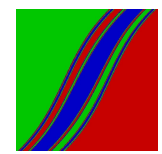
State Space Representations Are Not Unique: Step 2

Second, premultiply the transition equation by B to yield:

$$\underset{mx1}{(B\ \alpha_t)} = \underset{mxm}{(B\ T\ B^{-1})}\ \underset{mx1}{(B\ \alpha_{t-1})} + \underset{mxg}{(B\ R)}\ \underset{gx1}{\eta_t}$$

$$\underset{1x1}{y_t} = \underset{1xm}{(Z\ B^{-1})}\ \underset{mx1}{(B\ \alpha_t\ )} + \underset{1x1}{\zeta_t}$$

(Equivalent State Space Representation)



AR(1) State Space Representation

$$y_t = \phi\, y_{t-1} + \eta_t$$

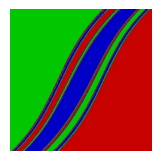
$$\eta_t \sim WN(0, \sigma_\eta^2)$$

Already in state space form!

$$\alpha_t = \phi\, \alpha_{t-1} + \eta_t$$

$$y_t = \alpha_t$$

$$(T = \phi,\ R = 1,\ Z = 1,\ Q = \sigma_\eta^2,\ h = 0)$$



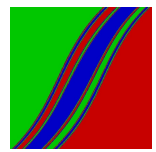
MA(1) in State Space Form

$$y_t = \eta_t + \theta \, \eta_{t-1}$$

$$\eta_t \sim WN(0, \sigma_\eta^2)$$

$$\begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ \theta \end{pmatrix} \eta_t$$

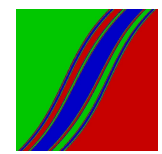
$$y_t = (1, \, 0) \, \alpha_t = \alpha_{1t}$$



MA(1) in State Space Form

Why? Recursive substitution from the bottom up yields:

$$\alpha_t = \begin{pmatrix} y_t \\ \theta\eta_t \end{pmatrix}$$



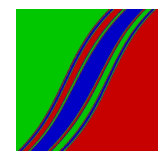
MA(q) in State Space Form

$$y_t \;=\; \eta_t \;+\; \theta_1\, \eta_{t-1} \;+\; \ldots \;+\; \theta_q\, \eta_{t-q}$$

$$\eta_t \;\sim\; WN\; N(0, \sigma_\eta^2)$$

$$\begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \vdots \\ \alpha_{q+1,t} \end{pmatrix} = \begin{pmatrix} 0 & \\ 0 & I_q \\ \vdots & \\ 0 & 0' \end{pmatrix} \begin{pmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \\ \vdots \\ \alpha_{q+1,t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_q \end{pmatrix} \eta_t$$

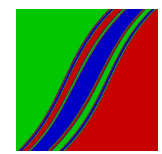
$$y_t \;=\; (1,\; 0,\; \ldots,\; 0)\, \alpha_t \;=\; \alpha_{1t}$$



MA(q) in State Space Form

Recursive substitution from the bottom up yields:

$$\alpha_t \equiv \begin{pmatrix} \theta_q \eta_{t-q} + \ldots + \theta_1 \eta_{t-1} + \eta_t \\ \vdots \\ \theta_q \eta_{t-1} + \theta_{q-1} \eta_t \\ \theta_q \eta_t \end{pmatrix} = \begin{pmatrix} y_t \\ \vdots \\ \theta_q \eta_{t-1} + \theta_{q-1} \eta_t \\ \theta_q \eta_t \end{pmatrix}$$



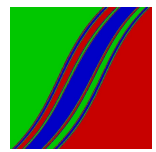
AR(p) in State Space Form

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \eta_t$$

$$\eta_t \sim WN(0, \sigma_\eta^2)$$

$$\alpha_t = \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \vdots \\ \alpha_{pt} \end{pmatrix} = \begin{pmatrix} \phi_1 & \\ \phi_2 & I_{p-1} \\ \vdots & \\ \phi_p & 0' \end{pmatrix} \begin{pmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \\ \vdots \\ \alpha_{p,t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \eta_t$$

$$y_t = (1, 0, \ldots, 0)\, \alpha_t = \alpha_{1t}$$

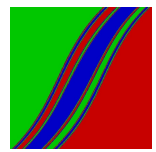


AR(p) in State Space Form

Recursive substitution from the bottom up yields:

$$\alpha_t = \begin{pmatrix} \alpha_{1t} \\ \vdots \\ \alpha_{p-1,t} \\ \alpha_{pt} \end{pmatrix} = \begin{pmatrix} \phi_1\alpha_{1,t-1} + \ldots + \phi_p\alpha_{1,t-p} + \eta_t \\ \vdots \\ \phi_{p-1}\alpha_{1,t-1} + \phi_p\alpha_{1,t-2} \\ \phi_p\alpha_{1,t-1} \end{pmatrix}$$

$$= \begin{pmatrix} y_t \\ \vdots \\ \phi_{p-1}y_{t-1} + \phi_p y_{t-2} \\ \phi_p y_{t-1} \end{pmatrix}$$



ARMA(p,q) in State Space Form

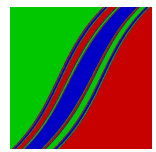
$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \eta_t + \theta_1 \eta_{t-1} + ... + \theta_q \eta_{t-q}$$

$$\eta_t \sim WN(0, \sigma_\eta^2)$$

Let $m = max(p, q+1)$ and write as ARMA$(m, m-1)$:

$$(\phi_1,\ \phi_2,\ ...,\ \phi_m) \ = \ (\phi_1,\ ...,\ \phi_p,\ 0,\ ...,\ 0)$$

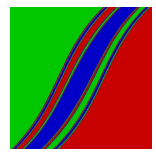
$$(\theta_1,\ \theta_2,\ ...,\ \theta_{m-1}) \ = \ (\theta_1,\ ...,\ \theta_q,\ 0,\ ...,\ 0)$$



ARMA(p,q) in State Space Form

$$\alpha_t = \begin{pmatrix} \phi_1 & \\ \phi_2 & I_{m-1} \\ \vdots & \\ \phi_m & 0' \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-1} \end{pmatrix} \eta_t$$

$$y_t = (1, 0, ..., 0)\, \alpha_t$$

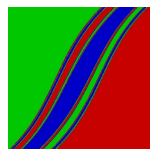


ARMA(p,q) in State Space Form

Recursive substitution from the bottom up yields:

$$\begin{pmatrix} \alpha_{1t} \\ \vdots \\ \alpha_{m-1,t} \\ \alpha_{mt} \end{pmatrix} = \begin{pmatrix} \phi_1\alpha_{1,t-1} + \phi_p\alpha_{1,t-p} + \eta_t + \theta_1\eta_{t-1} + \cdots + \theta_q\eta_{t-q} \\ \vdots \\ \phi_{m-1}\alpha_{1,t-1} + \alpha_{m,t-1} + \theta_{m-2}\eta_t \\ \phi_m\alpha_{1,t-1} + \theta_{m-1}\eta_t \end{pmatrix}$$

$$= \begin{pmatrix} y_t \\ \vdots \\ \phi_{m-1}y_{t-1} + \phi_m y_{t-2} + \theta_{m-1}\eta_{t-1} + \theta_{m-2}\eta_t \\ \phi_m y_{t-1} + \theta_{m-1}\eta_t \end{pmatrix}$$



Multivariate State Space

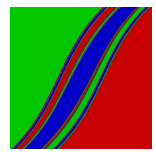
(Same framework, $N > 1$ observables)

$$\underset{mx1}{\alpha_t} = \underset{mxm}{T}\ \underset{mx1}{\alpha_{t-1}} + \underset{mxg}{R}\ \underset{gx1}{\eta_t}$$

$$\underset{Nx1}{y_t} = \underset{Nxm}{Z}\ \underset{mx1}{\alpha_t} + \underset{Nx1}{\zeta_t}$$

$$\begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix} \sim WN\left(\underline{0}, \text{diag}(\underbrace{Q}_{g\times g}, \underbrace{H}_{N\times N})\right)$$

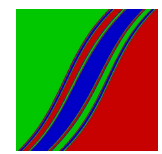
$$E(\alpha_0\ \eta_t') = 0_{mxg} \qquad E(\alpha_0\ \zeta_t') = 0_{mxN}$$



N-Variable VAR(p)

$$\underset{Nx1}{y_t} = \underset{NxN}{\Phi_1} y_{t-1} + \ldots + \Phi_p y_{t-p} + \underset{Nx1}{\eta_t}$$

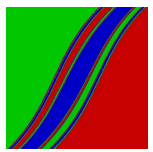
$$\eta_t \sim WN(0, \Sigma)$$



State Space Representation

$$\underset{Npx1}{\begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \vdots \\ \alpha_{pt} \end{pmatrix}} = \underset{NpxNp}{\begin{pmatrix} \Phi_1 & \\ \Phi_2 & I_{N(p-1)} \\ \vdots & \\ \Phi_p & 0' \end{pmatrix}} \underset{Npx1}{\begin{pmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \\ \vdots \\ \alpha_{p,t-1} \end{pmatrix}} + \underset{NPxN}{\begin{pmatrix} I_N \\ 0_{NxN} \\ \vdots \\ 0_{NxN} \end{pmatrix}} \eta_t$$

$$\underset{Nx1}{y_t} = \underset{NxNp}{(I_N, \quad 0_N \quad , \ ..., \ 0_N)} \underset{Npx1}{\alpha_t}$$

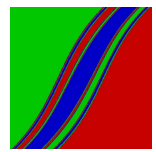


N-Variable VARMA(p,q)

$$\underset{Nx1}{y_t} = \underset{NxN}{\Phi_1}\, y_{t-1} + \ldots + \underset{NxN}{\Phi_p}\, y_{t-p}$$

$$+ \eta_t + \underset{NxN}{\Theta_1}\, \eta_{t-1} + \ldots + \underset{NxN}{\Theta_q}\, \eta_{t-q}$$

$$\eta_t \sim WN(0, \Sigma)$$

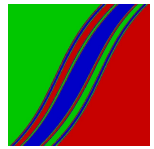


N-Variable VARMA(p,q)

$$\underset{Nm\times 1}{\alpha_t} = \begin{pmatrix} \Phi_1 & \\ \Phi_2 & I_{N(m-1)} \\ \vdots & \\ \Phi_m & 0_{N\times N(m-1)} \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} I \\ \Theta_1 \\ \vdots \\ \Theta_{m-1} \end{pmatrix} \eta_t$$

$$y_t = (I,\ 0,\ ...,\ 0)\ \alpha_t = \alpha_{1t}$$

$$\text{where } m = \max(p,\ q+1)$$



Linear Regression

Transition:

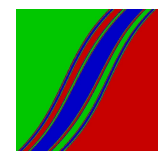
$$\alpha_t \;=\; \alpha_{t-1}$$

Measurement:

$$y_t \;=\; x_t'\alpha_t \;+\; \zeta_t$$

$$(\,T = I,\;\; R = 0,\;\; Z_t = x_t',\;\; H = \sigma_\zeta^2\,)$$

Note the time-varying system matrix.



Linear Regression with Time-Varying Coefficients

Transition:

$$\alpha_t \;=\; \phi\,\alpha_{t-1} \;+\; \eta_t$$

Measurement:

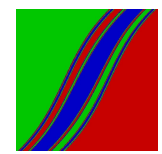
$$y_t \;=\; x_t'\alpha_t \;+\; \zeta_t$$

$$(T = \phi,\; R = I,\; Q = cov(\eta_t),\; Z_t = x_t',\; H = \sigma_\zeta^2)$$

– Gradual evolution of tastes, technologies and institutions

– Lucas critique

– Stationary or non-stationary



Linear Regression with ARMA(p,q) Disturbances

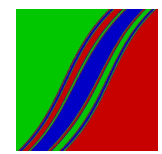
$$y_t = \beta x_t + u_t$$

$$u_t = \phi_1 u_{t-1} + \ldots + \phi_p u_{t-p} + \eta_t + \phi_1 \eta_{t-1} + \ldots + \theta_q \eta_{t-q}$$

$$\alpha_t = \begin{pmatrix} \phi_1 & \\ \phi_2 & I_{m-1} \\ \vdots & \\ \phi_m & 0' \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-1} \end{pmatrix} \eta_t$$

$$y_t = (1,\ 0,\ \ldots,\ 0)\alpha_t + \beta x_t$$

$$\text{where } m = \max(p,\ q+1)$$



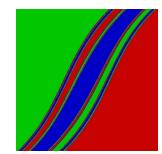
Signal + Noise Model
"Unobserved Components"

$$y_t = x_t + \zeta_t$$

$$x_t = \phi x_{t-1} + \eta_t$$

$$\begin{pmatrix} \zeta_t \\ \eta_t \end{pmatrix} \sim WN\left(0, \begin{pmatrix} \sigma_\zeta^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right)$$

$$(\alpha_t = x_t,\ T = \phi,\ R = 1,\ Z = 1,\ Q = \sigma_\eta^2,\ H = \sigma_\zeta^2)$$

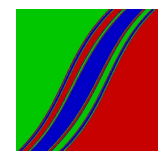


Cycle + Seasonal + Noise

$$y_t = c_t + s_t + \zeta_t$$

$$c_t = \phi\, c_{t-1} + \eta_{ct}$$

$$s_t = \gamma\, s_{t-4} + \eta_{st}$$

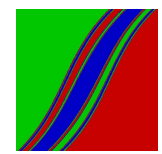


Cycle + Seasonal + Noise

Transition equations for the cycle and seasonal:

$$\alpha_{ct} = \phi\, \alpha_{c,t-1} + \eta_{ct}$$

$$\alpha_{st} = \begin{pmatrix} 0 & \\ 0 & I_3 \\ 0 & \\ \gamma & 0' \end{pmatrix} \alpha_{s,t-1} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \eta_{st}$$



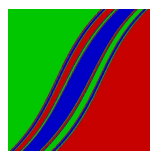
Cycle + Seasonal + Noise

Stacking transition equations gives the grand transition equation:

$$\begin{pmatrix} \alpha_{st} \\ \alpha_{ct} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \gamma & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi \end{pmatrix} \begin{pmatrix} \alpha_{s,t-1} \\ \alpha_{c,t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{st} \\ \eta_{ct} \end{pmatrix}$$

Finally, the measurement equation is:

$$y_t = (1,\ 0,\ 0,\ 0,\ 1) \begin{pmatrix} \alpha_{st} \\ \alpha_{ct} \end{pmatrix} + \zeta_t$$



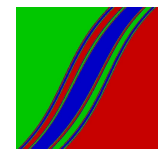
Dynamic Factor Model – Single AR(1) Factor

(White noise idiosyncratic factors uncorrelated with each other and uncorrelated with the factor at all leads and lags...)

$$\begin{pmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix} f_t + \begin{pmatrix} \zeta_{1t} \\ \vdots \\ \zeta_{Nt} \end{pmatrix}$$

$$f_t = \phi f_{t-1} + \eta_t$$

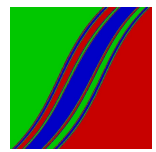
Already in state-space form!



Dynamic Factor Model – Single ARMA(p,q) Factor

$$\begin{pmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix} f_t + \begin{pmatrix} \zeta_{1t} \\ \vdots \\ \zeta_{Nt} \end{pmatrix}$$

$$\Phi(L)\ f_t \ = \ \Theta(L)\ \eta_t$$

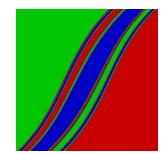


Dynamic Factor Model – Single ARMA(p,q) Factor

State vector for f is state vector for system.

System transition:

$$\alpha_t = \begin{pmatrix} \phi_1 & \\ \phi_2 & I_{m-1} \\ \vdots & \\ \phi_m & 0' \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-1} \end{pmatrix} \eta_t$$

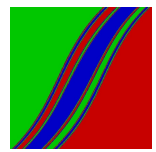


Dynamic Factor Model – Single ARMA(p,q) Factor

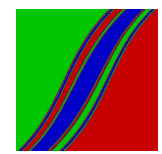
System measurement:

$$\begin{pmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix} (1,\ 0,\ ...,\ 0)\ \alpha_t + \begin{pmatrix} \zeta_{1t} \\ \vdots \\ \zeta_{Nt} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ \vdots & & & \\ \lambda_N & 0 & ... & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} \zeta_{1t} \\ \vdots \\ \zeta_{Nt} \end{pmatrix}$$



Part III: The Kalman Filter



State Space Representation

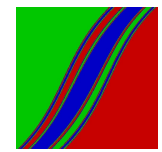
$$\underset{mx1}{\alpha_t} = \underset{mxm}{T}\ \underset{mx1}{\alpha_{t-1}} + \underset{mxg}{R}\ \underset{gx1}{\eta_t}$$

$$\underset{Nx1}{y_t} = \underset{Nxm}{Z}\ \underset{mx1}{\alpha_t} + \underset{Nx1}{\zeta_t}$$

$$\begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix} \sim WN\left(0, \text{diag}(\underbrace{Q}_{g\times g}, \underbrace{H}_{N\times N})\right)$$

$$E(\alpha_0\ \eta_t') = 0_{mxg}$$

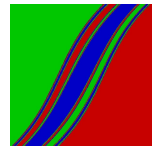
$$E(\alpha_0\ \zeta_t') = 0_{mxN}$$



The Filtering "Thought Experiment"

– Prediction and updating

– "Online", "ex ante", using only real-time available data $\tilde{y}_t$
to extract α_t and predict α_{t+1},
where $\tilde{y}_t = \{y_1, ..., y_t\}$

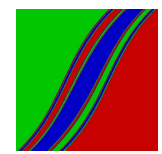


Statement of the Kalman Filter

I. Initial state estimate and MSE

$$a_0 = E(\alpha_0)$$

$$P_0 = E(\alpha_0 - a_0)(\alpha_0 - a_0)'$$



Statement of the Kalman Filter

II. Prediction Recursions

$$a_{t/t-1} = T\, a_{t-1}$$

$$P_{t/t-1} = T\, P_{t-1}\, T' + R\, Q\, R'$$

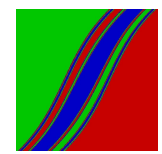
III. Updating Recursions

$$a_t = a_{t/t-1} + P_{t/t-1}\, Z'\, F_t^{-1}\, (y_t - Z a_{t/t-1})$$

$$(\text{where } F_t = Z\, P_{t/t-1}\, Z' + H)$$

$$P_t = P_{t/t-1} - P_{t/t-1}\, Z'\, F_t^{-1}\, Z\, P_{t/t-1}$$

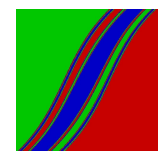
$$t = 1, ..., T$$



State-Space Representation in Density Form (Assuming Normality)

$$\alpha_t | \alpha_{t-1} \sim N(T\alpha_{t-1},\ RQR')$$

$$y_t | \alpha_t \sim N(Z\alpha_t,\ H)$$



Kalman Filter in Density Form (Assuming Normality)

Initialize at a_0, P_0

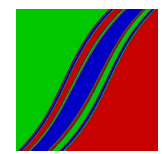
State prediction:

$$\alpha_t|\tilde{y}_{t-1} \sim N(a_{t/t-1},\ P_{t/t-1})$$
$$a_{t/t-1} = Ta_{t-1}$$
$$P_{t/t-1} = TP_{t-1}T' + RQR'$$

Update:

$$\alpha_t|\tilde{y}_t \sim N(a_t,\ P_t)$$
$$a_t = a_{t/t-1} + K_t(y_t - Za_{t/t-1})$$
$$P_t = P_{t/t-1} - K_tZP_{t/t-1}$$

where $\tilde{y}_t = \{y_1, ..., y_t\}$



Useful Result 1: Conditional Expectation is MVUE Extraction Under Normality

Suppose that

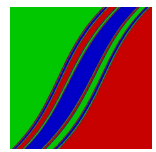
$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\mu, \Sigma\right)$$

where x is unobserved and y is observed.

Then

$$E(x|y) = argmin_{\hat{x}(y)} \int \int (x - \hat{x}(y))^2 \; f(x, y) \; dx \; dy,$$

where $\hat{x}(y)$ is any unbiased extraction of x based on y



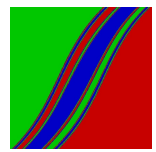
Useful Result 2: Properties of the Multivariate Normal

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\mu, \Sigma\right) \quad \mu = (\mu_x, \mu_y)' \qquad \Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}$$

$$\Longrightarrow x|y \sim \quad N\left(\mu_{x|y} \ , \ \Sigma_{x|y}\right)$$

$$\mu_{x|y} \ = \ \mu_x \ + \ \Sigma_{xy} \, \Sigma_{yy}^{-1} \, (y \ - \ \mu_y)$$

$$\Sigma_{x|y} \ = \ \Sigma_{xx} \ - \ \Sigma_{xy} \, \Sigma_{yy}^{-1} \, \Sigma_{yx}$$



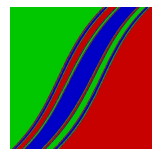
Constructive Derivation of the Kalman Filter Under Normality

$$\text{Let } E_t(\cdot) \equiv E(\cdot\,|\Omega_t), \text{ where } \Omega_t \equiv \{y_1, ..., y_t\}.$$

Time 0 "update" (initialization):

$$a_0 = E_0(\alpha_0) = E(\alpha_0)$$

$$P_0 = var_0(\alpha_0) = E\left[(\alpha_0 - a_0)(\alpha_0 - a_0)'\right]$$



Derivation of the Kalman Filter, Continued...

Time 0 prediction

At time 1 we know that:

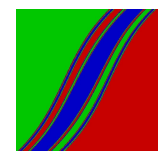
$$\alpha_1 = T\alpha_0 + R\eta_1$$

Now take expectations conditional on time-0 information:

$$E_0(\alpha_1) = TE_0(\alpha_0) + RE_0(\eta_1)$$

$$= Ta_0$$

$$= a_{1/0}$$



Derivation of the Kalman Filter, Continued...

Time 0 prediction covariance matrix

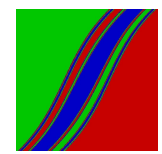
$$E_0\left((\alpha_1 - a_{1/0})\,(\alpha_1 - a_{1/0})'\right)$$

$$= E_0\left((\alpha_1 - Ta_0)\,(\alpha_1 - Ta_0)'\right) \quad \text{(subst. } a_{1/0}\text{)}$$

$$= E_0\left((T(\alpha_0 - a_0) + R\eta_1)\,(T(\alpha_0 - a_0) + R\eta_1)'\right) \quad \text{(subst. } \alpha_1\text{)}$$

$$= TP_0\,T' + RQR' \quad \text{(using } E(\alpha_0\,\eta_t') = 0\ \forall t\text{)}$$

$$= P_{1/0}$$



Derivation of the Kalman Filter, Continued...

Time 1 updating

We will derive the distribution of:

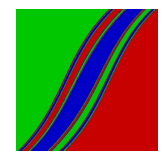
$$\left(\begin{array}{c} \alpha_1 \\ y_1 \end{array} \right) \Big| \Omega_0$$

and then convert to

$$\alpha_1 | (\Omega_0 \cup y_1)$$

or

$$\alpha_1 | \Omega_1$$

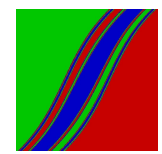


Derivation of the Kalman Filter, Continued...

Means:

$$E_0(\alpha_1) = a_{1/0}$$

$$E_0(y_1) = Za_{1/0}$$



Derivation of the Kalman Filter, Continued...

Variance-Covariance Matrix:

$$var_0(\alpha_1) = E_0\left((\alpha_1 - a_{1/0})\,(\alpha_1 - a_{1/0})\right) = P_{1/0}$$

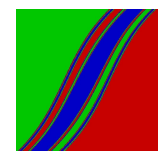
$$var_0(y_1) = E_0\left((y_1 - Za_{1/0})\,(y_1 - Za_{1/0})'\right)$$

$$= E_0\left((Z(\alpha_1 - a_{1/0}) + \zeta_1)\,(Z(\alpha_1 - a_{1/0}) + \zeta_1)'\right)$$

$$= Z\,P_{1/0}Z' + H \text{ (using } \zeta \perp \eta)$$

$$cov_0\,(\alpha_1, y_1) = E_0(\alpha_1 - a_{1/0})\,(Z\,(\alpha_1 - a_{1/0}) + \zeta_1)'$$

$$= P_{1/0}Z' \text{ (using } \zeta \perp \eta)$$



Derivation of the Kalman Filter, Continued...

Hence:

$$\begin{pmatrix} \alpha_1 \\ y_1 \end{pmatrix} \Big| \Omega_0 \sim N\left(\begin{pmatrix} a_{1/0} \\ Z a_{1/0} \end{pmatrix}, \begin{pmatrix} P_{1/0} & P_{1/0}\, Z' \\ Z P_{1/0} & Z P_{1/0}\, Z' + H \end{pmatrix} \right)$$

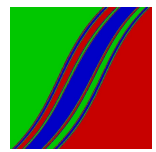
Now by Useful Result 2, $\alpha_1 \,|\, \Omega_0 \cup y_1 \sim N(a_1,\ P_1)$

$$a_1 = a_{1/0} + P_{1/0}\, Z'\, F_1^{-1}\, (y_1 - Z a_{1/0})$$

$$P_1 = P_{1/0} - P_{1/0}\, Z'\, F_1^{-1}\, Z\, P_{1/0}$$

$$(F_1 = Z\, P_{1/0}\, Z' + H)$$

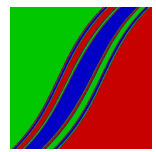
Repeating yields the Kalman filter.



What Have We Done?

Under normality,
we proved that the Kalman filter delivers
best predictions and extractions, for a standard and appropriate definition of "best".
"MVUE"

Dropping normality,
similar results continue to hold
(best *linear* predictions and extractions).
"BLUE"



Calculation of Initial Covariance Matrix $P_0 = \Gamma(0)$ When $a_0 = 0$

$$\alpha_t = T\alpha_{t-1} + R\eta_t$$

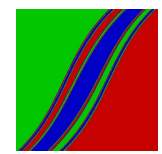
$$\implies P_0 = E(T\alpha_{t-1} + R\eta_t)(T\alpha_{t-1} + R\eta_t)' = TP_0T' + RQR'$$

$$\implies vec(P_0) = vec(TP_0T') + vec(RQR')$$

$$= (T \otimes T)vec(P_0) + vec(RQR')$$

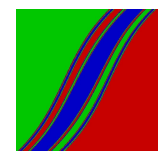
$$\implies (I - (T \otimes T))vec(P_0) = vec(RQR')$$

$$\implies vec(P_0) = (I - (T \otimes T))^{-1}vec(RQR')$$



The Smoothing Thought Experiment

– "offline", "ex post", using *all* data $\tilde{y}_T$ to extract α_t.



The Kalman Smoother

1. Kalman filter forward through the sample, $t = 1, \ ..., \ T$

2. Smooth backward, $t = T, \ (T-1), \ (T-2), \ ..., \ 1$

Initialize: $a_{T,T} = a_T, \ P_{T,T} = P_T$

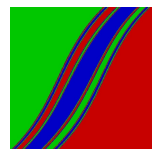
Then:

$$a_{t,T} = a_t + J_t(a_{t+1,T} - a_{t+1,t})$$

$$P_{t,T} = P_t + J_t(P_{t+1,T} - P_{t+1,t})J_t'$$

where

$$J_t = P_t T' P_{t+1,t}^{-1}$$



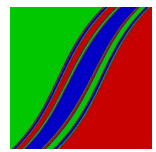
Point Prediction of y_t

Prediction:

$$y_{t/t-1} = Za_{t/t-1}$$

Prediction error:

$$v_t = y_t - Za_{t/t-1}$$



Density Prediction of y_t

$$y_t|\Omega_{t-1} \sim N(Za_{t/t-1}, F_t)$$

or equivalently

$$v_t|\Omega_{t-1} \sim N(0, F_t)$$

Normality follows from linearity of all transformations.

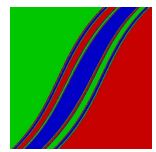
Conditional mean already derived.

Proof that the conditional covariance matrix is F_t:

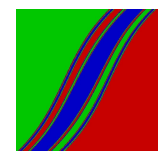
$$E_{t-1}v_t v_t' = E_{t-1}[Z(\alpha_t - a_{t/t-1}) + \zeta_t][Z(\alpha_t - a_{t/t-1}) + \zeta_t]'$$

$$= ZP_{t/t-1}Z' + H$$

$$= F_t$$



Part IV:
The Innovations (Steady-State) Representation



Combining State Vector Prediction and Updating

$$\text{(1) Prediction: } a_{t+1/t} = Ta_t$$

$$\text{(2) Update: } a_t = a_{t/t-1} + P_{t/t-1} Z' F_t^{-1} (y_t - Za_{t/t-1})$$

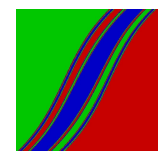
$$= a_{t/t-1} + K_t v_t$$

where

$$K_t = P_{t/t-1} Z' F_t^{-1}$$

Substituting (2) into (1):

$$a_{t+1/t} = Ta_{t/t-1} + TK_t v_t$$



Combining Covariance Matrix Prediction and Updating

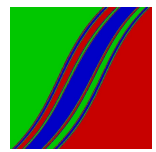
(1) Prediction: $P_{t+1/t} = T\,P_t\,T' + RQR'$

(2) Update: $P_t = P_{t/t-1} - K_t\,Z\,P_{t/t-1}$

Substitute (2) into (1):

$$P_{t+1/t} = T\,P_{t/t-1}\,T' - T\,K_t\,Z\,P_{t/t-1}\,T' + RQR'$$

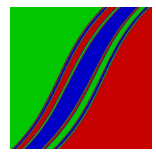
(Matrix Ricatti equation)



Why Care About Combining Prediction and Updating?

It leads us to the notion of steady state of the Kalman filter...

...which is the bridge from the Wold representation
to the state space representation



"Two-Shock" State Space Representation

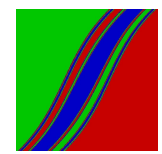
$$\alpha_t = T\alpha_{t-1} + R\eta_t$$

$$y_t = Z\alpha_t + \zeta_t$$

$$E(\eta_t\eta_t') = Q$$

$$E(\zeta_t\zeta_t') = H$$

(Nothing new)



"One-Shock" ("Prediction Error") Representation

We have seen that

$$a_{t+1|t} = Ta_{t|t-1} + TK_t\ v_t \quad \text{(transition)}$$

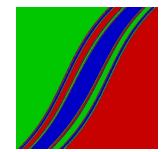
Moreover, it is tautologically true that

$$y_t = Z\ a_{t|t-1} + (y_t - Za_{t|t-1})$$

$$= Z\ a_{t|t-1} + v_t \quad \text{(measurement)}$$

Note that one-shock state space representation has time-varying system matrices:

- "R matrix" in transition equation is TK_t
- Covariance matrix of v_t is F_t



"Innovations" (Steady-State) Representation

If as $T \to \infty$, $P_{t/t-1} \to \bar{P}$,
where $\bar{P}$ solves the matrix Ricatti equation, then:

$$a_{t+1|t} = Ta_{t|t-1} + T\bar{K}\varepsilon_t$$

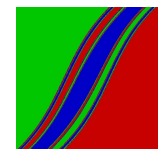
$$y_t = Z\, a_{t|t-1} + \varepsilon_t$$

where

$$\bar{K} = \bar{P}\, Z'\, \bar{F}^{-1}$$

$$E(\varepsilon_t \varepsilon_t') = \bar{F} = Z\bar{P}Z' + H$$

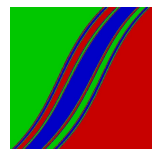
– Effectively Wold-Wiener-Kolmogorov prediction and extraction
– Prediction $y_{t+1/t}$ is now the projection of y_{t+1} on *infinite* past, and the finite-history prediction errors v_t are now the infinite-history Wold-Wiener-Kolmogorov innovations ε_t



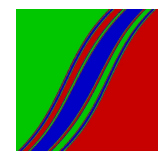
Remarks on the Steady State

1. Steady state $\bar{P}$ exists if:
   - the underlying two-shock system is time invariant
   - all eigenvalues of T are less than one
   - $P_{1|0}$ is positive semidefinite

2. Because the recursions for $P_{t|t-1}$ and K_t don't depend on the data, but only on P_0, we can calculate arbitrarily close approximations to $\bar{P}$ and $\bar{K}$ by letting the Kalman filter run



Likelihood Evaluation and Optimization



Gaussian Likelihood by Brute Force
Univariate Zero-Mean Gaussian $AR(1)$ Example

Process: $y_t = \phi y_{t-1} + \varepsilon_t$

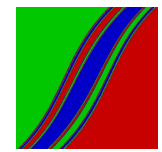
$$\varepsilon_T \sim iidN(0, \sigma^2)$$

$T \times 1$ Sample path:

$$y \sim N(\underline{0},\ \Sigma(\phi, \sigma^2)),$$

where

$$\Sigma_{ij}(\phi, \sigma^2) = \frac{\sigma^2}{1-\phi^2}\phi^{|i-j|}$$



Gaussian Likelihood by Brute Force
Univariate Zero-Mean Gaussian $AR(1)$ Example Continued

$$L(y; \phi, \sigma^2) = (2\pi)^{T/2} |\Sigma(\phi, \sigma^2)|^{-1/2} exp\left(-\frac{1}{2} y' \Sigma^{-1}(\phi, \sigma^2) y\right)$$

$$lnL(y; \phi, \sigma^2) = const - \frac{1}{2} ln|\Sigma(\phi, \sigma^2)| - \frac{1}{2} y' \Sigma^{-1}(\phi, \sigma^2) y$$

– Here Σ is easy to express analytically in terms of model parameters (ϕ, σ^2), but that only works in the simplest cases.

– In general Σ is *very* hard to express analytically in terms of model parameters.

– In any event, likelihood evaluation requires inversion of Σ, which is $T \times T$. *Very* hard except for very small T.

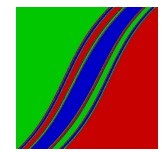
Gaussian Likelihood by Brute Force: General Case

$$L(y;\theta) = (2\pi)^{T/2}|\Sigma(\theta)|^{-1/2} exp\left(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(\theta)(y-\mu)\right)$$

$$lnL(y;\theta) = const - \frac{1}{2}ln|\Sigma(\theta)| - \frac{1}{2}(y-\mu)'\Sigma^{-1}(\theta)\ (y-\mu)$$

$T \times T$ matrix $\Sigma(\theta)$ is generally
very hard (read: impossible)
to calculate and invert

(And in the multivariate case $\Sigma(\theta)$ would be $NT \times NT$)



Gaussian Likelihood with Finesse: The Prediction-Error Likelihood Decomposition

In levels:

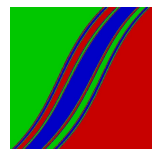
$$L(y_1, \ldots, y_T; \theta) = \prod_{t=1}^{T} L_t(y_t | y_{t-1}, \ldots, y_1; \theta)$$

$$L(v_1, \ldots, v_T; \theta) = \prod_{t=1}^{T} L_t(v_t; \theta)$$

In logs:

$$\ln L(y_1, \ldots, y_T; \theta) = \sum_{t=1}^{T} L_t(y_t | y_{t-1}, \ldots, y_1; \theta)$$

$$\ln L(v_1, \ldots, v_T; \theta) = \sum_{t=1}^{T} \ln L_t(v_t; \theta)$$



Prediction-Error Decomposition, Continued
Univariate Gaussian Case

$$\ln L = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln \sigma_t^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-\mu_t)^2}{\sigma_t^2}$$

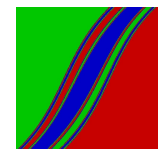
Now change the above "standard notation"
to "Kalman filter notation":

$$\ln L = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln F_t - \frac{1}{2}\sum_{t=1}^{T}\frac{v_t^2}{F_t}$$

Kalman filter delivers v_t and F_t!

No need for tedious analytic likelihood derivations!

No matrix inversion!



Prediction-Error Decomposition, Continued
Multivariate Gaussian Case

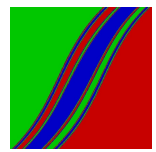
$$\ln L = -\frac{NT}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln|\Sigma_t| - \frac{1}{2}\sum_{t=1}^{T}(y_t-\mu_t)'\Sigma_t^{-1}(y_t-\mu_t)$$

or

$$\ln L = -\frac{NT}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln|F_t| - \frac{1}{2}\sum_{t=1}^{T}v_t'F_t^{-1}v_t$$

Kalman filter again delivers v_t and F_t.

Only the small $N \times N$ matrix F_t need be inverted.



Approximate (Asymptotic) Frequency Domain Gaussian Likelihood

We have:

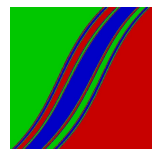
$$x_j = \frac{2\hat{f}(\omega_j)}{f(\omega_j;\theta)} \overset{d}{\rightarrow} \chi^2_2$$

where $f(\omega_j;\theta)$ is the spectral density and the χ^2_2 random variables are independent across frequencies

$$\omega_j = \frac{2\pi j}{T},\ j = 0, 1, ..., \frac{T}{2}$$

$\Rightarrow$ MGF of any one of the x_j's is

$$M_x(t) = \frac{1}{1-2t}$$



Approximate Frequency Domain Gaussian Likelihood, Continued

Now, because:

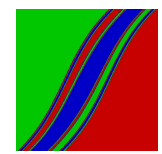
$$\hat{f}(\omega_j) \;=\; \frac{f(\omega_j;\theta)\; x_j}{2},$$

we can infer that:

$$M_{\hat{f}}(t) \;=\; M_x\left(\frac{f(\omega_j;\theta)}{2}\; t\right) = \frac{1}{1 - f(\omega_j;\theta)\; t}$$

This is the MGF of exponential rv with parameter $1/f(\omega_j;\theta)$. Hence the density (point likelihood) of $\hat{f}$ is:

$$I(\hat{f}(\omega_j);\theta) \;=\; \frac{1}{f(\omega_j;\theta)}\; e^{\frac{-\hat{f}(\omega_j)}{f(\omega_j;\theta)}}$$



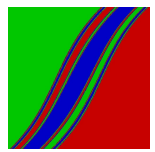
Approximate Frequency Domain Gaussian Likelihood, Continued

So the univariate asymptotic Gaussian log likelihood is:

$$\ln L(\hat{f}(\omega_j);\theta) = \sum_{j=0}^{T/2} \ln I(\hat{f}(\omega_j);\theta) = -\sum_{j=0}^{T/2} \ln f(\omega_j;\theta) - \sum_{j=0}^{T/2} \frac{\hat{f}(\omega_j)}{f(\omega_j;\theta)}$$

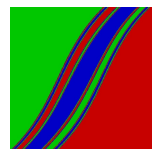
The multivariate asymptotic Gaussian log likelihood is:

$$\ln L(\hat{F};\theta) = -\sum_{j=0}^{T/2} \ln |F(\omega_j;\theta)| - trace\left(\sum_{j=0}^{T/2} F^{-1}(\omega_j;\theta)\,\hat{F}(\omega_j)\right)$$



Numerical Maximization of the Gaussian Likelihood

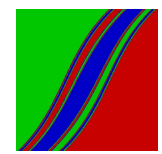
- The key is to be able to *evaluate* the likelihood for a given parameter configuration.
- Then we can climb uphill to *maximize* the likelihood.
- Now we will introduce methods for doing so.



Numerical Optimization: Basic Framework

Function $lnL(\theta)$ to be optimized w.r.t. θ,

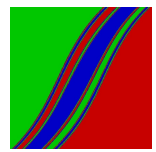
$\theta \in \Theta$, a compact subset of R^k



Crude Search ("Brute Force")

- Deterministic search: Search k dimensions at r locations in each dimension.
- Randomized Search: Repeatedly sample from Θ, repeatedly evaluating $lnL(\theta)$

  – Absurdly slow (curse of dimensionality)



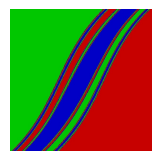
"Gradient-Based" Iterative Algorithms ("Line-Search")

Parameter vector at iteration m: $\theta^{(m)}$.

$\theta^{(m+1)} = \theta^{(m)} + C^{(m)}$, where $C^{(m)}$ is the **step**.

Gradient algorithms: $C^{(m)} = -t^{(m)} D^{(m)} s^{(m)}$

$t^{(m)}$ is the step length, or step size (a positive number)
$D^{(m)}$ is a positive definite direction matrix
$s^{(m)}$ is the score (gradient) vector evaluated at $\theta^{(m)}$



General Algorithm

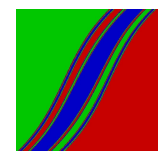
1. Specify $\theta^{(0)}$

2. Compute $D^{(m)}$ and $s^{(m)}$

3. Determine step length $t^{(m)}$
   (Often, at each step, choose $t^{(m)}$ to optimize the objective function ("variable step length"))

4. Compute $\theta^{(m+1)}$

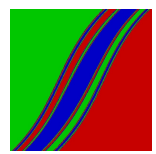
5. If convergence criterion not met, go to 2.



Convergence Criteria

$$\| \theta^{(m)} - \theta^{(m-1)} \| \text{ “small”}$$

$$\| s^{(m)} \| \text{ “small”}$$



Method of Steepest Decent

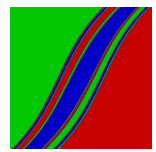
Use $D^{(m)} = I, t^{(m)} = 1, \ \forall \ m.$

Properties:

1. May converge to a critical point other than a minimum (of course)

2. Requires only first derivative of the objective function

3. Slow convergence



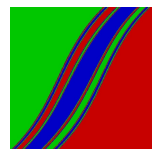
Newton's Method

Take $D^{(m)}$ as the inverse Hessian of $lnL(\theta)$ at $\theta^{(m)}$

$$D^{(m)} = H^{-1(m)} = \begin{pmatrix} \frac{\partial^2 lnL}{\partial \theta_1^2}|_{\theta^{(m)}} & \cdot & \cdot & \cdot & \frac{\partial^2 lnL}{\partial \theta_1 \partial \theta_k}|_{\theta^{(m)}} \\ & \cdot & & & \\ & \cdot & & & \\ & \cdot & & & \\ \frac{\partial^2 lnL}{\partial \theta_k \partial \theta_1}|_{\theta^{(m)}} & \cdot & \cdot & \cdot & \frac{\partial^2 lnL}{\partial \theta_k^2}|_{\theta^{(m)}} \end{pmatrix}^{-1}$$

Also take $t^{(m)} = 1$

Then $\theta^{(m+1)} = \theta^{(m)} - H^{-1(m)} s^{(m)}$



Derivation From Second-Order Taylor Expansion

Initial guess: $\theta^{(0)}$

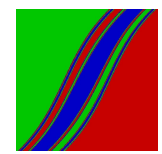
$$lnL(\theta) \approx lnL(\theta^{(0)}) + s^{(0)}(\theta - \theta^{(0)}) + \frac{1}{2}(\theta - \theta^{(0)})' H^{(0)}(\theta - \theta^{(0)})$$

F.O.C.:

$$s^{(0)} + H^{(0)}(\theta^* - \theta^{(0)}) = 0$$

or

$$\theta^* = \theta^{(0)} - H^{-1(0)} s^{(0)}$$



Properties of Newton

$lnL(\theta)$ quadratic $\Rightarrow$ full convergence in a single iteration

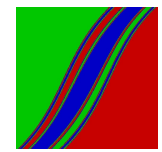
More generally, iterate to convergence:
$\theta^{(m+1)} = \theta^{(m)} - H^{-1(m)} s^{(m)}$

Quick Convergence

But there is a price:

Requires first *and* second derivatives of the objective function

Requires inverse Hessian at each iteration



The EM (Expectation/Maximization) Algorithm

Think of $\{\alpha_t\}_{t=0}^T$ as data that are unfortunately missing in

$$\alpha_t = T\alpha_{t-1} + \eta_t$$

$$y_t = Z\alpha_t + \zeta_t$$

$$t = 1, ..., T$$

Incomplete Data Likelihood:

$$lnL(\{y_t\}_{t=1}^T; \theta)$$

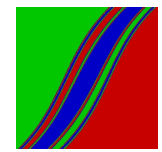
Complete Data Likelihood: (If only we had complete data!)

$$lnL\left(\{y_t\}_{t=1}^T, \{\alpha_t\}_{t=0}^T; \theta\right)$$

Expected Complete Data Likelihood:

$$lnL\left(\{y_t\}_{t=1}^T; \theta\right) \approx \mathbf{E}_\alpha\left[lnL\left(\{y_t\}_{t=1}^T, \{\alpha_t\}_{t=0}^T; \theta\right)\right]$$

EM iteratively constructs and maximizes the expected complete-data likelihood, which (amazingly) has same maximizer as the (relevant) incomplete-data likelihood.



EM has Strong Intuition

1. E Step:

Approximate a "complete data" situation by replacing $\{\alpha_t\}_{t=0}^{T}$ with $\{a_{t,T}\}_{t=0}^{T}$ from the Kalman smoother

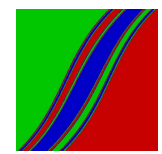
2. M Step:

Estimate parameters by running regressions:

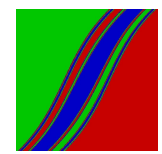
$$a_{t,T} \rightarrow a_{t-1,T}$$
$$y_t \rightarrow a_{t,T}$$

3. If convergence criterion not met, go to 1

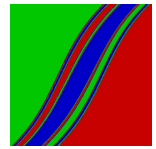
(Note: This slide provides some important intuition, but it also omits some important details.)



Simulation Methods in Econometrics



Simulation: The Basics



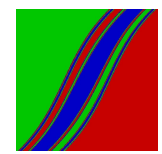
The Canonical Problem: $U(0,1)$

– Chaotic solutions to certain non-linear deterministic difference equations, appropriately scaled, appear indistinguishable from $U(0,1)$

– "Pseudo-random deviates"

– Assume that we have such a $U(0,1)$ pseudo-random number generator

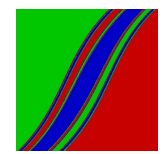
Given $U(0,\ 1)$, $U(\alpha,\ \beta)$ is immediate



Inverse cdf Method ("Inversion Methods")

Desired density: $f(y)$

1. Find the analytical c.d.f., $F(y)$, corresponding to $f(y)$
2. Generate T $U(0,1)$ deviates, $\{r_1, \ ..., \ r_T\}$
3. Calculate $\{F^{-1}(r_1), \ ..., \ F^{-1}(r_T)\}$



Graphical Representation of Inverse cdf Method

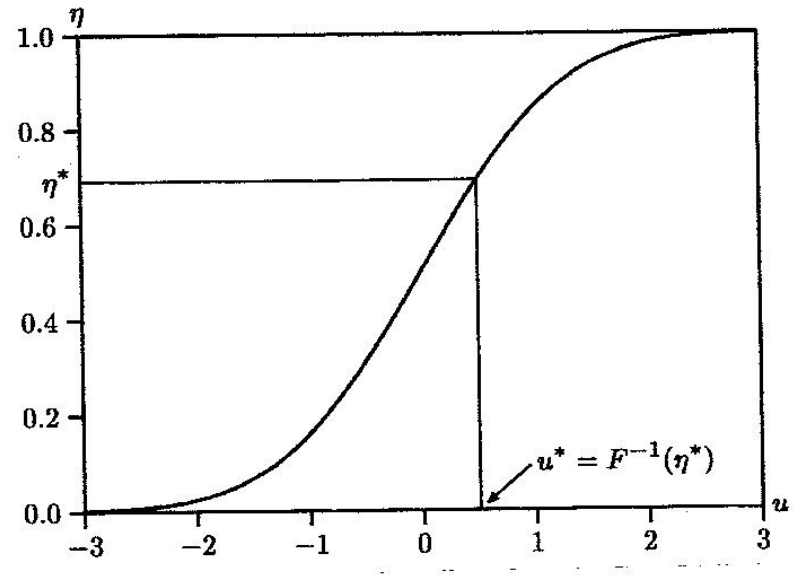
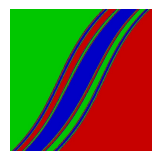


Figure: *Transforming from U(0,1) to f (from Davidson and MacKinnon, 1993)*



Example: Inverse cdf Method for exp(β) Deviates

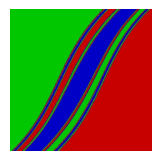
$f(y) = \beta e^{-\beta y}$ where $\beta > 0,\ y \geq 0$

$\Rightarrow F(y) = \int_0^y \beta e^{-\beta t}\, dt$

$= \frac{\beta e^{-\beta t}}{-\beta} \Big|_o^y = -e^{-\beta y} + 1 = 1 - e^{-\beta y}$

Hence $e^{-\beta y} = 1 - F(y)$ so $y = \frac{\ln(1 - F(y))}{-\beta}$

Then insert a $U(0,1)$ deviate for $F(y)$



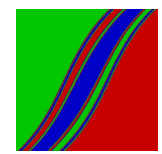
Complications

Analytic inverse cdf not always available
(e.g., $N(0,1)$ distribution).

- Approach 1: Evaluate the cdf numerically
- Approach 2: Use a different method

  e.g., CLT approximation:

  Take $\left(\sum_{i=1}^{12} U_i(0,1) \; - \; 6\right)$ for $N(0,1)$



An Efficient Gaussian Approach: Box-Muller

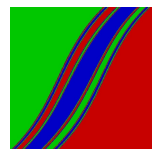
Let y_1 and y_2 be i.i.d. $U(0,1)$, and consider

$$z_1 = \sqrt{-2 \ln y_1}\ \cos(2\pi y_2)$$

$$z_2 = \sqrt{-2 \ln y_1}\ \sin(2\pi y_2)$$

Find the distribution of z_1 and z_2. We know that

$$f(z_1, z_2) = f(y_1, y_2) \cdot \left| \begin{array}{cc} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_1}{\partial z_2} \\ \frac{\partial y_2}{\partial z_1} & \frac{\partial y_2}{\partial z_2} \end{array} \right| = 1 \cdot \left| \begin{array}{cc} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_1}{\partial z_2} \\ \frac{\partial y_2}{\partial z_1} & \frac{\partial y_2}{\partial z_2} \end{array} \right|$$

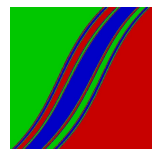


Box-Muller (Continued)

Here we have $y_1 = e^{-\frac{1}{2}(z_1^2+z_2^2)}$ and $y_2 = \frac{1}{2\pi}\ \arctan\left(\frac{z_2}{z_1}\right)$

$$\textit{Hence} \left| \begin{array}{cc} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_1}{\partial z_2} \\ \frac{\partial y_2}{\partial z_1} & \frac{\partial y_2}{\partial z_2} \end{array} \right| = \left(\frac{1}{\sqrt{2\pi}}\ e^{-z_1^2/2}\right)\left(\frac{1}{\sqrt{2\pi}}\ e^{-z_2^2/2}\right)$$

Bivariate density is the product of two $N(0,1)$ densities, so we have generated two independent $N(0,1)$ deviates.



Generating Deviates Derived from N(0,1)

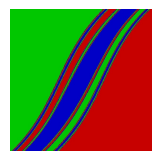
$$N(\mu, \sigma^2) = \mu + \sigma\, N(0,1)$$

$$\chi^2_1 = [N(0,1)]^2$$

$$\chi^2_d = \sum_{i=1}^{d} [N_i(0,1)]^2, \text{ where the } N_i(0,1) \text{ are independent}$$

$$t_d = N(0,1)/\sqrt{x^2_d/d}, \text{ where } N(0,1) \text{ and } \chi^2_d \text{ are independent}$$

$$F_{d_1,d_2} = \chi^2_{d_1}/d_1/\chi^2_{d_2}/d_2 \text{ where } \chi^2_{d_1} \text{ and } \chi^2_{d_2} \text{ are independent}$$



Multivariate Normal

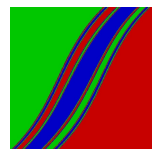
$N(0, I)$ (N-dimensional) – Just stack N $N(0, 1)$'s

$N(\mu, \Sigma)$ (N-dimensional)

Let $PP' = \Sigma$ (P is the Cholesky factor of Σ)

Let $y \sim N(0, I)$. Then $Py \sim N(0, \Sigma)$

To sample from $N(\mu, \Sigma)$, take $\mu + Py$



Simulating Time Series Processes

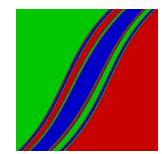
1. Nonparametric: Exact realization via Cholesky factorization of desired covariance matrix. One need only specify the autocovariances.

2. Parametric I: Exact realization via Cholesky factorization of covariance matrix corresponding to desired parametric model.

3. Parametric II: Approximate realization via arbitrary startup value. Burn in before sampling.

4. Parametric III: Exact realization via drawing startup values from stationary distribution.

Note: $VAR(1)$ simulation is key (state transition dynamics).



Accept-Reject Methods (Simple Example)

We want to sample from $f(y)$

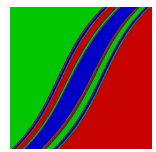
Draw:

$$\nu_1 \sim U(\alpha, \beta)$$

$$\nu_2 \sim U(0, h)$$

If ν_1, ν_2 lies under the density $f(y)$, then take $y = \nu_1$

Otherwise reject and repeat



Graphical Representation of Simple Accept-Reject

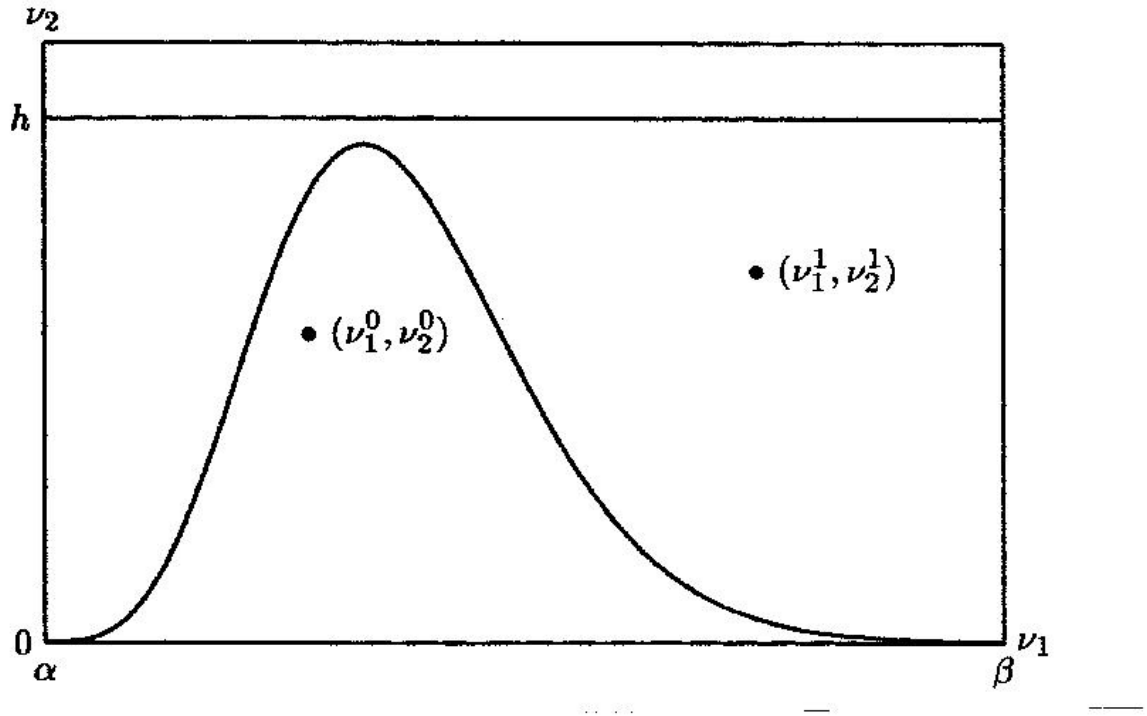
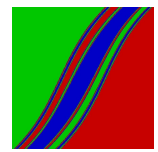


Figure: Simple Accept-Reject



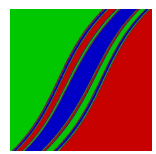
General Accept-Reject

We want to draw S values of y from $f(y)$, but we only know how to sample from $g(y)$.

Let M satisfy $\frac{f(y)}{g(y)} \leq M < \infty$, $\forall y$. Then:

1. Draw proposal $y^* \sim g(y)$.
2. Accept $y = y^*$ *w.p.* $\frac{f(y^*)}{g(y^*)M}$
3. Repeat until S proposals have been kept.

Note that the S draws so-generated are iid.



Markov Chain Monte Carlo (MCMC) Methods

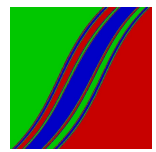
– Construct a Markov chain whose steady-state distribution is the distribution we want to sample from

– Run ("burn in") the chain to convergence and start sampling

Question: How might you assess convergence to steady state?

– Note that the draws so-generated will be serially correlated

Question: For what does the serial correlation "matter"? For what does it not matter?



MCMC I: Metropolis-Hastings Sampling

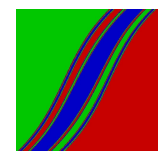
We want to draw S values of y from $f(y)$.

1. Draw y^* from proposal density $q(y; y^{(s-1)})$
2. Calculate the acceptance probability $\alpha(y^{(s-1)}, y^*)$
3. Set

$$y^s = \begin{cases} y^* \text{ w.p. } \alpha(y^{(s-1)}, y^*) & \text{“accept”} \\ y^{(s-1)} \text{ w.p. } 1 - \alpha(y^{(s-1)}, y^*) & \text{“reject”} \end{cases}$$

4. Repeat 1-3, $s = 1, ..., S$

Questions: What q to use in step 1, and what α to use in step 2?



Metropolis-Hastings "Independence Chain"

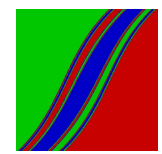
Draw proposal y^* from fixed proposal density $q^*(y)$.
So $q(y; y^{(s-1)})$ is just fixed at $q^*(y)$.

Acceptance probability that does the trick:

$$\alpha(y^{(s-1)}, y^*) = min\left[\frac{f(y = y^*)\ q^*\left(y = y^{(s-1)}\right)}{f(y = y^{(s-1)})\ q^*\left(y = y^*\right)},\ 1\right]$$

$$= min\left[\frac{\frac{f(y=y^*)}{f(y=y^{(s-1)})}}{\frac{q^*(y=y^*)}{q^*\left(y=y^{(s-1)}\right)}},\ 1\right].$$

If improvement in target greater than improvement in proposal, go there w.p.=1. Else go there w.p.<1.



Metropolis-Hastings "Random Walk Chain"

Random walk proposals:

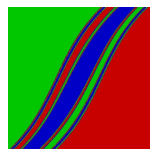
$$y^* = y^{(s-1)} + \varepsilon$$

ε is called the increment random variable. Commonly we draw from $\varepsilon \sim N(0, c^2)$, where c is chosen to ensure a "reasonable" acceptance rate (around 30%, say).

Then $q(y; y^{(s-1)})$ is $N(y^{(s-1)}, c^2)$.

Acceptance probability that does the trick:

$$\alpha(y^{(s-1)}, y^*) = min\left[\frac{f(y = y^*)}{f(y = y^{(s-1)})}, 1\right]$$

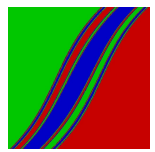


A Useful Property of Metropolis-Hastings

Metropolis requires evaluating only the kernel of the density of interest, because the acceptance probability is governed by the *ratio*

$$\frac{f(y = y^*)}{f(y = y^{(s-1)})}.$$

– This constrasts with generic accept-reject, which requires evaluation of the full density $f(y)$ (not just its kernel)



MCMC II: Gibbs Sampling

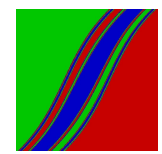
Consider first the bivariate case. We want to sample from $f(y) = f(y_1, y_2)$. Initialize ($j = 0$) using y_2^0.

Gibbs iteration $j = 1$:
a. Draw y_1^1 from $f(y_1|y_2^0)$
b. Draw y_2^1 from $f(y_2|y_1^1)$

Repeat $j = 2,\ 3,\ ....$

Notice that $\{y^1,\ y^2,\ ...\}$ is a Markov chain. Under very general conditions its stationary distribution is $f(y)$.

Useful if conditionals are known and "easy" to sample from, but joint/marginals are not. (This often happens in Bayesian analysis.)



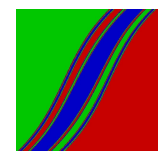
General Gibbs Sampling

We want to sample from $f(y) = f(y_1, y_2, ..., y_k)$

Initialize $(j = 0)$ using $y_2^0, y_3^0, ..., y_k^0$

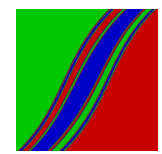
Gibbs iteration $j = 1$:
a. Draw y_1^1 from $f(y_1|y_2^0, ..., y_k^0)$
b. Draw y_2^1 from $f(y_2|y_1^1, y_3^0, ..., y_k^0)$
c. Draw y_3^1 from $f(y_3|y_1^1, y_2^1, y_4^0, ..., y_k^0)$
...
k. Draw y_k^1 from $f(y_k|y_1^1, ..., y_{k-1}^1)$

Repeat $j = 2,\ 3, ...$

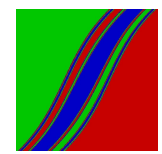


More Simulation:

Bayesian Analysis with Markov Chain Monte Carlo



I. Basic Issues



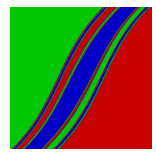
Frequentist vs. Bayesian Paradigms

Frequentist: $\hat{\theta}$ random, θ fixed. $\sqrt{T}(\hat{\theta}_{ML} - \theta) \rightarrow_d N$

Bayesian: $\hat{\theta}$ fixed, θ random. $\sqrt{T}(\theta - \hat{\theta}_{ML}) \rightarrow_d N$

Frequentist: Characterize the distribution of the random data ($\hat{\theta}$) conditional on fixed "true" θ. Focus on the likelihood max ($\hat{\theta}_{ML}$) and likelihood curvature in an ϵ-neighborhood of the max.

Bayesian: Characterize the distribution of the random θ conditional on fixed data ($\hat{\theta}$). Examine the entire likelihood.



Some Bayesian Pros and Cons (We Could Lengthen Both Lists...)

Pros:

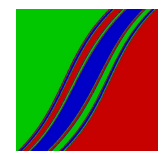
1. Feels sensible to focus on $p(\theta/y)$. Relative frequency in repeated samples replaced with subjective degree of belief conditional on the single sample actually obtained

2. Exact finite-sample full-density inference

Cons:

1. From where does the prior come? How to elicit prior distributions? Very difficult in all but the most trivial cases.

2. How to do an "objective" analysis? What is an "uninformative" prior? Uniform is definitely *not* uninformative...



Bayesian Computational Mechanics

Data $y \equiv \{y_1, \ldots, y_T\}$

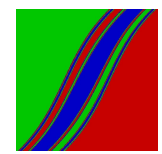
Bayes' Theorem:

$$f(\theta/y) = \frac{f(y/\theta)f(\theta)}{f(y)}$$

or

$$f(\theta/y) = c\ f(y/\theta)f(\theta)$$
where $c^{-1} = \int_\theta f(y/\theta)f(\theta)$

$$f(\theta/y) \propto f(y/\theta)f(\theta)$$
$$p(\theta/y) \propto L(\theta/y)g(\theta)$$
posterior $\propto$ likelihood $\cdot$ prior



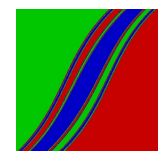
Bayesian Estimation

Full posterior density

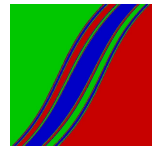
Highest posterior density intervals

Posterior mean, median, mode (depending on loss function)

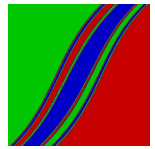
How to get the posterior density of interest?
(We will see...)



II. Bayesian Analysis of State-Space Models



A. First do Bayes for Gaussian Regression



Model and Standard Results

$$y = X\beta + \varepsilon$$
$$\varepsilon \sim iid\ N(0, \sigma^2 I)$$

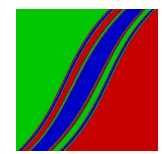
Recall the standard (frequentist) results:

$$\hat{\beta}_{ML} = (X'X)^{-1}X'y$$

$$\hat{\sigma}^2_{ML} = \frac{e'e}{T}$$

$$\hat{\beta}_{ML} \sim N\left(\beta, \sigma^2(X'X)^{-1}\right)$$

$$\frac{T\hat{\sigma}^2_{ML}}{\sigma^2} \sim \chi^2_{T-K}$$



Bayesian Inference for β/σ^2

Conjugate prior:

$$\beta \sim N(\beta_0, \Sigma_0)$$

$$g(\beta) \propto exp(-1/2(\beta - \beta_0)'\Sigma_0^{-1}(\beta - \beta_0))$$

Likelihood:

$$L(\beta/\sigma^2, y) \propto \left(\sigma^2\right)^{-T/2} exp(\tfrac{-1}{2\sigma^2}(y - X\beta)'(y - X\beta))$$

Posterior:

$$p(\beta/\sigma^2, y) \propto$$

$$exp(-1/2(\beta - \beta_0)'\Sigma_0^{-1}(\beta - \beta_0) - \tfrac{1}{2\sigma^2}(y - X\beta)'(y - X\beta))$$

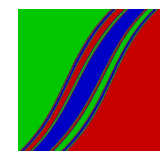
This is the kernel of a normal distribution:

$$\beta/\sigma^2, y \sim N(\beta_1, \Sigma_1)$$

where

$$\beta_1 = \Sigma_1(\Sigma_0^{-1}\beta_0 + \sigma^{-2}(X'X)\hat{\beta}_{ML})$$

$$\Sigma_1 = (\Sigma_0^{-1} + \sigma^{-2}(X'X))^{-1}$$



Gamma and Inverse Gamma Refresher

$$z_t \overset{iid}{\sim} N\left(0, \frac{1}{\delta}\right), \; x = \sum_{t=1}^{v} z_t^2 \;\Rightarrow\; x \sim \Gamma\left(x; \frac{v}{2}, \frac{\delta}{2}\right)$$

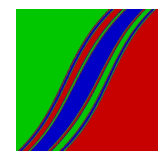
(Note $\delta = 1 \Rightarrow x \sim \chi^2_v$, so χ^2 is a special case of Γ)

$$\Gamma\left(x; \frac{v}{2}, \frac{\delta}{2}\right) \propto x^{\frac{v}{2}-1} exp\left(\frac{-x\delta}{2}\right)$$

$$E(x) = \frac{v}{\delta}$$

$$var(x) = \frac{2v}{\delta^2}$$

$x \sim \Gamma^{-1}(\frac{v}{2}, \frac{\delta}{2})$ ("inverse gamma") $\Leftrightarrow \; \frac{1}{x} \sim \Gamma(\frac{v}{2}, \frac{\delta}{2})$



Bayesian Inference for σ^2/β

Conjugate prior:

$$\frac{1}{\sigma^2} \sim \Gamma\left(\frac{v_0}{2}, \frac{\delta_0}{2}\right)$$

$$g\left(\frac{1}{\sigma^2}\right) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{v_0}{2}-1} exp\left(-\frac{\delta_0}{2\sigma^2}\right)$$

Likelihood:

$$L\left(\frac{1}{\sigma^2}/\beta, y\right) \propto \left(\sigma^2\right)^{-T/2} exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right)$$

Posterior:

$$p\left(\frac{1}{\sigma^2}/\beta, y\right) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{v_1}{2}-1} exp\left(\frac{-\delta_1}{2\sigma^2}\right)$$

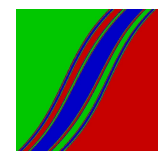
This is the kernel of a gamma distribution:

$$\frac{1}{\sigma^2}/\beta, y \sim \Gamma\left(\frac{v_1}{2}, \frac{\delta_1}{2}\right)$$

where

$$v_1 = v_0 + T$$

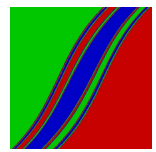
$$\delta_1 = \delta_0 + (y - X\beta)'(y - X\beta)$$



The Key Issue/Question

- We have the *conditional* posterior distributions ($p(\beta/\sigma^2, y)$, $p\left(\sigma^2/\beta, y\right)$), but they're not what we want.
- How do we get the *unconditional* (joint and marginal) posterior distributions that we *really* want: $p(\beta, \sigma^2/y)$, $p(\beta/y)$, $p(\sigma^2/y)$?

MCMC!



Gibbs Sampling from the Joint Posterior

0. Initialize: $\sigma^2 = (\sigma^2)^{(0)}$

Gibbs sampler at generic iteration j:

$j1$. Draw $\beta^{(j)}$ from $p(\beta^{(j)}/(\sigma^2)^{(j-1)}, y)$ $\quad (N(\beta_1, \Sigma_1))$

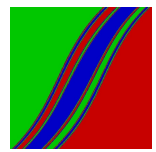
$j2$. Draw $(\sigma^2)^{(j)}$ from $p(\sigma^2/\beta^{(j)}, y)$ $\quad \left(\Gamma^{-1}\left(\frac{\nu_1}{2}, \frac{\delta_1}{2}\right)\right)$

Iterate to convergence (steady state of the Markov chain), and then estimate posterior moments of interest

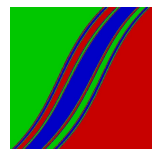
– Sample mean converges appropriately despite the serial correlation in the Markov chain

– Assessing precision requires robust s.e.'s (based on spectrum at frequency zero) due to the serial correlation in the Markov chain.

$$\left(\text{Recall}: \quad \sqrt{\mathrm{T}}(\bar{\mathrm{x}} - \mu) \xrightarrow{d} \mathrm{N}(0,\ \mathrm{g}(0))\right)$$



B. Now Move to Bayesian Analysis of General State-Space Models



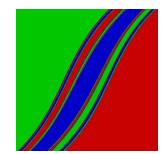
Recall the State-Space Model

$$\alpha_t = T\alpha_{t-1} + R\eta_t$$

$$y_t = Z\alpha_t + \zeta$$

$$\begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix} \overset{iid}{\sim} N\begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix}$$

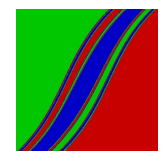
Let $\tilde{\alpha}_T = (\alpha_1', \ldots, \alpha_T')'$,
and collect all system parameters into a vector θ.



Recall the State-Space Model in Density Form

$$\alpha_t|\alpha_{t-1} \sim N(T\alpha_{t-1},\ RQR')$$

$$y_t|\alpha_t \sim N(Z\alpha_t,\ H)$$



Recall the Kalman Filter in Density Form

Initialize at a_0, P_0

State prediction:

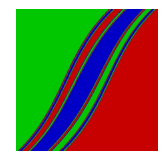
$$\alpha_t|\tilde{y}_{t-1} \sim N(a_{t/t-1},\ P_{t/t-1})$$
$$a_{t/t-1} = Ta_{t-1}$$
$$P_{t/t-1} = TP_{t-1}T' + RQR'$$

State update:

$$\alpha_t|\tilde{y}_t \sim N(a_t,\ P_t)$$
$$a_t = a_{t/t-1} + K_t(y_t - Za_{t/t-1})$$
$$P_t = P_{t/t-1} - K_tZP_{t/t-1}$$

Data prediction:

$$y_t|\tilde{y}_{t-1} \sim N(Za_{t/t-1},\ F_t)$$



Recall the (Essence of the) EM Algorithm for State-Space Models

1. E Step:

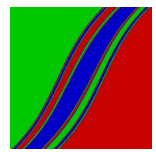
Approximate a "complete data" situation by replacing $\{\alpha_t\}_{t=0}^{T}$ with $\{a_{t,T}\}_{t=0}^{T}$ from the Kalman smoother

2. M Step:

Estimate parameters by running regressions:

$$a_{t,T} \rightarrow a_{t-1,T}$$
$$y_t \rightarrow a_{t,T}$$

3. If convergence criterion not met, go to 1



Recall Multivariate Regression

$$y_{it} = x_t' \beta^i + \varepsilon_{it}$$

$$(\epsilon_{1,t}, ...\epsilon_{N,t})' \overset{iid}{\sim} N(0, \Sigma)$$

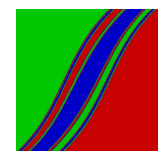
$$i = 1, ..., N$$

$$t = 1, ...T$$

or

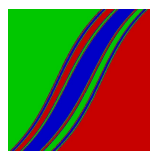
$$\underbrace{Y}_{T \times N} = \underbrace{X}_{T \times K} \underbrace{B}_{K \times N} + \underbrace{E}_{T \times N}$$

OLS is still $(X'X)^{-1}X'Y$



The Key to Moving Forward...

- Treat the vector $\tilde{\alpha}_T$ as a *parameter*, along with system matrices θ
- Use Gibbs to draw from posterior $\tilde{\alpha}_T, \theta \,/\, \tilde{y}_T$ by iterating on $\tilde{\alpha}_T \,/\, \theta, \tilde{y}_T$ and $\theta \,/\, \tilde{\alpha}_T, \tilde{y}_T$
- Note that we draw from two large blocks $\tilde{\alpha}_T \,/\, \theta, \tilde{y}_T$ (one draw) and $\theta \,/\, \tilde{\alpha}_T, \tilde{y}_T$ (one draw) "Multi-move Gibbs sampler"
- Massively more efficient than cycling through a "one at a time" Gibbs sampler



Carter-Kohn "Multimove" Gibbs Sampler

Let $\tilde{y}_T = (y_1', \ldots, y_T')'$

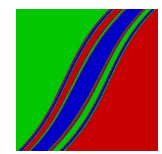
0. Initialize $\theta^{(0)}$

Gibbs sampler at generic iteration j:

$j1$. Draw from posterior $\tilde{\alpha}_T^{(j)} / \theta^{(j-1)}, \tilde{y}_T$ ("hard")

$j2$. Draw from posterior $\theta^{(j)} / \tilde{\alpha}_T^{(j)}, \tilde{y}_T$ ("easy")

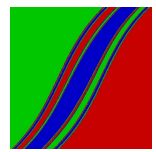
Iterate to convergence, and then estimate posterior moments of interest



Let's First Explain Step $j2$ $(\theta^{(j)}|\tilde{\alpha}_T^{(j)}, \tilde{y}_T)$ ("easy")

Conditional upon an $\tilde{\alpha}_T^{(j)}$ draw, obtaining a $\theta^{(j)}$ draw is just a Bayesian multivariate regression problem (i.e., we need to draw from the posterior of the multivariate regression parameters).

(We have already seen how to do Bayesian univariate regression. That is, we know how to draw from the posterior of univariate regression parameters using a conjugate (normal-gamma) prior. We can easily extend to draw from the posterior of multivariate regression parameters using a conjugate (normal-Wishart) prior.)



Now do Step $j1$ ($\tilde{\alpha}_T^{(j)}/\theta^{(j-1)}, \tilde{y}_T$) ("hard")

For notational simplicity we write $p(\tilde{\alpha}_T/\tilde{y}_T)$, suppressing the conditioning on θ, but it is of course still there.

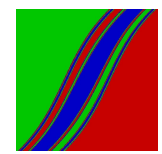
$$p(\tilde{\alpha}_T/\tilde{y}_T) = p(\alpha_T/\tilde{y}_T)\ p(\tilde{\alpha}_{T-1}/\alpha_T, \tilde{y}_T)$$

$$= p(\alpha_T/\tilde{y}_T)\ p(\alpha_{T-1}/\alpha_T, \tilde{y}_T)\ p(\tilde{\alpha}_{T-2}/\alpha_{T-1}, \alpha_T, \tilde{y}_T)$$

$$= \ldots$$

$$= p(\alpha_T/\tilde{y}_T)\ \Pi_{t=1}^{(T-1)} p(\alpha_t/\alpha_{t+1}, \tilde{y}_t)$$

So, to draw from $p(\tilde{\alpha}_T/\tilde{y}_T)$, we need to be able to draw from $p(\alpha_T/\tilde{y}_T)$ and $p(\alpha_t/\alpha_{t+1}, \tilde{y}_t), t = 1, \ldots, (T-1)$



Multimove Gibbs sampler, Continued

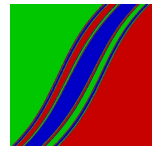
The key is to *work backward*:

Draw from $p(\alpha_T/\tilde{y}_T)$,
then from $p(\alpha_{T-1}/\alpha_T, \tilde{y}_{T-1})$,
then from $p(\alpha_{T-2}/\alpha_{T-1}, \tilde{y}_{T-2})$,
etc.

Time T draw is easy. From the Kalman filter,
$p(\alpha_T/\tilde{y}_T)$ is $N(a_T, P_T)$
(where we have the usual formulas for a_T and P_T)

Earlier-time draws require a bit of new work:
How to draw from $p(\alpha_t/\alpha_{t+1}, \tilde{y}_t)$, $t = (T-1), ..., 1$?

– Note that the CK smoother requires first running the Kalman filter. Prior info for $\tilde{\alpha}_T$ enters through choice of a_0 and P_0.



Multimove Gibbs sampler, Continued

CK show that the posterior density $p(\alpha_t/\alpha_{t+1}, \tilde{y}_t)$ is Gaussian:

$$\alpha_t/\alpha_{t+1}, \tilde{y}_t \sim N(a_{t/t,\alpha_{t+1}}, P_{t/t,\alpha_{t+1}})$$

where

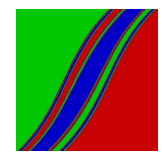
$$a_{t/t,\alpha_{t+1}} = E(\alpha_t/\tilde{y}_t, \alpha_{t+1}) = E(\alpha_t|a_t, \alpha_{t+1})$$

$$= a_t + P_t T'(TP_t T' + Q)^{-1}(\alpha_{t+1} - Ta_t)$$

$$P_{t/t,\alpha_{t+1}} = cov(\alpha_t/\tilde{y}_t, \alpha_{t+1}) = cov(\alpha_t|a_t, \alpha_{t+1})$$

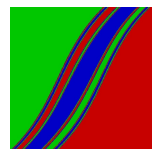
$$= P_t - P_t T'(TP_t T' + Q)^{-1} TP_t$$

$$t = (T-1), ..., 1.$$



Remarks

- Multi-move Gibbs resembles EM for state-space models, iterating between an "α step" and a "θ step"
- More than a superficial resemblance; both explore likelihood
  - EM explores likelihood as part of getting to a max
  - Multi-move Gibbs explores likelihood as part of exploring a posterior



Non-linear and/or Non-Gaussian

This small block is just to stimulate curiosity, and to introduce some non-linear and/or non-Gaussian models.

(Take 722.)

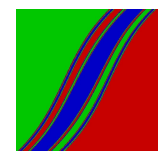
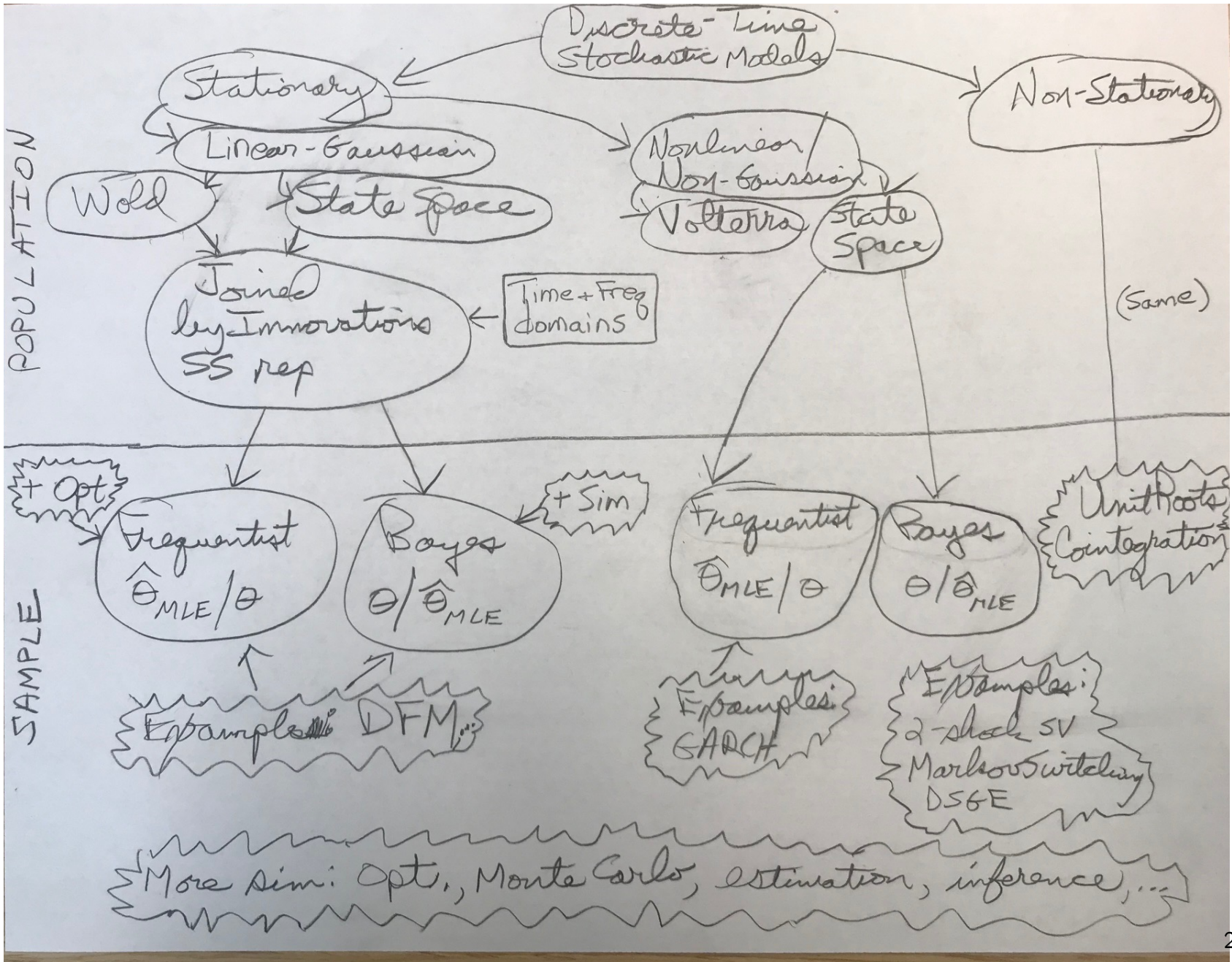


Figure: A Map


Discrete-Time Stochastic Models
Stationary
Non-Stationary
Linear-Gaussian
Nonlinear/Non-Gaussian
Wold
State Space
Volterra
State Space
POPULATION
Joined by Innovations SS rep
Time + Freq domains
(Same)
+ Opt
+ Sim
Frequentist $\hat{\theta}_{MLE}/\theta$
Bayes $\theta/\hat{\theta}_{MLE}$
Frequentist $\hat{\theta}_{MLE}/\theta$
Bayes $\theta/\hat{\theta}_{MLE}$
Unit Roots Cointegration
SAMPLE
Examples: DFM...
Examples: GARCH
Examples: 2-shock SV Markov Switching DSGE
More sim: Opt., Monte Carlo, estimation, inference, ...

Earlier: Wold-Wiener-Kolmogorov Representation/Theory
Now: Volterra Representation/Theory

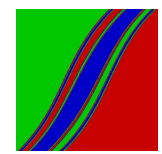
$$y_t = \sum_{m_1=0}^{\infty} h_1(m_1)\varepsilon_{t-m_1}$$

$$+ \sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty} h_2(m_1, m_2)\varepsilon_{t-m_1}\varepsilon_{t-m_2}$$

$$+ \sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty}\sum_{m_3=0}^{\infty} h_3(m_1, m_2, m_3)\varepsilon_{t-m_1}\varepsilon_{t-m_2}\varepsilon_{t-m_3}$$

$$+ \ldots$$

where $\varepsilon_t \sim N(0, 1)$



State Space Models

Linear / Gaussian:

$$\alpha_t = T\alpha_{t-1} + R\eta_t, \quad y_t = Z\alpha_t + \zeta_t, \quad \eta_t \sim N(0, Q), \quad \zeta_t \sim N(0, H)$$

Linear / Non-Gaussian:

$$\alpha_t = T\alpha_{t-1} + R\eta_t, \quad y_t = Z\alpha_t + \zeta_t, \quad \eta_t \sim D^{\eta}, \quad \zeta_t \sim D^{\zeta}$$

Non-Linear / Gaussian:

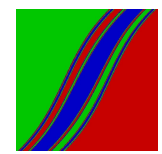
$$\alpha_t = Q(\alpha_{t-1}, \eta_t), \quad y_t = G(\alpha_t, \zeta_t), \quad \eta_t \sim N(0, Q), \quad \zeta_t \sim N(0, H)$$

Non-Linear / Non-Gaussian:

$$\alpha_t = Q(\alpha_{t-1}, \eta_t), \quad y_t = G(\alpha_t, \zeta_t), \quad \eta_t \sim D^{\eta}, \quad \zeta_t \sim D^{\zeta}$$

Non-Linear / Non-Gaussian, Specialized to Additive:

$$\alpha_t = Q(\alpha_{t-1}) + \eta_t, \quad y_t = G(\alpha_t) + \zeta_t, \quad \eta_t \sim D^{\eta}, \quad \zeta_t \sim D^{\zeta}$$



Example:
Regime Switching Model (Non-Linear / Gaussian)

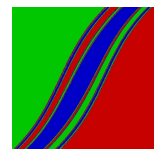
$$\begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix} = \begin{pmatrix} \phi & 0 \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

$$y_t = \mu_0 + \delta\, I(\alpha_{2t} > 0) + (1, 0) \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix}$$

$$\eta_{1t} \sim N^{\eta_1} \quad \eta_{2t} \sim N^{\eta_2} \quad \eta_{1t} \perp \eta_{2t}$$

Extensions to:

– Richer α_1 dynamics (governing the observed y)
– Richer α_2 dynamics (governing the latent regime)
– Richer η_t distribution (e.g., η_{2t} asymmetric)
– More than two states
– Switching also on dynamic parameters, volatilities, etc.
– Multivariate

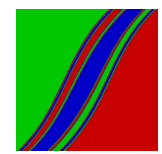


Example:
Two-Shock Stochastic Volatility Model
(Non-Linear / Gaussian Form)

$$h_t = \omega + \beta h_{t-1} + \eta_t \quad \text{(transition)}$$

$$r_t = \sqrt{e^{h_t}}\, \zeta_t \quad \text{(measurement)}$$

$$\eta_t \sim N(0, \sigma_\eta^2), \qquad \zeta_t \sim N(0, 1)$$



Example:
Two-Shock Stochastic Volatility Model
(Linear / Non-Gaussian Form)

$$h_t = \omega + \beta h_{t-1} + \eta_t \quad \text{(transition)}$$

$$2ln|r_t| = h_t + 2ln|\zeta_t| \quad \text{(measurement)}$$

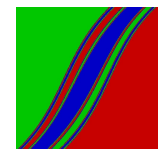
or

$$h_t = \omega + \beta h_{t-1} + \eta_t$$

$$y_t = h_t + u_t$$

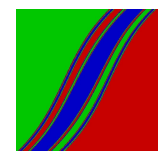
$$\eta_t \sim N(0, \sigma_\eta^2), \quad u_t \sim D^u$$

– A "signal plus (non-Gaussian) noise" components model



Example:
One-Shock Stochastic Volatility Model (GARCH)

Coming soon, in detail...

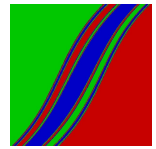


Nonlinear and/or Non-Gaussian Filtering

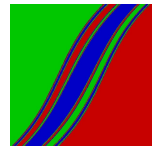
– Kalman filter BLUE but not MVUE

– Fully-optimal filtering gets complicated
Different filters (e.g., particle filter) needed for full optimality

(Again, take 722.)



One-Shock Stochastic Volatility Model (ARCH/GARCH)



Stock Returns

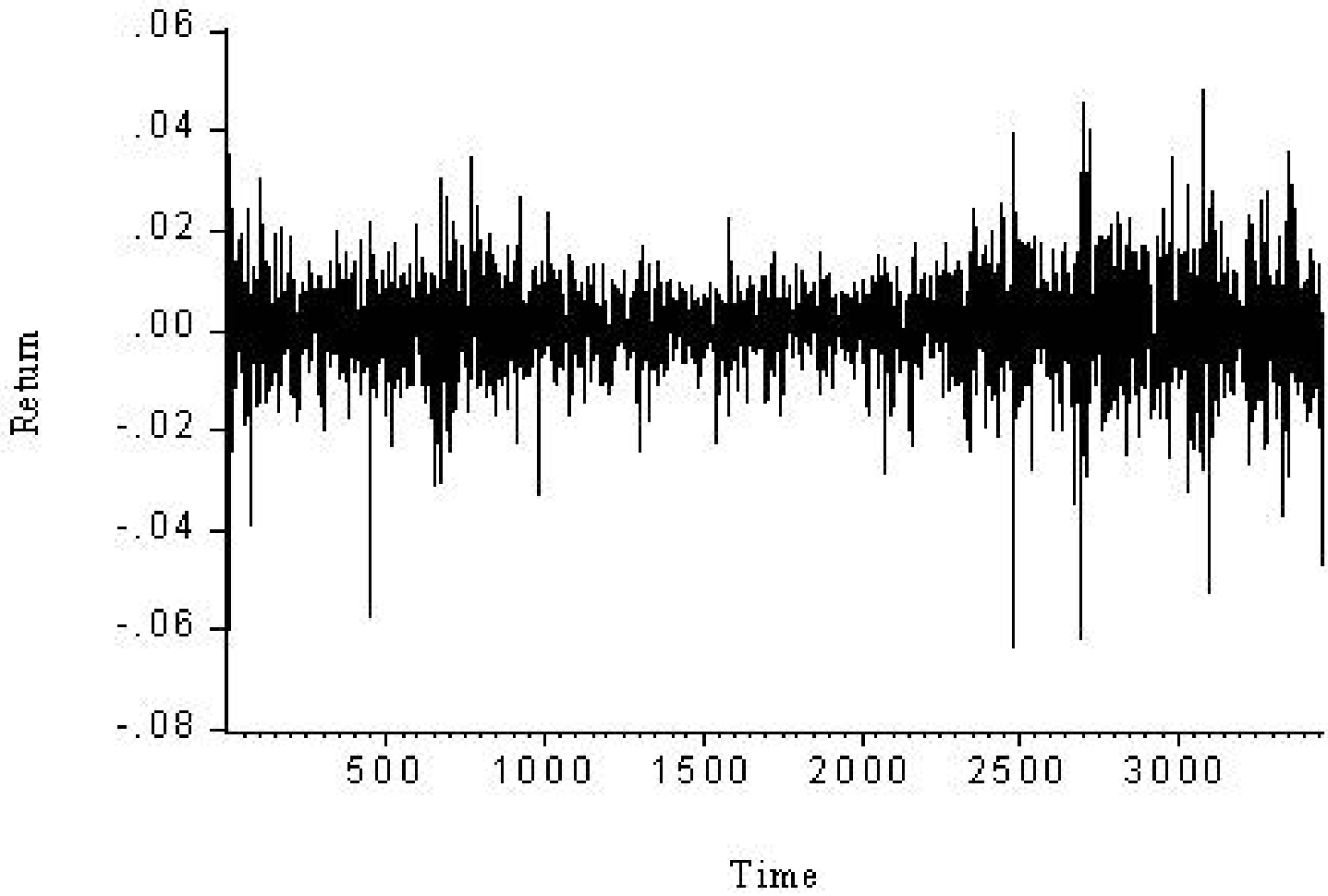
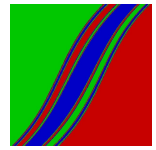


Figure: Time Series of Daily NYSE Returns.



Key Fact 1: Returns are Approximately White Noise

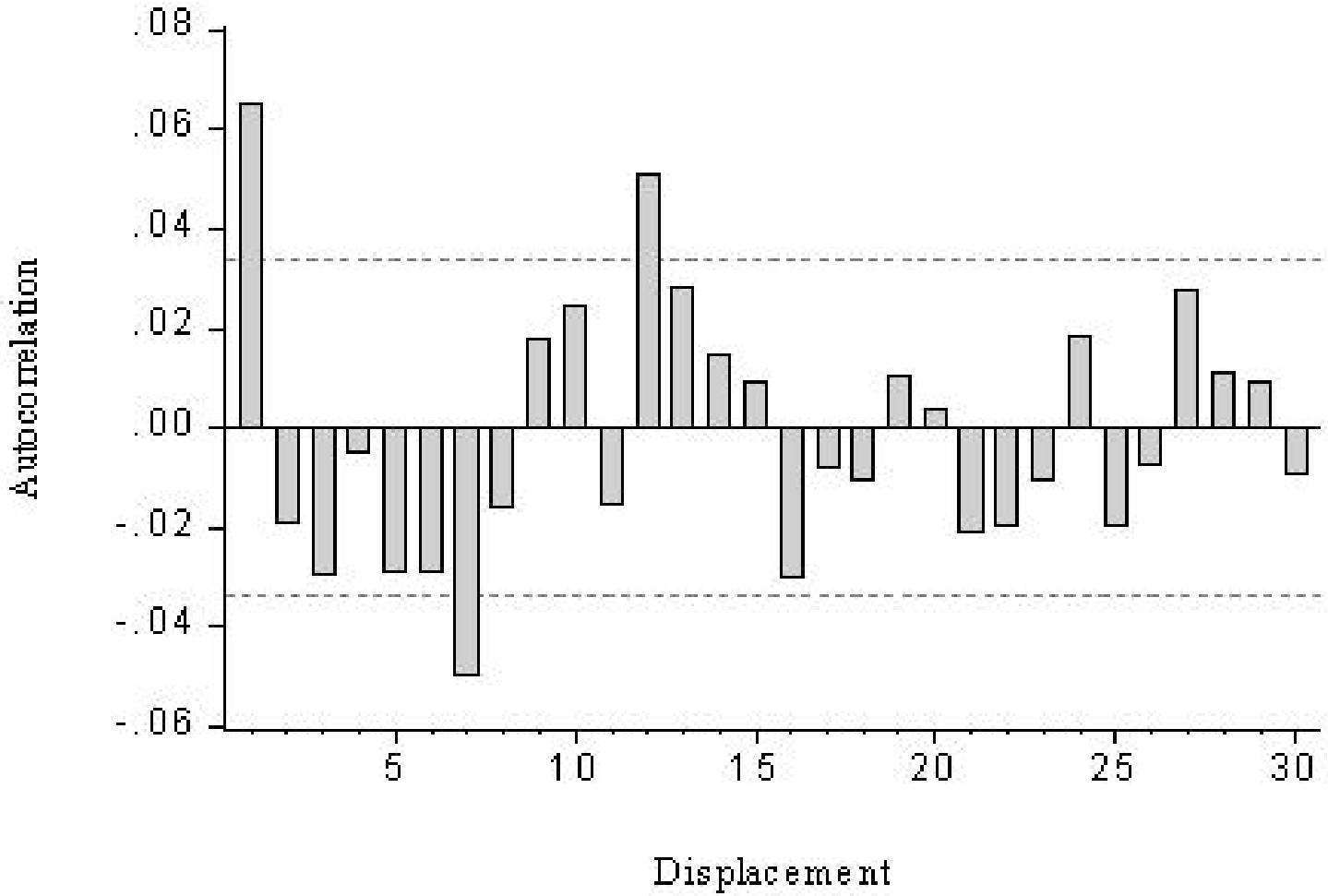
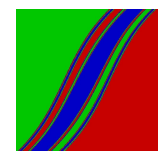


Figure: Correlogram of Daily NYSE Returns.



Key Fact 2: Returns are not Gaussian

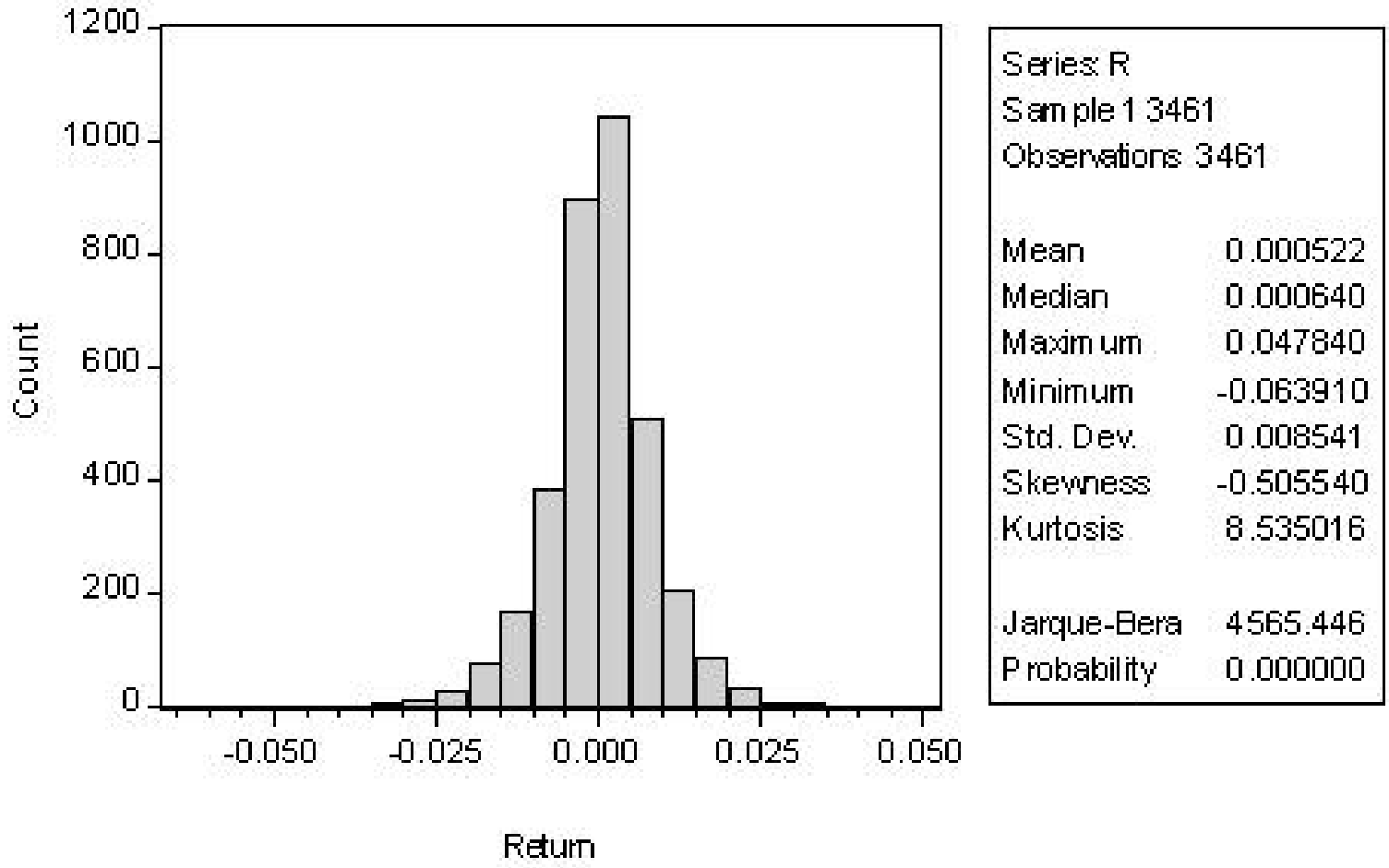
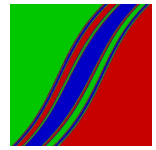


Figure: Histogram and Statistics for Daily NYSE Returns.



Unconditional Volatility Measures

Unconditional Variance: $\sigma^2 = E(r_t - \mu)^2$
(or standard deviation: σ)

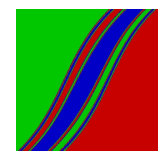
Unconditional Mean Absolute Deviation: $MAD = E|r_t - \mu|$

Unconditional Interquartile Range: $IQR = 75\% - 25\%$

Unconditional $p\%$ Value at Risk (VaR^p)): x s.t. $P(r_t < x) = p$

Unconditional Outlier probability: $P|r_t - \mu| > 5\sigma$ (for example)

Unconditional Tail index: γ s.t. $P(r_t > r) = kr^{-\gamma}$



Key Fact 3: Returns are Conditionally Heteroskedastic (And the Volatility Dynamics are Highly Persistent)

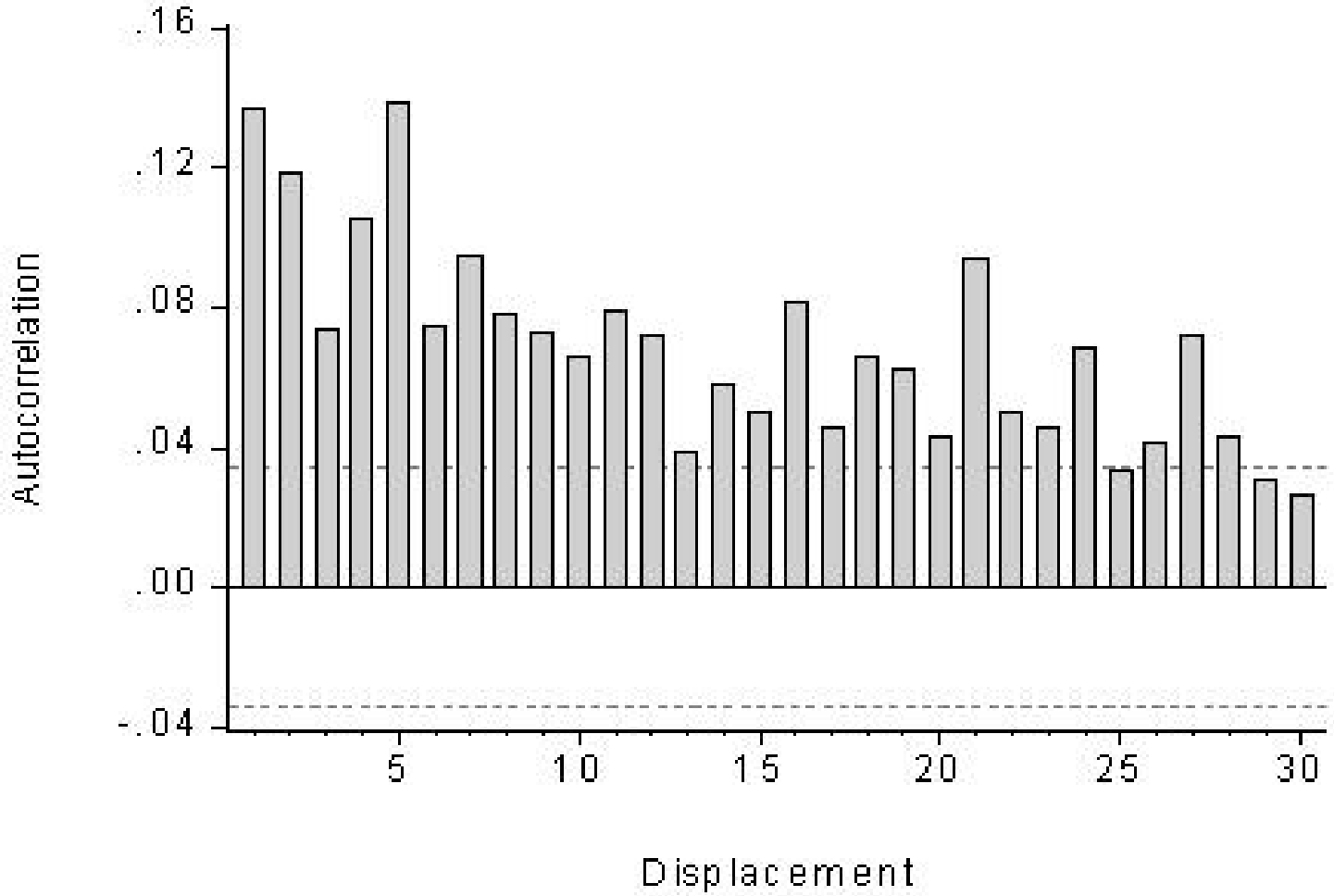


Figure: Correlogram of Daily Squared NYSE Returns.

Conditional Return Distributions

Consider $f(r_t)$ *vs.* $f(r_t|\Omega_{t-1})$

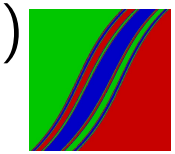
Key 1: $E(r_t|\Omega_{t-1})$ is approximately constant

Returns are approximately serially uncorrelated, and approximately free of additional non-linear conditional mean dependence.

Key 2: $var(r_t|\Omega_{t-1})$ is *not* constant!

Squared returns serially correlated, often with very slow decay.

Bottom line: Returns are (approximately) non-Gaussian weak white noise, serially uncorrelated but nevertheless serially dependent, with the non-linear serial dependence operating not through $E(r_t|\Omega_{t-1})$ but rather through $var(r_t|\Omega_{t-1})$ ("*volatility dynamics*").



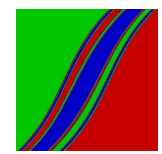
The Standard Model
(Linearly Indeterministic Process with *iid* Innovations)

$$y_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

$$\varepsilon \sim iid(0,\ \sigma_\varepsilon^2)$$

$$\sum_{i=0}^{\infty} b_i^2 < \infty, \quad b_0 = 1$$

Uncond. mean: $E(y_t) = 0$ (constant)
Uncond. variance: $E(y_t - E(y_t))^2 = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} b_i^2$ (constant)
Cond. mean: $E(y_t \mid \Omega_{t-1}) = \sum_{i=1}^{\infty} b_i \varepsilon_{t-i}$ (varies)
Cond. variance: $E([y_t - E(y_t \mid \Omega_{t-1})]^2 \mid \Omega_{t-1}) = \sigma_\varepsilon^2$ (constant)



The Standard Model, Continued
h-Step-Ahead Least Squares Forecasting

$$E(y_{t+h} \mid \Omega_t) = \sum_{i=0}^{\infty} b_{h+i}\varepsilon_{t-i}$$

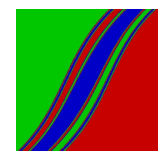
Associated prediction error:

$$y_{t+h} - E(y_{t+h} \mid \Omega_t) = \sum_{i=0}^{h-1} b_i\varepsilon_{t+h-i}$$

Conditional prediction error variance:

$$E([y_{t+h} - E(y_{t+h} \mid \Omega_t)]^2 \mid \Omega_t) = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} b_i^2$$

Key: Depends only on h, not on Ω_t



$ARCH(1)$ Process

$$r_t|\Omega_{t-1} \sim N(0, h_t)$$

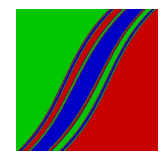
$$h_t = \omega + \alpha r_{t-1}^2$$

$$E(r_t) = 0$$

$$E(r_t - E(r_t))^2 = \frac{\omega}{(1-\alpha)}$$

$$E(r_t|\Omega_{t-1}) = 0$$

$$E([r_t - E(r_t|\Omega_{t-1})]^2|\Omega_{t-1}) = \omega + \alpha r_{t-1}^2$$



Nonlinear State-Space Representation

$$r_t \mid \Omega_{t-1} \sim N(0, \omega + \alpha r_{t-1}^2)$$

or

$$r_t = \sqrt{\omega + \alpha r_{t-1}^2}\, z_t$$

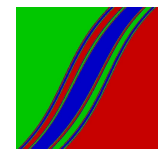
$$z_t \sim iid\ N(0,1)$$

Trivial state-space representation:

$$\alpha_t = \sqrt{\omega + \alpha \alpha_{t-1}^2}\, \eta_t$$

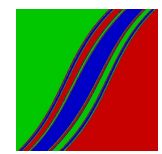
$$r_t = \alpha_t$$

$$\eta_t \sim iid\ N(0,1)$$



Unconditionally Fat Tails

ARCH produces an unconditional distribution that is symmetric but with fatter tails than the conditional (Gaussian) distribution.



Tractable Likelihood

$$L\left(\theta; r_T, \ldots, r_1\right) \approx f\left(r_T|\Omega_{T-1}; \theta\right) f\left(r_{T-1}|\Omega_{T-2}; \theta\right) \ldots f\left(r_2|\Omega_1; \theta\right),$$

where $\theta = (\omega, \alpha)'$.
(It is approximate because we drop the initial marginal $f(r_1; \theta)$).

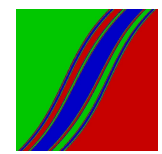
With Gaussian conditional densities,

$$f\left(r_t|\Omega_{t-1}; \theta\right) = \frac{1}{\sqrt{2\pi}} h_t(\theta)^{-1/2} \exp\left(-\frac{1}{2}\frac{r_t^2}{h_t(\theta)}\right),$$

we have:

$$\ln L\left(\theta; r_T, \ldots, r_1\right) \approx const - \frac{1}{2}\sum_{t=2}^{T} \ln h_t(\theta) - \frac{1}{2}\sum_{t=2}^{T} \frac{r_t^2}{h_t(\theta)},$$

where $h_t(\theta) = \omega + \alpha r_{t-1}^2$.



GARCH(1,1) Process

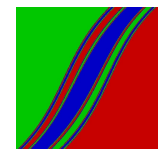
$$r_t \mid \Omega_{t-1} \sim N(0, h_t)$$
$$h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}$$

$$E(r_t) = 0$$
$$E(r_t - E(r_t))^2 = \frac{\omega}{(1-\alpha-\beta)}$$
$$E(r_t|\Omega_{t-1}) = 0$$
$$E([r_t - E(r_t \mid \Omega_{t-1})]^2 \mid \Omega_{t-1}) = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}$$

GARCH(1,1) back substitution yields:

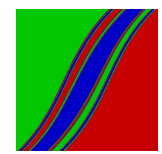
$$h_t = \frac{\omega}{1-\beta} + \alpha \sum_{j=1}^{\infty} \beta^{j-1} r_{t-j}^2$$

$$ARCH(\infty)!$$



Variations on the ARCH/GARCH Theme

- Asymmetric Response and the Leverage Effect
- Fat-Tailed Conditional Densities



Asymmetric Response and the Leverage Effect: Threshold GARCH

Standard GARCH: $h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}$

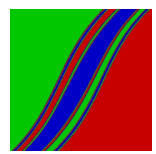
TARCH: $h_t = \omega + \alpha r_{t-1}^2 + \gamma r_{t-1}^2 D_{t-1} + \beta h_{t-1}$

$$D_t = \begin{cases} 1 \ \textit{if} \ \ r_t < 0 \\ 0 \ \textit{otherwise} \end{cases}$$

positive return (good news): α effect on volatility

negative return (bad news): $\alpha + \gamma$ effect on volatility

$\gamma \neq 0$: Asymetric news response
$\gamma > 0$: "Leverage effect"



Fat-Tailed Conditional Densities: t-GARCH

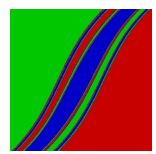
If r is conditionally Gaussian, then $\frac{r_t}{\sqrt{h_t}} \sim N(0,1)$

But often with high-frequency data, $\frac{r_t}{\sqrt{h_t}} \sim$ *fat tailed*

So take:

$$r_t = h_t^{1/2} z_t$$

$$z_t \overset{iid}{\sim} \frac{t_d}{std(t_d)}$$



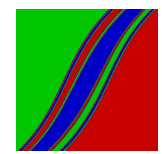
GARCH(1,1) for NYSE Returns

Dependent Variable: R
Method: ML - ARCH (Marquardt)

Sample: 1 3461
Included observations: 3461
Convergence achieved after 19 iterations
Variance backcast: ON

Coefficient	Std. Error	z-Statistic	Prob.	
C	0.000640	0.000127	5.036942	0.0000
Variance Equation				
C	1.06E-06	1.49E-07	7.136840	0.0000
ARCH(1)	0.067410	0.004955	13.60315	0.0000
GARCH(1)	0.919714	0.006122	150.2195	0.0000

Figure: GARCH(1,1) Estimation, Daily NYSE Returns.



Fitted Volatility

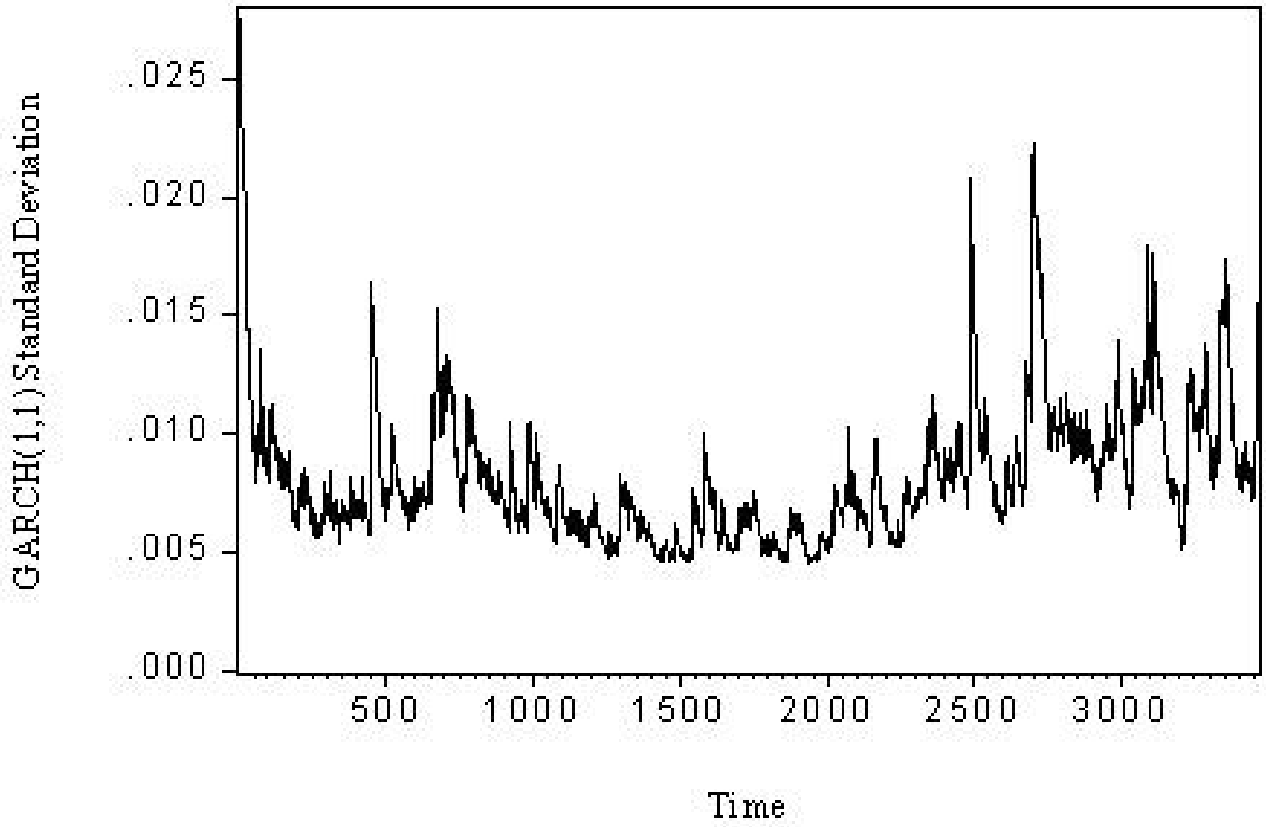
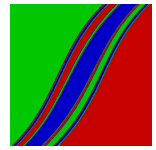


Figure: Estimated Conditional Standard Deviation, Daily NYSE Returns.



Recall: Correlogram of Squared Returns

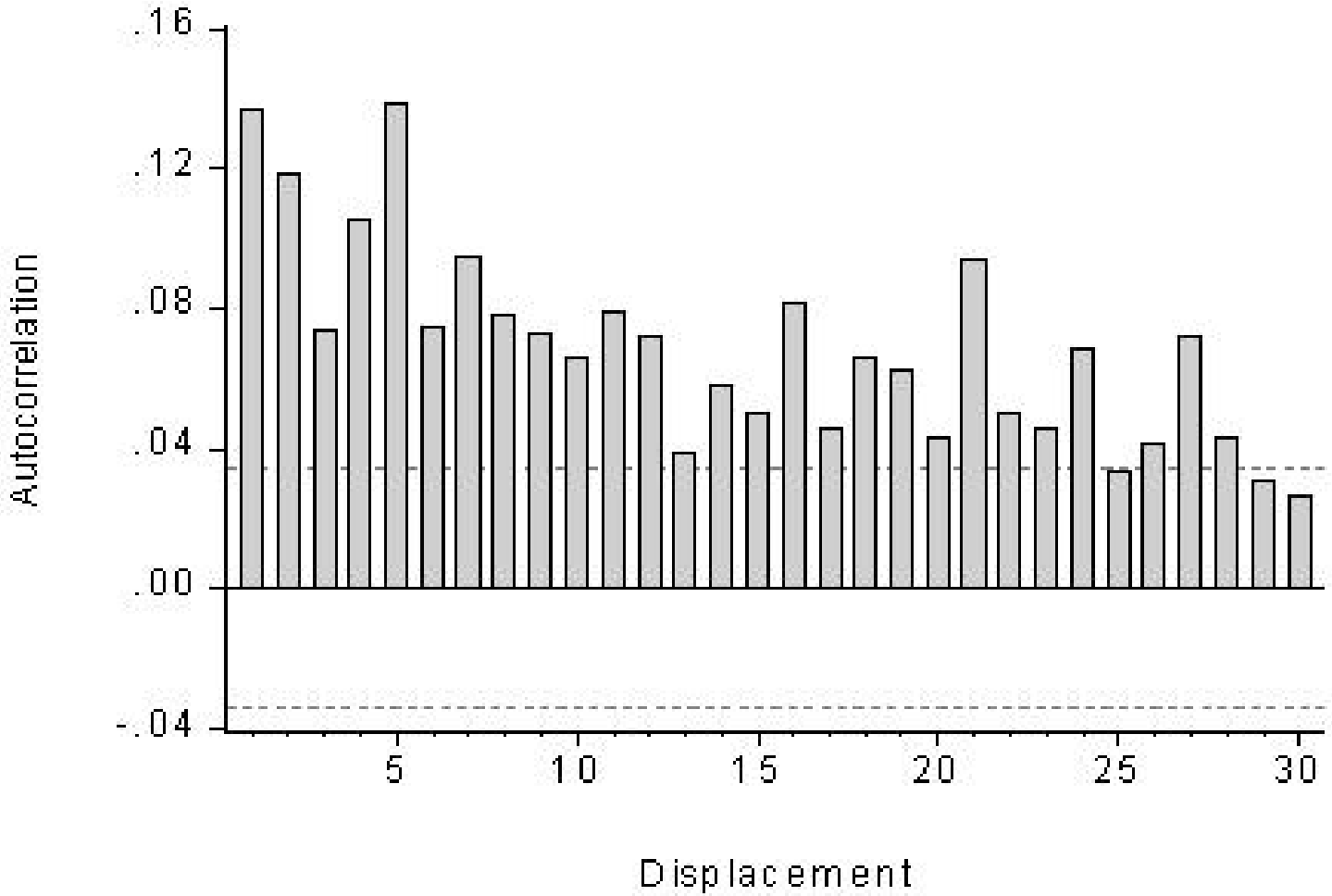
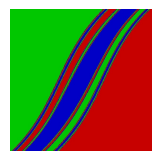


Figure: Correlogram of Squared Daily NYSE Returns.



Compare: Correlogram of Squared *Standardized* Returns

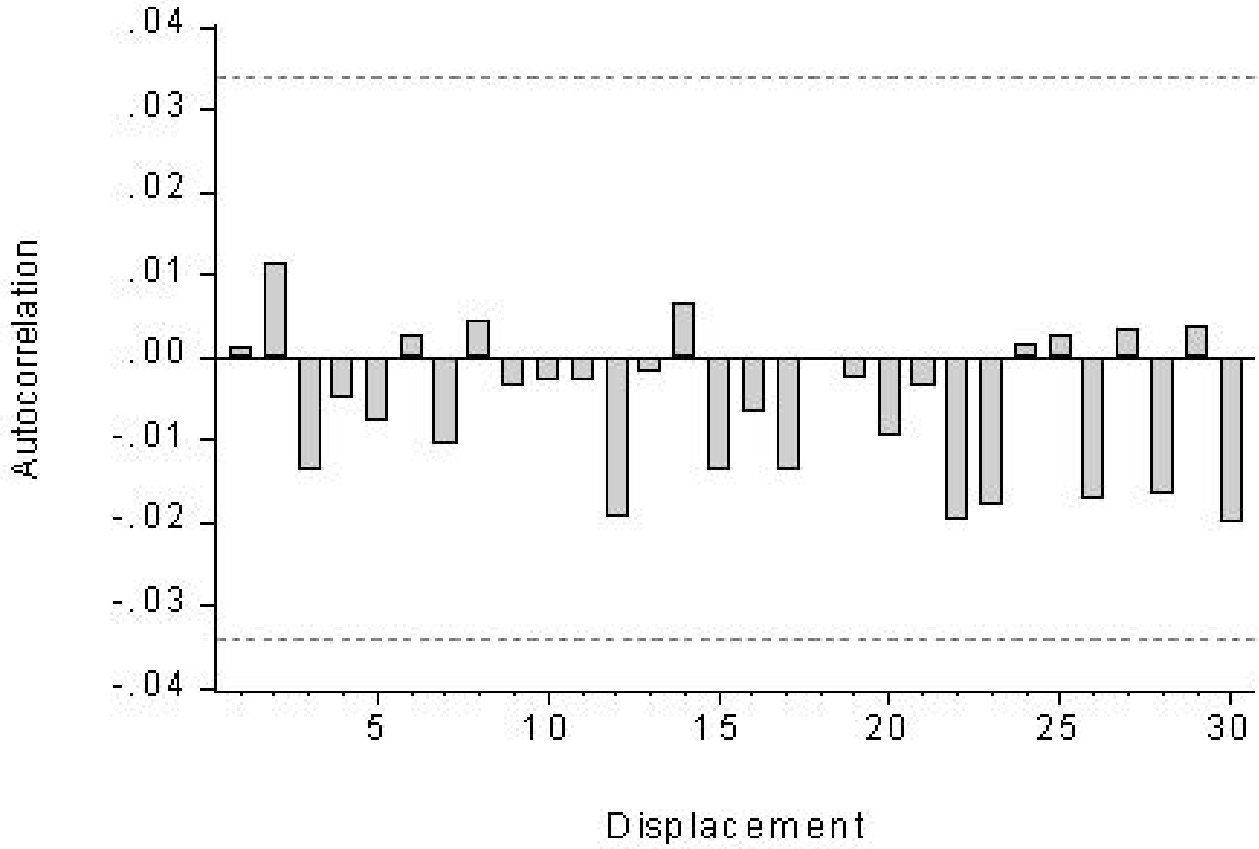
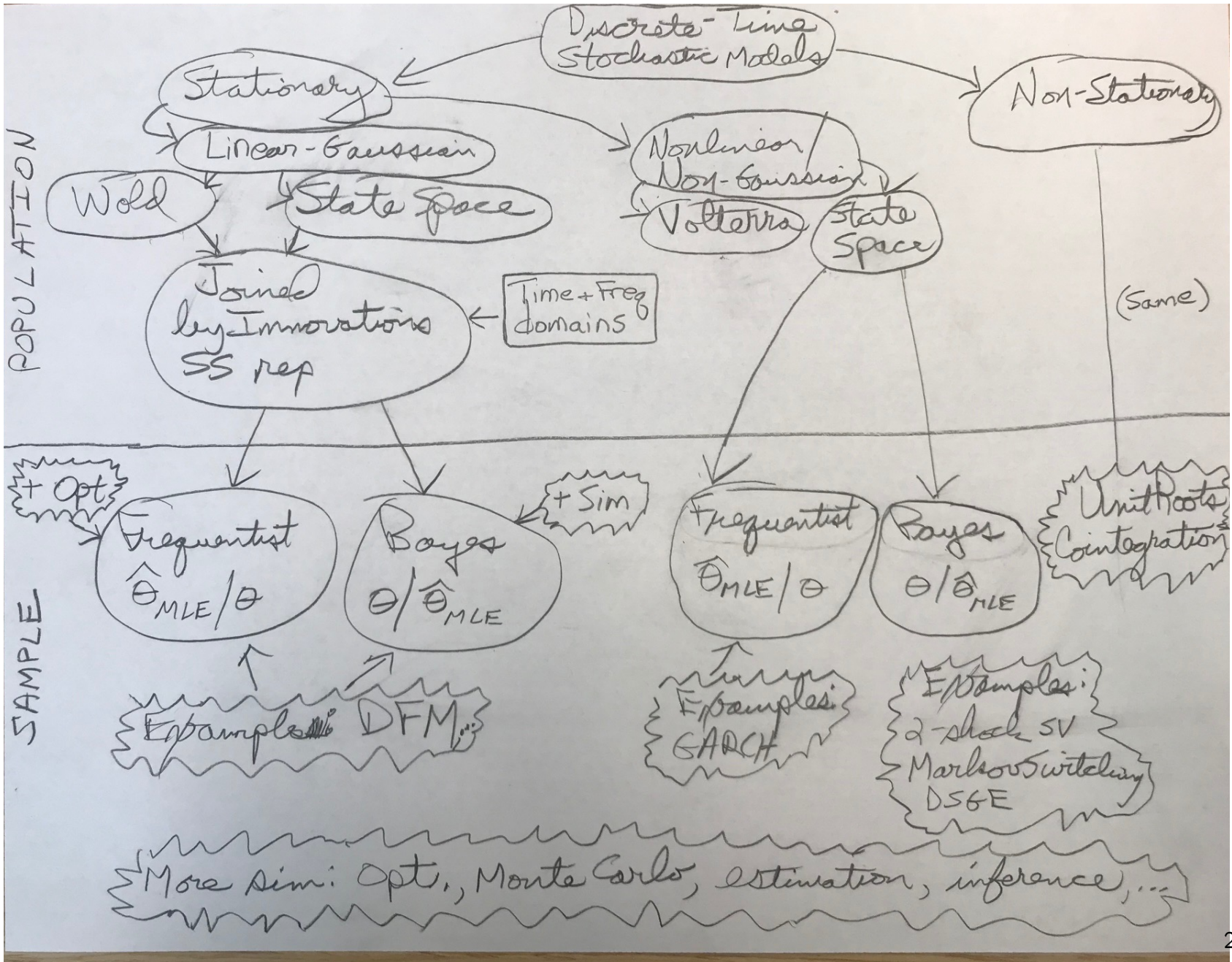
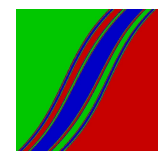


Figure: Correlogram of Squared Standardized Daily NYSE Returns.

Figure: A Map


Discrete-Time Stochastic Models
Stationary
Non-Stationary
Linear-Gaussian
Nonlinear/Non-Gaussian
Wold
State Space
Volterra
State Space
POPULATION
Joined by Innovations SS rep
Time + Freq domains
(Same)
+ Opt
+ Sim
Frequentist $\hat{\theta}_{MLE}/\theta$
Bayes $\theta/\hat{\theta}_{MLE}$
Frequentist $\hat{\theta}_{MLE}/\theta$
Bayes $\theta/\hat{\theta}_{MLE}$
Unit Roots Cointegration
SAMPLE
Examples: DFM...
Examples: GARCH
Examples: 2-shock SV Markov Switching DSGE
More sim: opt., Monte Carlo, estimation, inference, ...

Integration and Cointegration



Random Walks

Random walk:

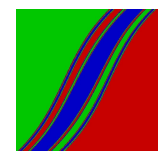
$$y_t = y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

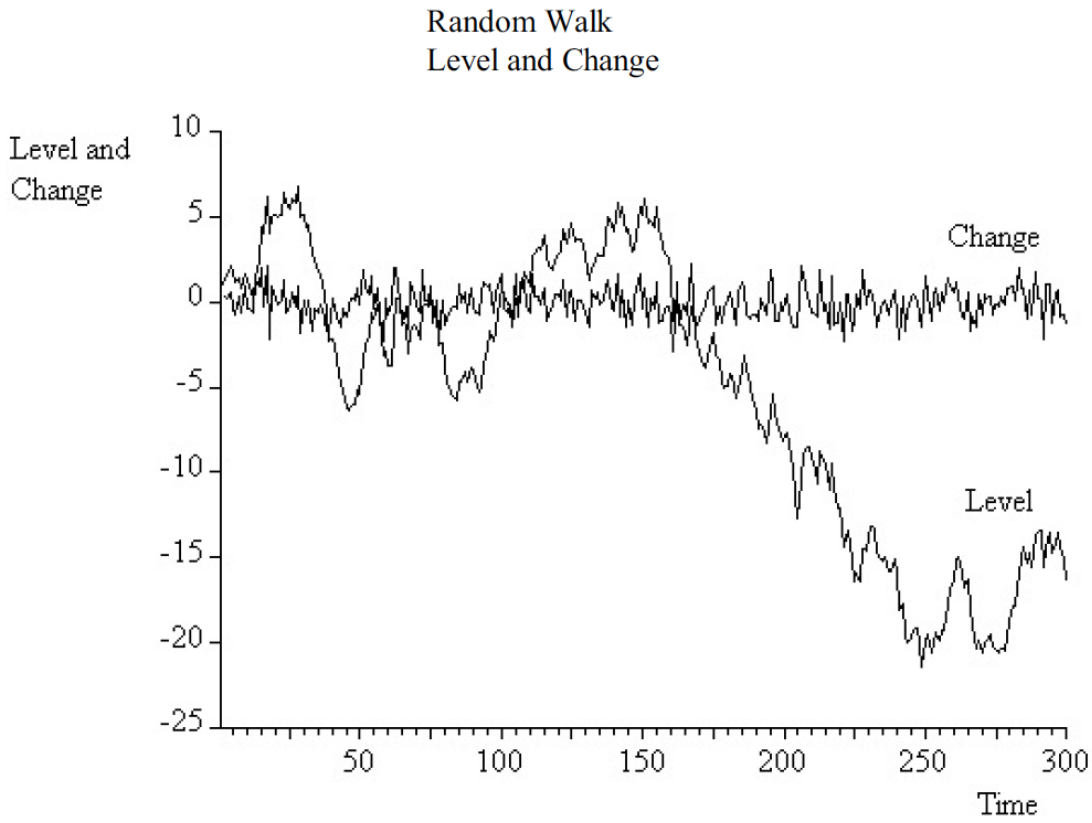
Random walk with drift:

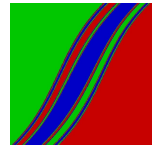
$$y_t = \delta + y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

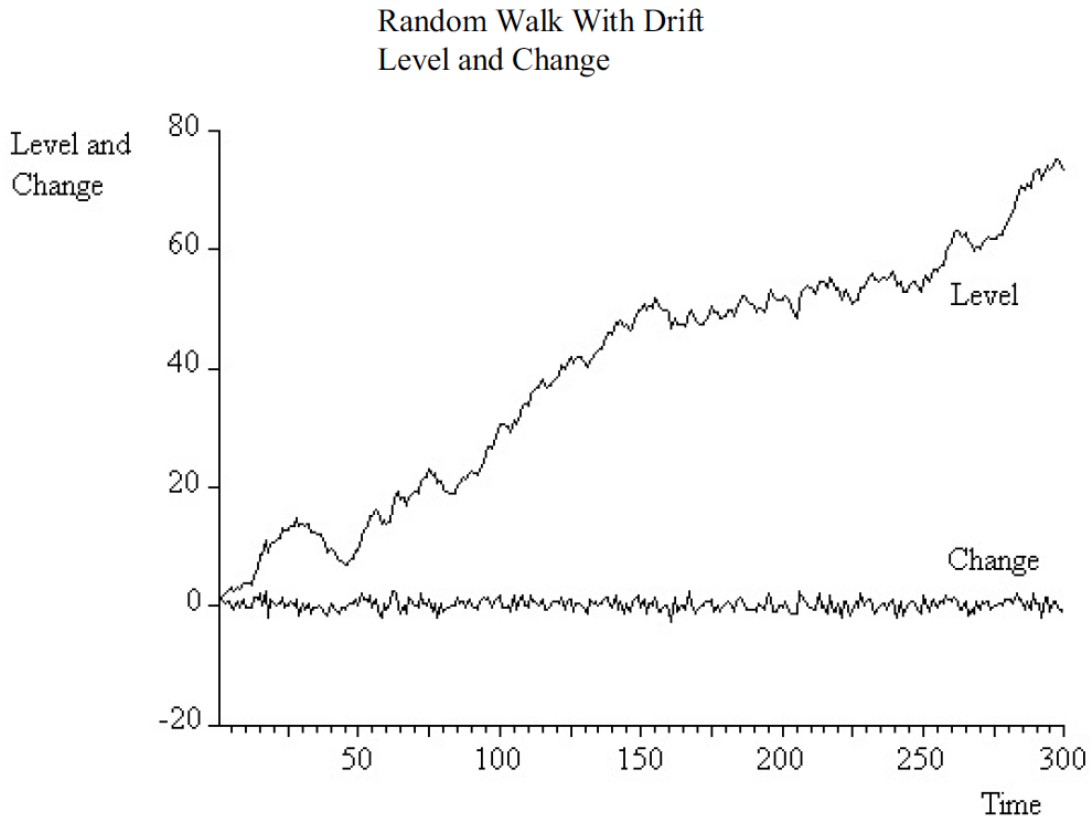


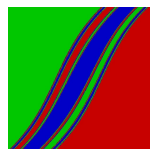
Random Walk – Level and Change





Random Walk With Drift – Level and Change





Properties of the Random Walk

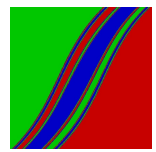
$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

(shocks perfectly persistent)

$$E(y_t) = y_0$$

$$var(y_t) = t\sigma^2$$

$$\lim_{t \to \infty} var(y_t) = \infty$$



Properties of the Random Walk with Drift

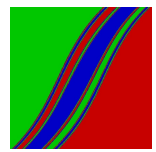
$$y_t = t\delta + y_0 + \sum_{i=1}^{t} \varepsilon_i$$

(shocks again perfectly persistent)

$$E(y_t) = y_0 + t\delta$$

$$var(y_t) = t\sigma^2$$

$$\lim_{t \to \infty} var(y_t) = \infty$$



Forecasting a Random Walk with Drift

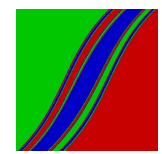
$$y_t = \delta + y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim WN$$

Optimal forecast:

$$y_{T+h,T} = \delta h + y_T$$

Forecast does not revert to trend



Forecasting a Linear Trend + Stationary AR(1)

$$y_t = a + \delta t + u_t$$

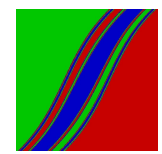
$$u_t = \phi u_{t-1} + v_t$$

$$v_t \sim WN$$

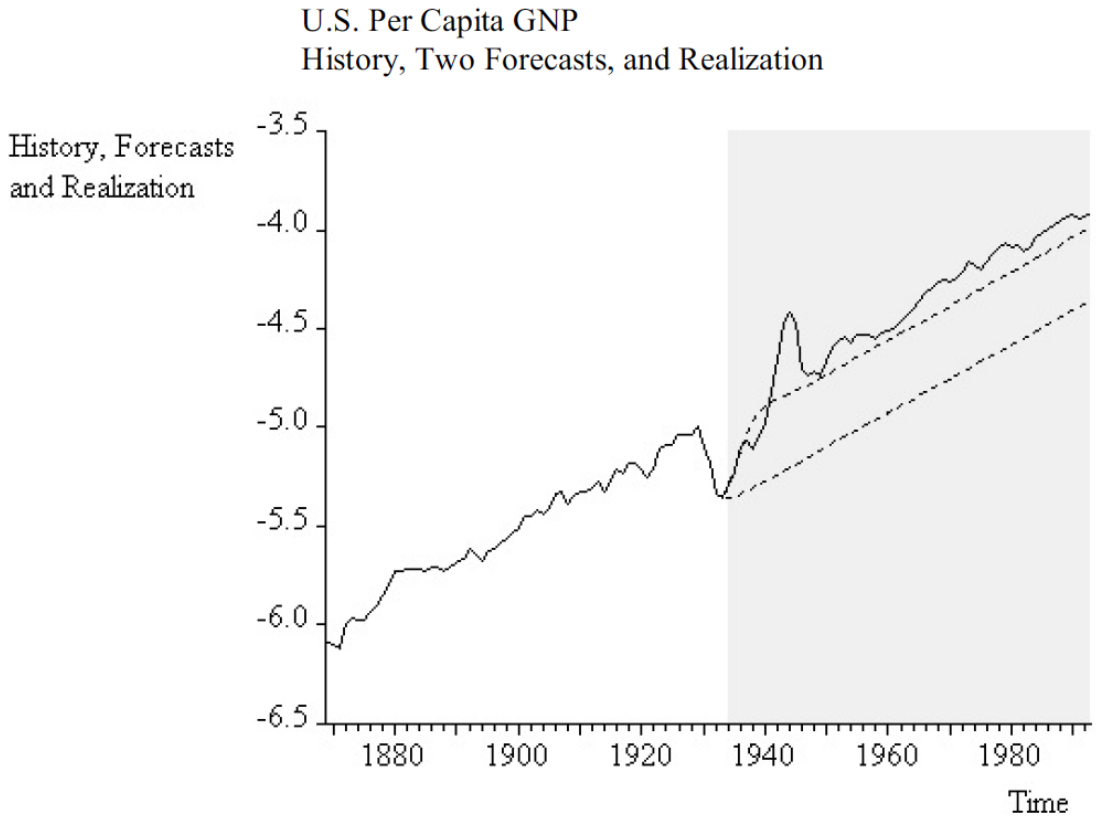
Optimal forecast:

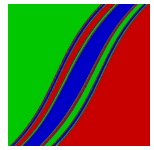
$$y_{T+h,T} = a + \delta(T + h) + \phi^h u_T$$

Forecast reverts to trend



U.S. Per Capita GNP
History, Two Forecasts, and Realization





Some Language...

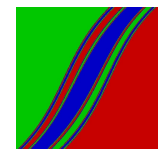
"Random walk with drift" vs. "stat. $AR(1)$ around linear trend"

"unit root" vs. "stationary root"

"Difference stationary" vs. "trend stationary"

"Stochastic trend" vs. "deterministic trend"

"$I(1)$" vs. "$I(0)$"



Unit Root Distribution in the AR(1) Process

DGP:

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

LS regression run:

$$y_t \rightarrow y_{t-1}$$

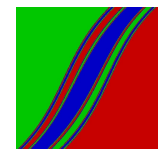
Then we have:

$$T(\hat{\phi}_{LS} - 1) \overset{d}{\rightarrow} DF$$

"Superconsistent"

Not Gaussian, even asymptotically

DF tabulated by Monte Carlo



Studentized Version

DGP:

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

LS regression run:

$$y_t \to y_{t-1}$$

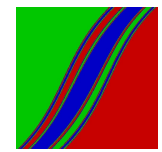
Form the "t-stat":

$$\hat{\tau} = \frac{\hat{\phi} - 1}{s\sqrt{\frac{1}{\sum_{t=2}^{T} y_{t-1}^2}}}$$

"Random walk" vs. "stationary $AR(1)$"

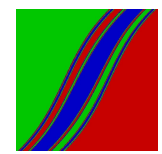
Not t in finite samples, and not $N(0,1)$ asymptotically

Again, tabulate by Monte Carlo



Example of Monte Carlo for the Dickey-Fuller Distribution: Tabulating the Null Distribution of $\hat{\tau}$

1. Draw T $N(0,1)$ innovations $\varepsilon_1, ..., \varepsilon_T$
2. Convert $\varepsilon_1, ..., \varepsilon_T$ into $y_1, ..., y_T$ using $y_t = y_{t-1} + \varepsilon_t$ and $y_0 = 0$
3. Run DF regression $y_t \rightarrow y_{t-1}$ and get $\hat{\tau}$
4. Repeat 100000 times, yielding $\{\hat{\tau}^i\}_{i=1}^{100000}$
5. Sort the $\hat{\tau}^i$'s and compute percentiles



Nonzero Mean Under the Alternative

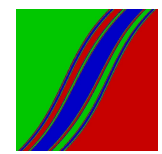
$$(y_t - \mu) \;=\; \phi(y_{t-1} - \mu) \;+\; \varepsilon_t$$

$$y_t \;=\; \alpha \;+\; \phi y_{t-1} \;+\; \varepsilon_t$$

where $\alpha = \mu(1 - \phi)$

"Random walk" vs. "stationary $AR(1)$"

Studentized statistic $\hat{\tau}_\mu$



Linear Trend Under the Alternative

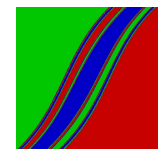
$$(y_t - a - b\,t) = \phi(y_{t-1} - a - b\,(t-1)) + \varepsilon_t$$

$$y_t = \alpha + \beta t + \phi y_{t-1} + \varepsilon_t$$

where $\alpha = a(1-\phi) + b\phi$ and $\beta = b(1-\phi)$

"Random walk with drift" vs. "stat. $AR(1)$ around linear trend"
"Difference stationary" vs. "trend stationary"
"Stochastic trend" vs. "deterministic trend"

Studentized statistic $\hat{\tau}_\tau$



$AR(p)$
"Augmented Dickey Fuller" ("ADF")

Any univariate $AR(p)$,

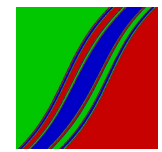
$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + u_t,$$

can be written as

$$\Delta y_t = -\pi y_{t-1} + \sum_{i=1}^{p-1} b_i \Delta y_{t-i} + u_t.$$

– In the unit root case, $\pi = 0$ ($AR(p-1)$ in changes).
– Use standard automatically-computed t-stat
(which of course does not have the t-distribution)

DF "trick form": regress change on lagged level and lagged changes ("augmentation lags")



Multivariate Problem: Spurious Time-Series Regressions

Regress a persistent variable on an *unrelated* persistent variable:

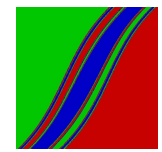
$$y_{1t} \to c, \; y_{2t}$$

(Canonical case: y_1, y_2 independent driftless random walks)

$$\frac{\hat{\beta}}{\sqrt{T}} \overset{d}{\to} RV \; (\hat{\beta} \text{ diverges})$$

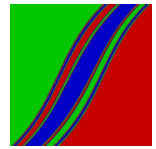
$$\frac{t}{\sqrt{T}} \overset{d}{\to} RV \; (t \text{ diverges})$$

$$R^2 \overset{d}{\to} RV \; (\text{not zero})$$



When are I(1) Levels Regressions *Not* Spurious?

Answer: When the variables are cointegrated.

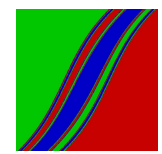


Cointegration

Consider an N-dimensional variable y:

$$y \sim CI(d, b) \text{ if}$$

1. $y_i \sim I(d),\ i = 1, \ldots, N$
2. $\exists$ 1 or more linear combinations $z_t = \alpha' y_t$ s.t. $z_t \sim I(d - b), b > 0$



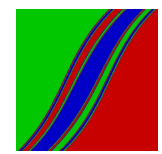
Leading Case

$$y \sim CI(1,1) \text{ if}$$

$$(1)\ y_i \sim I(1),\ i = 1, \ldots, N$$

$$(2)\ \exists \text{ 1 or more linear combinations}$$

$$z_t = \alpha' y_t \text{ s.t. } z_t \sim I(0)$$

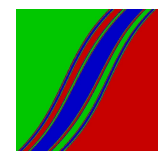


Example

$$y_{1t} = y_{1,t-1} + v_t \quad (I(1))$$

$$y_{2t} = y_{1,t-1} + \varepsilon_t \quad (I(1))$$

$$\Longrightarrow \; (y_{2t} - y_{1t}) = \varepsilon_t - v_t \quad (I(0))$$



Cointegration and "Attractor Sets"

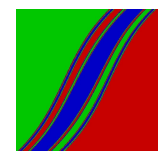
y_t is N-dimensional but does not wander randomly in $\mathbb{R}^N$

$\alpha' y_t$ is attracted to an $(N - R)$-dimensional subspace of $\mathbb{R}^N$

N: space dimension

R: number of cointegrating relationships

Attractor dimension $= N - R$
("number of underlying unit roots")
("number of common trends")



Example

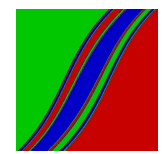
3-dimensional $VAR(p)$, all variables $I(1)$

$R = 0 \Leftrightarrow$ no cointegration $\Leftrightarrow$ y wanders throughout $\mathbb{R}^3$

$R = 1 \Leftrightarrow$ 1 cointegrating vector $\Leftrightarrow$ y attracted to a 2-Dim hyperplane in $\mathbb{R}^3$ given by $\alpha' y = 0$

$R = 2 \Leftrightarrow$ 2 cointegrating vectors $\Leftrightarrow$ y attracted to a 1-Dim hyperplane (line) in $\mathbb{R}^3$ given by intersection of two 2-Dim hyperplanes, $\alpha_1' y = 0$ and $\alpha_2' y = 0$

$R = 3 \Leftrightarrow$ 3 cointegrating vectors $\Leftrightarrow$ y attracted to a 0-Dim hyperplane (point) in $\mathbb{R}^3$ given by the intersection of three 2-Dim hyperplanes, $\alpha_1' y = 0$, $\alpha_2' y = 0$ and $\alpha_3' y = 0$
(Covariance stationary around $E(y)$)



Cointegration Motivation: Dynamic Factor Model

Consider a simple bivariate example:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} f_t + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

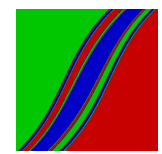
$$f_t = f_{t-1} + \eta_t$$

"Common trend" f_t

Note that $\underbrace{\frac{y_{1t}}{\lambda_1}}_{I(1)} - \underbrace{\frac{y_{2t}}{\lambda_2}}_{I(1)} = \underbrace{\frac{\varepsilon_{1t}}{\lambda_1} - \frac{\varepsilon_{2t}}{\lambda_2}}_{I(0)}$

So we have a linear combination of $I(1)$'s that is $I(0)$

– Immediate generalization to N-dimensional system with $N-R$ $I(1)$ factors (common trends)



Cointegration Motivation: Optimal Forecasting

$I(1)$ variables always co-integrated with their optimal forecasts

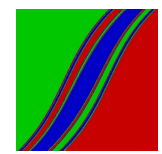
Example:

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

$$y_{t+h|t} = y_t$$

$$\implies \underbrace{y_{t+h}}_{I(1)} - \underbrace{y_{t+h|t}}_{I(1)} = \underbrace{\sum_{i=1}^{h} \varepsilon_{t+i}}_{I(0)}$$

(finite MA, always covariance stationary)



Cointegration Motivation: Long-Run Relation Augmented with Short-Run Dynamics

Simple Bivariate AR Case (ECM):

$$\Delta y_{1,t} = \alpha \Delta y_{1,t-1} + \beta \Delta y_{2,t-1} - \gamma(y_{1,t-1} - \delta y_{2,t-1}) + u_t$$

$$= \alpha \Delta y_{1,t-1} + \beta \Delta y_{2,t-1} - \gamma z_{t-1} + u_t$$

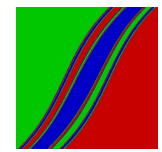
General AR Case (VECM):

$$A(L)\Delta y_t = -\gamma z_{t-1} + u_t$$

where:

$$A(L) = I - A_1 L - ... - A_p L^p$$

$$z_t = \alpha' y_t$$

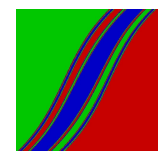


Recall Univariate ADF Regression

Any univariate $AR(p)$ can be written as

$$\Delta y_t = -\pi y_{t-1} + \sum_{i=1}^{p-1} b_i \Delta y_{t-i} + u_t.$$

In the unit root case, $\pi = 0$.



$VAR(p)$
Multivariate ADF Regression

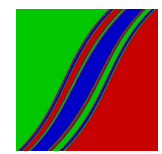
Any $VAR(p)$ can be written as

$$\Delta y_t = -\underbrace{\Pi}_{N\times N} y_{t-1} + \sum_{i=1}^{p-1} B_i \Delta y_{t-i} + u_t.$$

Now more possibilities for Π:

– Zero rank, full rank (like univariate $\pi = 0$ or $\pi \neq 0$)

– Intermediate rank (impossible in univariate)

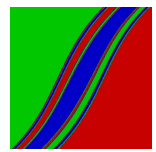


Integration/Cointegration Status

- $Rank(\Pi) = 0$
  0 cointegrating vectors, N underlying unit roots
  (all variables appropriately specified in differences)

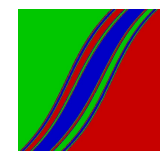
- $Rank(\Pi) = N$
  N cointegrating vectors, 0 unit roots
  (all variables appropriately specified in levels)

- $Rank(\Pi) = R \quad (0 < R < N)$
  R cointegrating vectors, $N - R$ unit roots



Granger Representation Theorem

$$y_t \sim VECM \quad \Longleftrightarrow \quad y_t \sim CI(1,1)$$



VECM $\Leftarrow$ Cointegration

We can always write the ADF form:

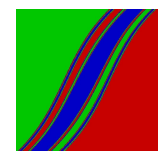
$$\Delta y_t = \sum_{i=1}^{p-1} B_i \Delta y_{t-i} - \Pi y_{t-1} + u_t$$

But under cointegration, $rank(\Pi) = R < N$, so

$$\underset{N \times N}{\Pi} = \underset{N \times R}{\gamma} \quad \underset{R \times N}{\alpha'}$$

$$\Longrightarrow \quad \Delta y_t = \sum_{i=1}^{p-1} B_i \Delta y_{t-i} - \gamma \alpha' y_{t-1} + u_t$$

$$= \sum_{i=1}^{p-1} B_i \Delta y_{t-i} - \gamma z_{t-1} + u_t$$



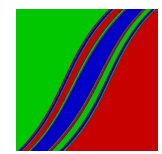
VECM $\Rightarrow$ Cointegration

$$\Delta y_t \;=\; \sum_{i=1}^{p-1} B_i\, \Delta y_{t-i} \;-\; \gamma\, \alpha'\, y_{t-1} \;+\; u_t$$

Premultiply by α':

$$\alpha' \Delta y_t \;=\; \alpha' \sum_{i=1}^{p-1} B_i\; \Delta y_{t-i} \;-\; \underbrace{\alpha'\gamma}_{\text{full rank}}\; \alpha' y_{t-1} \;+\; \alpha'\; u_t$$

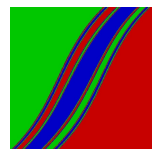
So equation balance requires that $\alpha' y_{t-1}$ be stationary.



Stationary-Nonstationary Decomposition

$$\underset{(N \times N)}{M'} \quad \underset{(N \times 1)}{y} = \begin{pmatrix} \alpha' \\ (R \times N) \\ \\ \delta \\ (N-R) \times N \end{pmatrix} y = \begin{pmatrix} CI \text{ combs} \\ \\ com. \text{ trends} \end{pmatrix},$$

where the rows of δ are orthogonal to the columns of γ



Intuition

The system is

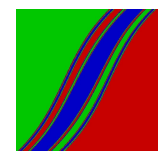
$$\Delta\, y_t \;=\; \sum_{i=1}^{p-1} B_i\, \Delta\, y_{t-i} \;\; - \gamma\, \alpha'\, y_{t-1} \;+\; \mu_t$$

Transforming the system by δ yields

$$\delta\, \Delta\, y_t \;=\; \sum_{i=1}^{p-1} \delta\, B_i\, \Delta\, y_{t-i} \;\; - \underbrace{\delta\, \gamma}_{\text{0 by orthogonality}} \alpha'\, y_{t-1} \;+\; \delta\, \mu_t$$

So δ isolates that part of the VECM
that is appropriately specified as a VAR in differences.

Note that if we *start* with M′y, then the observed series is
$(\mathrm{M}')^{-1}$ M′y, so nonstationarity is spread throughout the system.



Example

$$y_{1t} = y_{1t-1} + u_{1t}$$

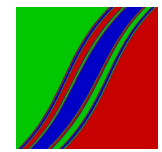
$$y_{2t} = y_{1t-1} + u_{2t}$$

Levels form:

$$\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} L \right) \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

Dickey-Fuller form:

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = - \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$



Example, Continued

$$\Pi = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix} = \gamma\alpha'$$

$$M' = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \textit{first`row just } \alpha' \\ \textit{second row orthogonal to } \gamma \end{pmatrix}$$

$$M' \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} u_{2t} - u_{1t} \\ y_{1t} \end{pmatrix} = \begin{pmatrix} y_{2t} - y_{1t} \\ y_{1t} \end{pmatrix}$$

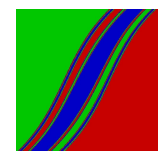
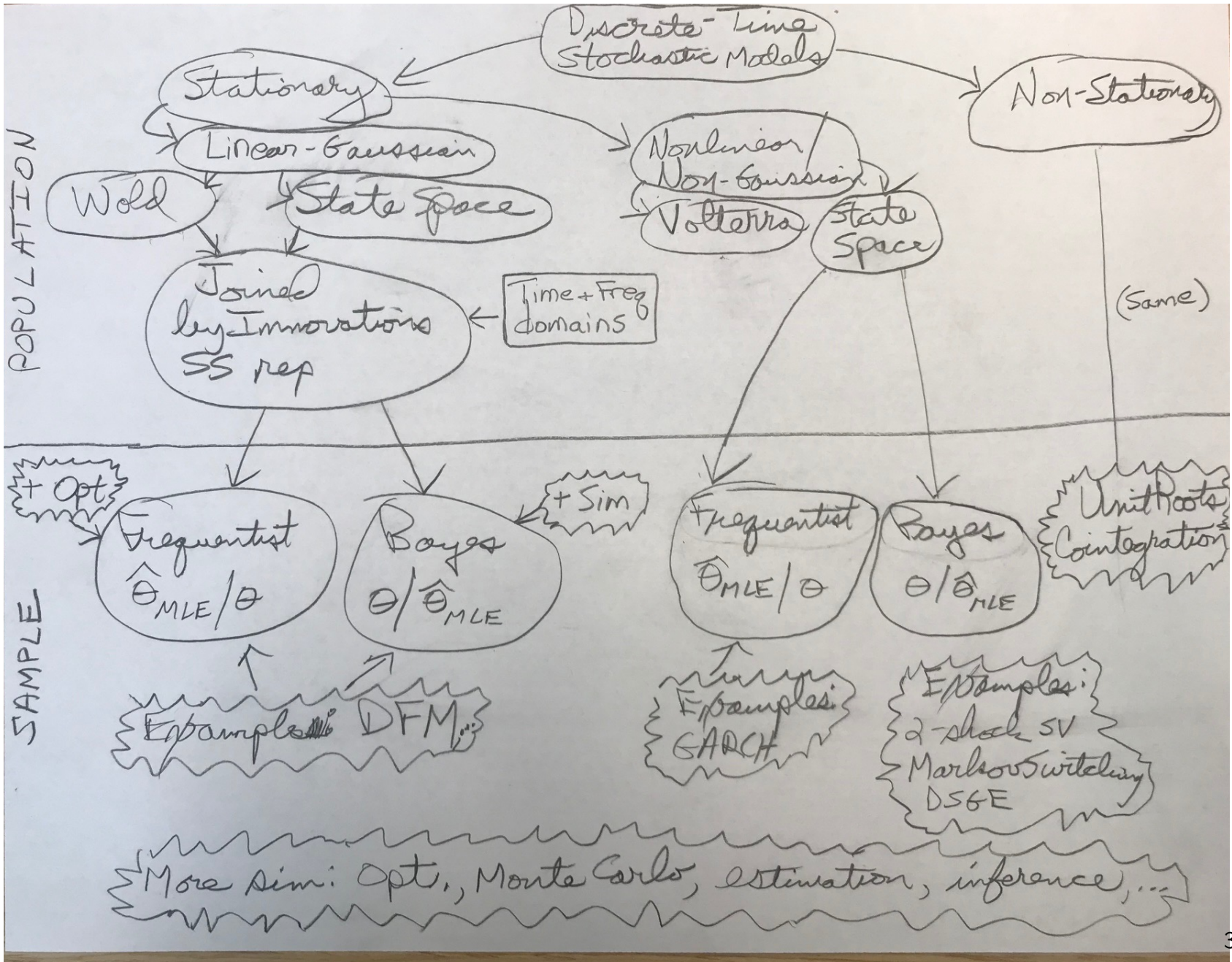
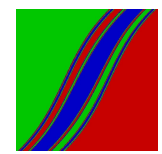


Figure: A Map


Discrete-Time Stochastic Models
Stationary
Non-Stationary
Linear-Gaussian
Nonlinear/Non-Gaussian
Wold
State Space
Volterra
State Space
POPULATION
Joined by Innovations SS rep
Time + Freq domains
(Same)
+ Opt
+ Sim
Frequentist $\hat{\theta}_{MLE}/\theta$
Bayes $\theta/\hat{\theta}_{MLE}$
Frequentist $\hat{\theta}_{MLE}/\theta$
Bayes $\theta/\hat{\theta}_{MLE}$
Unit Roots Cointegration
SAMPLE
Examples: DFM...
Examples: GARCH
Examples: 2-shock SV Markov Switching DSGE
More sim: Opt., Monte Carlo, estimation, inference, ...

More Simulation:
Global Optimization by Simulation

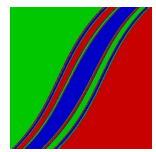


On Local vs. Global Optima

1. Try many startup values (sounds trivial but very important)

2. At the end of it all, use extreme value theory to assess the likelihood that the local optimum is global ("Veall's Method")

3. Actually use a global optimizer



Assessing Whether a Local Optimum is Global, Cont'd

$$\theta \in \Theta \subset R^k$$

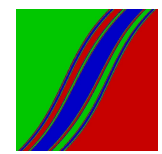
$lnL(\theta)$ is continuous

$lnL(\theta^*)$ is the unique finite global max of $lnL(\theta), \theta \in \Theta$

$H(\theta^*)$ exists and is nonsingular

$lnL(\hat{\theta})$ is a local max

Develop *statistical inference* for θ^*



Assessing Whether a Local Optimum is Global, Cont'd

Draw $\{\theta_i\}_{i=1}^N$ uniformly from Θ and form $\{lnL(\theta_i)\}_{i=1}^N$

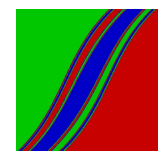
lnL_1 first order statistic, lnL_2 second order statistic

$$P[lnL(\theta^*) \in (lnL_1, lnL^\alpha)] = (1-\alpha), \text{ as } N \to \infty,$$

where

$$lnL^\alpha = lnL_1 + \frac{lnL_1 - lnL_2}{(1-\alpha)^{\frac{-2}{k}} - 1}$$

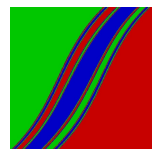
.



Global Optimization

Summary of Local Optimization:

1. initial guess $\theta^{(0)}$
2. **while** stopping criteria not met **do**
3. select $\theta^{(c)} \in N(\theta^{(m)})$ (Classically: use gradient)
4. **if** $\Delta \equiv lnL(\theta^{(c)}) - lnL(\theta^{(m)}) > 0$ **then** $\theta^{(m+1)} = \theta^{(c)}$
5. end while



Simulated Annealing
(Illustrated Here for a Discrete Parameter Space)

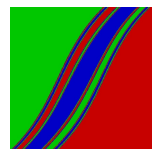
Framework:

1. A set Θ, and a real-valued function *lnL* (satisfying regularity conditions) defined on Θ. Let $\Theta^* \subset \Theta$ be the set of global maxima of *lnL*

2. $\forall \theta^{(m)} \in \Theta$, a set $N(\theta^{(m)}) \subset \Theta - \theta^{(m)}$, the set of neighbors of $\theta^{(m)}$

3. A nonincreasing function, $T(m) : N \to (0, \infty)$ ("the cooling schedule"), where $T(m)$ is the "temperature" at iteration m

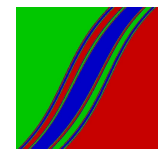
4. An initial guess, $\theta^{(0)} \in \Theta$



Simulated Annealing Algorithm

1. initial guess $\theta^{(0)}$
2. **while** stopping criteria not met **do**
3. select $\theta^{(c)} \in N(\theta^{(m)})$
4. **if** $\Delta > 0$ or $\exp(\Delta / T(m)) > U(0,1)$ **then** $\theta^{(m+1)} = \theta^{(c)}$
5. **end while**

Note the extremes:
$T = 0$ implies no randomization (like classical gradient-based)
$T = \infty$ implies complete randomization (like random search)



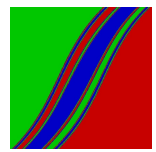
A (Heterogeneous) Markov Chain

$$\text{If } \theta^{(c)} \notin N(\theta^{(m)}) \text{ then}$$

$$P(\theta^{(m+1)} = \theta^{(c)} | \theta^{(m)}) = 0$$

$$\text{If } \theta^{(c)} \in N(\theta^{(m)}) \text{ then}$$

$$P(\theta^{(m+1)} = \theta^{(c)} \,|\, \theta^{(m)}) \;=\; exp\,(\min[0,\ \Delta / T(m)])$$

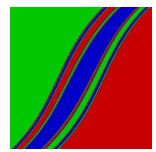


Convergence of a Global Optimizer

Definition. We say that the simulated annealing algorithm converges if

$$\lim_{m\to\infty} \; P[\theta^{(m)} \in \Theta^*] \; = \; 1.$$

Definition: We say that $\theta^{(m)}$ communicates with Θ^* at depth d if there exists a path in Θ (with each element of the path being a neighbor of the preceding element) that starts at $\theta^{(m)}$ and ends at some element of Θ^*, such that the smallest value of lnL along the path is $lnL(\theta^{(m)}) - d$.



Convergence of Simulated Annealing

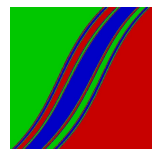
Theorem: Let d^* be the smallest number such that every $\theta^{(m)} \in \Theta$ communicates with Θ^* at depth d^*. Then the simulated annealing algorithm converges if and only if, as $m \to \infty$,

$$T(m) \to 0$$
and
$$\sum exp(-d^*/T(m)) \to \infty.$$

Problems: How to choose T, and moreover we don't know d^*

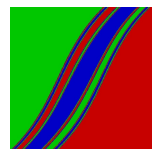
Popular choice of cooling function: $T(m) = \frac{1}{\ln m}$

Regarding speed of convergence, little is known



More Simulation:
Econometric Theory by Simulation

(Monte Carlo Methods)



Monte Carlo

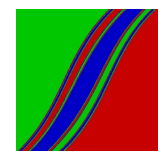
Key: Solve deterministic problems by simulating stochastic analogs, with the analytical unknowns reformulated as parameters to be estimated.

Many important discoveries made by Monte Carlo.

Also, numerous *mistakes avoided* by Monte Carlo!

The pieces:

(I) Experimental Design
(II) Simulation (including variance reduction techniques)
(III) Analysis: Response surfaces (which also reduce variance)



(I) Experimental Design

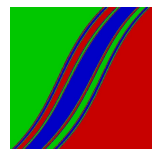
- Data-Generating Process (DGP)
- Objective
  - e.g., MSE of an estimator:

$$E[(\theta - \hat{\theta})^2] \;=\; g(\theta, T)$$

  - e.g., Power function of a test:

$$\pi \;=\; g(\theta, T)$$

- Selection of $(\theta,\ T)$ Configurations to Explore
- Number of Monte Carlo Repetitions (N)



(II) Simulation

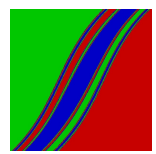
Running example: Monte Carlo integration

(Most of the things we examine by Monte Carlo
are expectations, and hence integrals!
(Estimator MSE, test size and power, etc.)

Canonical definite integral: $\mu = \int_0^1 m(y)dy$

Key insight:

$$\mu = \int_0^1 m(y)dy = E(m(y))$$
$$y \sim U(0, 1)$$



"Direct Simulation", General Case

$$\mu = E(m(y)) = \int m(y) f(y) dy$$

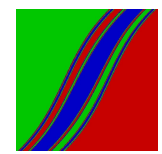
– Indefinite integral, arbitrary function $m(\cdot)$, arbitrary density $f(y)$

Draw $y_i \sim f(\cdot)$, and then form $m(y_i)$. Repeat. Then form:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} m(y_i).$$

Immediately,

$$\sqrt{N}(\hat{\mu} - \mu) \quad \stackrel{d}{\rightarrow} \quad N$$



Direct Simulation, Leading Special Case (Mean)

$$\mu = E(y) = \int y\, f(y)\, dy$$

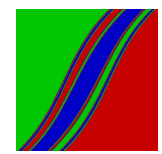
– Indefinite integral, $m(y) = y$, arbitrary density $f(y)$

Draw $y_i \sim f(\cdot)$. Repeat. Then form:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} y_i.$$

Immediately,

$$\sqrt{N}(\hat{\mu} - \mu) \quad \stackrel{d}{\rightarrow} \quad N$$



*In*direct Simulation:
Importance Sampling to Facilitate Sampling

Sampling from $f(\cdot)$ may be difficult. So change to:

$$\mu = \int y \frac{f(y)}{g(y)} g(y) dy$$

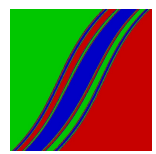
where the "importance sampling density" $g(\cdot)$ is easy to sample

Draw many $y_i \sim g(\cdot)$, and then form:

$$\hat{\mu}_* = \frac{1}{N} \sum_{i=1}^{N} y_i \frac{f(y_i)}{g(y_i)} = \sum_{i=1}^{N} w_i y_i$$

– Avg of $f(y)$ draws replaced by *weighted* avg of $g(y)$ draws

$$\sqrt{N}(\hat{\mu}_* - \mu) \quad \stackrel{d}{\rightarrow} \quad N$$

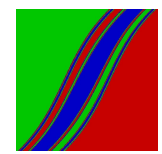


(Much) More on Indirect Simulation: Variance Reduction

"Variance-Reduction Techniques"

– Same accuracy for smaller N

– Greater accuracy for same N



Importance Sampling to Achieve Variance Reduction

$$\text{Again}: \ \mu = \int y f(y) dy = \int y \frac{f(y)}{g(y)} g(y) dy,$$

$$\text{And again}: \ \hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} y_i \ \ (y_i \text{ drawn from } f)$$

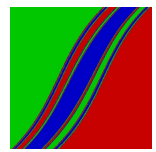
$$\text{And again}: \ \sqrt{N}(\hat{\mu} - \mu) \to_d \ N(0, \ \sigma^2)$$

$$\text{And again}: \ \hat{\mu}_* = \frac{1}{N} \sum_{i=1}^{N} y_i \frac{f(y_i)}{g(y_i)} \ \ (y_i \text{ drawn from } g)$$

$$\text{And again}: \ \sqrt{N}(\hat{\mu}_* - \mu) \ \ \to_d \ N(0, \ \sigma_*^2)$$

The new point: If $g(y)$ is chosen judiciously, $\sigma_*^2 \ll \sigma^2$

Key: Pick $g(y)$ s.t. $\frac{yf(y)}{g(y)}$ has small variance



Importance Sampling Example

Let $y \sim N(0,1)$, and estimate the mean of $I(y > 1.96)$:

$$\mu = E(I(y > 1.96)) = P(y > 1.96) = \int \underbrace{I(y > 1.96)}_{y} \; \underbrace{\phi(y)}_{f(y)} \; dy$$

$$\hat{\mu} = \sum_{i=1}^{N} \frac{I(y_i > 1.96)}{N} \quad \text{(with variance } \sigma^2\text{)}$$

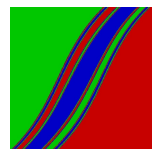
Use importance sampler:

$$g(y) = N(1.96,\, 1)$$

$$P(y > 1.96) = \int I(y > 1.96) \frac{\phi(y)}{g(y)} \; g(y) \; dy$$

$$\hat{\mu}_* = \frac{\sum_{i=1}^{N} I(y_i > 1.96) \frac{\phi(y_i)}{g(y_i)}}{N} \quad \text{(with variance } \sigma_*^2\text{)}$$

$$\frac{\sigma_*^2}{\sigma^2} \approx 0.06$$



Control Variates

Instead of simulating the mean of y, just simulate the mean of $(y - c(y))$, where the "control function" $c(y)$ has known mean and is highly correlated with y (i.e., $c(y)$ is simple enough to integrate analytically and flexible enough to absorb most of the variation in y).

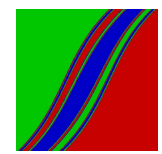
$$\mu = \int y f(y) dy = \int c(y) f(y) dy + \int (y - c(y)) f(y) dy$$

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

$$\hat{\mu}_* = \int c(y) f(y) dy + \frac{1}{N} \sum_{i=1}^{N} (y_i - c(y_i))$$

$$\sqrt{N}(\hat{\mu}_* - \mu) \rightarrow_d N(0, \sigma_*^2)$$

If $c(y)$ is chosen judiciously, $\sigma_*^2 \ll \sigma^2$



Control Variate Example

Suppose we want $\mu = E(e^y) = \int_0^1 e^y dy$, for $y \sim U(0,1)$

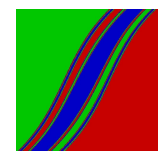
$$\text{Immediately}: \quad \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} e^{y_i}$$

Alternatively, use control variate: $c(y) = 1 + 1.7y$

$$\text{Immediately}: \quad \int_0^1 c(y)dy = \left(y + \frac{1.7}{2}y^2\right)\Bigg|_0^1 = 1.85$$

$$\text{So}: \quad \hat{\mu}_* = 1.85 + \frac{1}{N}\sum_{i=1}^{N}(e^{y_i} - (1 + 1.7y_i))$$

$$\frac{\hat{\sigma}_*^2}{\hat{\sigma}^2} \approx .01$$



Antithetic Variates

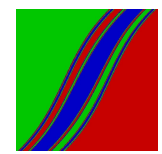
Average negatively-correlated unbiased estimators of μ

(Unbiasedness maintained, variance reduced)

How to get negative correlation?

– If $y \sim U(0,1)$, then so too is $(1-y)$
(so y and $1-y$ equally likely)

– If $y \sim D$, for *any* zero-mean symmetric distribution D,
then so too is $-y$ (so y and $-y$ equally likely)



How Antithetics Work (Zero-Mean Symmetric Case)

$$\mu = \int y f(y) dy$$

$$\text{Direct}: \quad \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i \ (\hat{\mu} \text{ is based on } y_i, i = 1, ..., N)$$

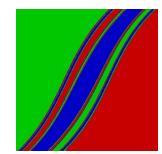
$$\text{Antithetic}: \quad \hat{\mu}_* = \frac{1}{2}\hat{\mu}_{(y)} + \frac{1}{2}\hat{\mu}_{(-y)}$$

$\hat{\mu}_{(y)}$ is based on $y_i, \ i = 1, ..., N/2$,
$\hat{\mu}_{(-y)}$ is based on $-y_i, \ i = 1, ..., N/2$

$$\sqrt{N}(\hat{\mu}_* - \mu) \to_d N(0, \sigma_*^2)$$

$$\sigma_*^2 = \frac{1}{4} var\left(\hat{\mu}_{(y)}\right) + \frac{1}{4} var\left(\hat{\mu}_{(-y)}\right) + \frac{1}{2} \underbrace{cov\left(\hat{\mu}_{(y)}, \hat{\mu}_{(-y)}\right)}_{< 0}$$

Often $\sigma_*^2 \ll \sigma^2$

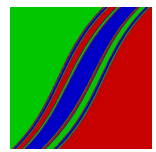


Common Random Numbers

Here we exploit *positively*-correlated estimates.

We have focused on estimation of a single integral. But interest often centers on *difference* (or ratio) of two integrals.

The key: Estimate each integral using the *same* random numbers. Then the (positively-correlated) simulation errors will tend to cancel from differences or ratios!



How Common Random Numbers Work (e.g., in Comparing MSE's of Two Estimators)

Two estimators $\hat{\delta}_1, \hat{\delta}_2$; true parameter δ_0

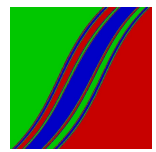
We want to compare MSE's: $E(\hat{\delta}_1 - \delta_0)^2$ *vs.* $E(\hat{\delta}_2 - \delta_0)^2$

Expected difference: $\mu = E\left((\hat{\delta}_1 - \delta_0)^2 \; - \; (\hat{\delta}_2 - \delta_0)^2\right)$

$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \left((\hat{\delta}_{1i} - \delta_0)^2 \; - \; (\hat{\delta}_{2i} - \delta_0)^2\right)$ (using indep r.n.'s)

$\hat{\mu}_* = \frac{1}{N}\sum_{i=1}^{N} \left((\hat{\delta}_{1i} - \delta_0)^2 \; - \; (\hat{\delta}_{2i} - \delta_0)^2\right)$ (using comm. r.n.'s)

$$\sigma_*^2 = \frac{1}{N} var\left((\hat{\delta}_1 - \delta_0)^2\right) + \frac{1}{N} var\left((\hat{\delta}_2 - \delta_0)^2\right) - \frac{2}{N} \underbrace{cov\left((\hat{\delta}_1 - \delta_0)^2, (\hat{\delta}_2 - \delta_0)^2\right)}_{>0}$$



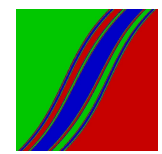
Often $\sigma_*^2 \ll \sigma^2$.

(III) Response surfaces

1. Direct Response Surfaces

2. Indirect Responses Surfaces:

   - Clear and informative graphical presentation
   - Imposition of known asymptotic results
     (e.g., power $\to 1$ as $T \to \infty$)
   - Imposition of known features of functional form
     (e.g. power $\in [0,1]$)
   - Variance reduction (!)



Example: Assessing Finite-Sample Test Size For a Fixed Alternative

$$\alpha = P\left(s > s^* | T, H_0 \text{ true}\right) = g(T)$$

(α is empirical size, s is test statistic, s^* is asymptotic c.v.)

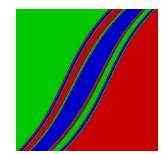
$$\hat{\alpha} = \frac{\#rej}{N}$$

$$\hat{\alpha} \sim N\left(\alpha, \frac{\alpha(1-\alpha)}{N}\right)$$

$$\text{Equivalently}: \quad \hat{\alpha} = \alpha + \varepsilon = g(T) + \varepsilon,$$

$$\text{where} \quad \varepsilon \sim N\left(0, \frac{g(T)(1-g(T))}{N}\right)$$

Note the heteroskedasticity: variance of ε changes with T



Example: Assessing Finite-Sample Test Size For a Fixed Alternative

Enforce analytically known structure on $\hat{\alpha}$.
Common approach:

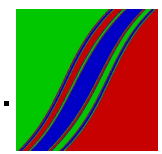
$$\hat{\alpha} = \underbrace{\alpha_0 + T^{-\frac{1}{2}} \left(c_0 + \sum_{i=1}^{p} c_i T^{\frac{-i}{2}} \right)}_{\alpha} + \varepsilon$$

α_0 is nominal size, which obtains as $T \to \infty$. Second term is the vanishing size distortion.

Response surface regression:

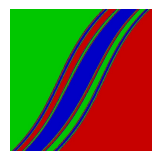
$$(\hat{\alpha} - \alpha_0) \to T^{-\frac{1}{2}}, T^{-1}, T^{-\frac{3}{2}}, ...$$

Disturbance will be approximately normal but heteroskedastic.
So use GLS or robust standard errors (or nothing – still consistent).



More Simulation:
Estimation by Simulation

(Simulated Method of Moments and Indirect Inference)



GMM

k-dimensional parameter θ

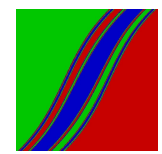
$$\hat{\theta}_{GMM} = argmin_\theta \; d(\theta)' \; \Sigma \; d(\theta)$$

where

$$d(\theta) = \begin{pmatrix} m_1(\theta) - \hat{m}_1 \\ m_2(\theta) - \hat{m}_2 \\ \vdots \\ m_r(\theta) - \hat{m}_r \end{pmatrix}$$

The $m_i(\theta)$ are model moments and the $\hat{m}_i$ are data moments.
MM: $k = r$ and the $m_i(\theta)$ calculated analytically
GMM: $k < r$ and the $m_i(\theta)$ calculated analytically

- Inefficient relative to MLE, but useful when likelihood is not available (and for other reasons, as we'll see)



Simulated Method of Moments (SMM)

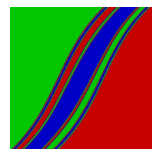
($k \leq r$ and the $m_i(\theta)$ calculated by simulation)

- Analytic model moments may be unavailable. So calculate model moments by simulation. (Use common r.n.'s when iterating to minimize the quadratic form. Why?)

- SMM: if you can simulate, you can estimate
  - If you understand a model you can simulate it, and if you can simulate it you can estimate it consistently. Eureka!

  - No need to work out what might be very complex likelihoods even if they are in principle "available."

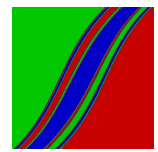
  - MLE efficiency lost may be a small price for SMM tractability gained.



SMM Under Misspecification

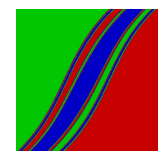
All econometric models are misspecified.
GMM/SMM has special appeal from that perspective.

- Under correct specification any consistent estimator (e.g., MLE or GMM/SMM) takes you to the right place asymptotically, and MLE has the extra benefit of efficiency.
- Under misspecification, consistency becomes an issue, quite apart from the secondary issue of efficiency. Best DGP approximation for one purpose may be very different from best for another.
- GMM/SMM is appealing in such situations, because it forces thought regarding which moments $M = \{m_1(\theta), ..., m_r(\theta)\}$ to match, and then by construction it is consistent for the M-optimal approximation.



SMM Under Misspecification, Continued

- In contrast, pseudo-MLE ties your hands. Gaussian pseudo-MLE, for example, is consistent for the KLIC-optimal approximation (1-step-ahead mean-squared prediction error).
- The bottom line: under misspecification MLE may not be consistent for what you want, whereas by construction GMM is consistent for what you want (once you *decide* what you want).



Indirect Inference

k-dimensional economic model parameter θ
$\delta > k$-dimensional auxiliary model parameter β

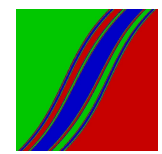
$$\hat{\theta}_{IE} = argmin_\theta \; d(\theta)' \; \Sigma \; d(\theta)$$

where

$$d(\theta) = \begin{pmatrix} \hat{\beta}_1(\theta) - \hat{\beta}_1 \\ \hat{\beta}_2(\theta) - \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_\delta(\theta) - \hat{\beta}_\delta \end{pmatrix}$$

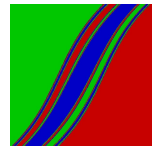
$\hat{\beta}_i(\theta)$ are est. params. of aux. model fit to simulated model data
$\hat{\beta}_i$ are est. params. of aux. model fit to real data

– Consistent for true θ if economic model correctly specified
– Consistent for pseudo-true θ otherwise



More Simulation: Inference by Simulation

(Bootstrap)



Simplest (iid) Case

$$\{x_t\}_{t=1}^{T} \sim iid(\mu, \sigma^2) \text{ (not necessarily Gaussian!)}$$

100α percent confidence interval for μ:

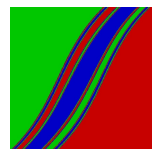
$$I = [\bar{x}_T - u_{(1+\alpha)/2}\frac{\sigma(x)}{\sqrt{T}}, \quad \bar{x}_T - u_{(1-\alpha)/2}\frac{\sigma(x)}{\sqrt{T}}]$$

$$\bar{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t$$

$$\sigma(x) = \sqrt{E(x-\mu)^2}$$

$$u_\alpha \text{ solves } P\left(\frac{(\bar{x}_T - \mu)}{\frac{\sigma}{\sqrt{T}}} \leq u_\alpha\right) = \alpha$$

Exact interval, regardless of the underlying distribution.



Operational Version

$$I = [\bar{x}_T - \hat{u}_{(1+\alpha)/2}\frac{\hat{\sigma}(x)}{\sqrt{T}}, \bar{x}_T - \hat{u}_{(1-\alpha)/2}\frac{\hat{\sigma}(x)}{\sqrt{T}}]$$

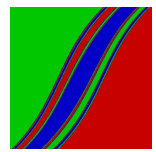
$$\hat{\sigma}^2(x) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x}_T)^2$$

$$\hat{u}_\alpha \text{ solves } P\left(\frac{(\bar{x}_T - \mu)}{\frac{\hat{\sigma}(x)}{\sqrt{T}}} \leq \hat{u}_\alpha\right) = \alpha$$

Classic (Gaussian) example:

$$I = \bar{x}_T \pm t_{(1-\alpha)/2}\frac{\hat{\sigma}(x)}{\sqrt{T}}$$

Bootstrap approach: No need to assume Gaussian data.



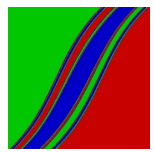
"Percentile Bootstrap"

$$Root: \ S = \frac{(\bar{x}_T - \mu)}{\frac{\sigma}{\sqrt{T}}}$$

$$Root \ c.d.f.: \ H(z) = P\left(\frac{(\bar{x}_T - \mu)}{\frac{\sigma}{\sqrt{T}}} \leq z\right)$$

1. Draw $\{x_t^{(j)}\}_{t=1}^{T}$ with replacement from $\{x_t\}_{t=1}^{T}$
2. Compute $\frac{\bar{x}_T^{(j)} - \bar{x}_T}{\frac{\hat{\sigma}(x)}{\sqrt{T}}}$
3. Repeat many times and build up the sampling distribution of $\frac{\bar{x}_T^{(j)} - \bar{x}_T}{\frac{\hat{\sigma}(x)}{\sqrt{T}}}$ which is an approximation to the distribution of $\frac{\bar{x}_T - \mu}{\frac{\sigma}{\sqrt{T}}}$

"Russian doll principle"



Percentile Bootstrap, Continued

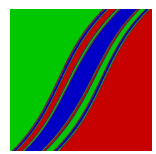
Bootstrap estimator of H(z):

$$\hat{H}(z) \;=\; P\left(\frac{(\bar{x}_T^{(j)} - \bar{x}_T)}{\frac{\hat{\sigma}(x)}{\sqrt{T}}} \;\leq\; z\right)$$

Translates into bootstrap 100α percent CI:

$$\hat{I} \;=\; [\bar{x}_T - \hat{u}_{(1+\alpha)/2}\frac{\hat{\sigma}(x)}{\sqrt{T}}, \bar{x}_T - \hat{u}_{(1-\alpha)/2}\frac{\hat{\sigma}(x)}{\sqrt{T}}]$$

$$\textit{where } P\left(\frac{(\bar{x}_T^{(j)} - \bar{x}_T)}{\frac{\hat{\sigma}(x)}{\sqrt{T}}} \;\leq\; \hat{u}_\alpha\right) \;=\; \hat{H}(\hat{u}_\alpha) \;=\; \alpha$$



"Percentile-t" Bootstrap

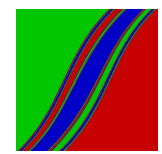
$$S = \frac{(\bar{x}_T - \mu)}{\frac{\hat{\sigma}(x)}{\sqrt{T}}}$$

$$H(z) = P\left(\frac{(\bar{x}_T - \mu)}{\frac{\hat{\sigma}(x)}{\sqrt{T}}} \leq z\right)$$

$$\hat{H}(z) = P\left(\frac{(\bar{x}_T^{(j)} - \bar{x}_T)}{\frac{\hat{\sigma}(x^{(j)})}{\sqrt{T}}} \leq z\right)$$

$$\hat{I} = [\bar{x}_T - \hat{u}_{(1+\alpha)/2}\frac{\hat{\sigma}(x)}{\sqrt{T}},\ \bar{x}_T - \hat{u}_{(1-\alpha)/2}\frac{\hat{\sigma}(x)}{\sqrt{T}}]$$

$$P\left(\frac{(\bar{x}_T^{(j)} - \bar{x}_T)}{\frac{\hat{\sigma}(x^{(j)})}{\sqrt{T}}} \leq \hat{u}_\alpha\right) = \alpha$$

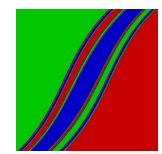


Percentile-t Bootstrap, Continued

Key insight:

Percentile: $\bar{x}_T^{(j)}$ changes across bootstrap replications

Percentile-t: both $\bar{x}_T^{(j)}$ *and* $\hat{\sigma}(x^{(j)})$ change across bootstrap replications



Key Bootstrap Property: Asymptotic Validity

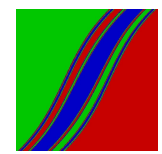
Real-world root:

$$S \xrightarrow{d} D \quad (\text{as } T \to \infty)$$

Bootstrap-world root:

$$S^* \xrightarrow{d} D^* \quad (\text{as } T, N \to \infty)$$

Bootstrap consistent ("valid," "first-order valid") if $D = D^*$.
Holds under regularity conditions.



Aren't There Simpler ways to do Asymptotically-Valid Inference for the Mean?

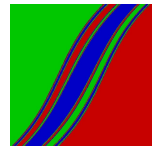
Of course. *But*:

1. Bootstrap idea extends mechanically to much more complicated cutting-edge econometric models

2. Bootstrap can deliver higher-order refinements (e.g., percentile-t)

3. Monte Carlo indicates that bootstrap often does very well in finite samples (not unrelated to 2, but does not require 2)

4. Many variations and extensions of the basic bootstraps



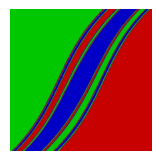
Percentile Bootstrap for Stationary Time Series

Before:

1. Use $S = \frac{(\bar{x}_T - \mu)}{\frac{\hat{\sigma}(x)}{\sqrt{T}}}$
2. Draw $\{x_t^{(j)}\}_{t=1}^{T}$ with replacement from $\{x_t\}_{t=1}^{T}$

Issues:

1. Inappropriate standardization of S for dynamic data. So replace $\hat{\sigma}(x)$ with $2\pi f_x^*(0)$, where $f_x^*(0)$ is a consistent estimator of the spectral density of x at frequency 0.
2. Inappropriate to draw $\{x_t^{(j)}\}_{t=1}^{T}$ with replacement for dynamic data. What to do?



Non-Parametric Solution: Block Bootstrap

Overlapping blocks of size b in the sample path:

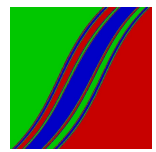
$$\xi_t = (x_t, ..., x_{t+b-1}), t = 1, ..., T - b + 1$$

Draw k blocks (where $T = kb$) from $\{\xi_t\}_{t=1}^{T-b+1}$:

$$\xi_1^{(j)}, ..., \xi_k^{(j)}$$

Concatenate: $(x_1^{(j)}, ..., x_T^{(j)}) = (\xi_1^{(j)} ... \xi_k^{(j)})$

Consistent if $b \to \infty$ *as* $T \to \infty$ *with* $b/T \to 0$

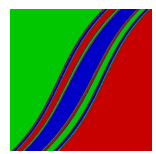


Parametric Solution: Parametric Bootstrap

$AR(1)$ example:

$$x_t = \phi x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \ iid$$

1. Regress $x_t \rightarrow x_{t-1}$ to get $\hat{\phi}$, and save residuals, $\{e_t\}_{t=1}^T$
2. Draw $\{\varepsilon_t^{(j)}\}_{t=1}^T$ with replacement from $\{e_t\}_{t=1}^T$
3. Draw $x_0^{(j)}$ from $\{x_t\}_{t=1}^T$
4. Generate $x_t^{(j)} = \hat{\phi} x_{t-1}^{(j)} + \varepsilon_t^{(j)}, t = 1, ..., T$
5. Regress $x_t^{(j)} \rightarrow x_{t-1}^{(j)}$ to get $\hat{\phi}^{(j)}$
6. Repeat $j = 1, ..., R$, and build up the distribution of $\hat{\phi}^{(j)}$



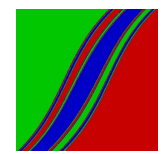
General State-Space Parametric Time Series Bootstrap

Recall the prediction-error state space representation:

$$a_{t+1/t} = Ta_{t/t-1} + TK_t v_t$$

$$y_t = Za_{t/t-1} + v_t$$

1. At the estimated system parameter values $\hat{\theta}$, run the Kalman filter to get the corresponding 1-step-ahead prediction errors $v_t \sim (0, F_t)$ and standardize them to $u_t = \Omega_t^{-1} v_t \sim (0, I)$, where $\Omega_t \Omega_t' = F_t$.
2. Draw $\{u_t^{(j)}\}_{t=1}^T$ with replacement from $\{u_t\}_{t=1}^T$ and convert to $\{v_t^{(j)}\}_{t=1}^T = \{\Omega_t u_t^{(j)}\}_{t=1}^T$.
3. Using the prediction-error draw $\{v_t^{(j)}\}_{t=1}^T$, simulate the model, obtaining $\{y_t^{(j)}\}_{t=1}^T$.
4. Estimate the model, obtaining $\hat{\theta}^{(j)}$.
5. Repeat $j = 1, ..., R$, build up the distribution of $\hat{\theta}^{(j)}$.



Some Contrasts, for Reflection...

– Wold-Wiener-Kolmogorov vs. state space

– Univariate vs. multivariate

– Time Domain vs. frequency domain

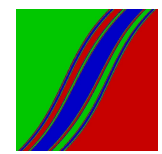
– Frequentist MLE vs. Bayesian posterior analysis

– Linear Gaussian vs. non-linear non-Gaussian

– Conditional mean vs. conditional variance

– Stationary vs. non-stationary

– Simulation sprinkled throughout
(incl. MCMC for Bayesian analysis, Monte Carlo and variance reduction, global optimization, simulated MM, bootstrap, ...)



For Reference: Hamilton TOC