Giacomini.

DGP, Forecasts?

How make inference about Y_t Dynamic Behavior

in Review : Y_t interest of

y_t - observation

$(Y_1 \dots Y_T)$ sample, iid ⇒ crossection.

DGP ≠ Model

true + no way we can get to know it.

time series: dependent, may be not identical.

Properties: p.d.f Y_t: $f(y_t)$ unconditional density

$f(y_t | \Omega_{t-1})$ conditional on history.

focus on first 2 moments, „information set"

$$E(y_t | \Omega_{t-1}) \quad Var(y_t | \Omega_{t-1})$$

* Parametric models: $E(y_t | \Omega_{t-1}) = g(\Omega_{t-1}, \beta)$ $\hat\beta$ sample estimate

↑population parameters

Asy. distr. theory: $\hat\beta$? $T \to \infty$

Properties: consistency, asy normality

$\hat\beta \xrightarrow{p} \beta$ (LLN) $?(\hat\beta - \beta) \sim N$ (CLT)

„Dependence

A: Heterogeneity ⇒ 1) $\bar{y}_T \xrightarrow[T\to\infty]{p} \mu$

$\bar{y}_T = \frac{1}{T} \Sigma y_t$, $\mu = E(y_t)$

2) $\sqrt{T}(\bar{y}_T - \mu) \xrightarrow{d} N(0, Var(\sqrt{T}, \bar{y}_T))$

Heterogeneity: how much does the uncondit. distr. change over time.

Dependence : how much y_t depends on its past.

	Homo	Heter.
Indep	iid	inid
Dep.	station ergod	MIXING 1990+

→ loosest A, giving CLT.

↑ stationary ergodicity.

Strict Stationarity - joint distrib. → needed for CLT

$f(y_t, y_{t+j_1} \dots y_{t+j_n})$ does not depend on t.

Covariance Stationarity - uncondit mean and var do not depend on t.

defined only if already cov-stat. → Ergodicity: γ_j goes to 0 fast enough.

$\gamma_j \leftarrow E(y_t - \mu)(y_{t-j} - \mu)$

Mixing - asy. independence $f(y_t, y_{t+j}) \xrightarrow[j\to\infty]{} f \cdot f$ (?)

(zillion times)

OLS in T-S: $y_t = \underset{(K\times 1)}{x_t'}\beta + u_t$ $\hat\beta = \beta + \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right)$

Prop: 1) $y_T \xrightarrow{d} y$, $g(\cdot)$ cont ⇒ $g(y_t) \xrightarrow{d} g(y)$

2) $y_T \xrightarrow{d} y$, $x_T \xrightarrow{p} c$, ⇒ $y_T + x_T \xrightarrow{d} y + c$, $y_T \cdot x_T \xrightarrow{d} y \cdot c$

3) $X \sim N(0, \Omega)$, $\underset{k\times k}{A}$ matrix, $AX \sim N(0, A\Omega A')$

Th (Consistency of $\hat\beta$) 1) $\frac{1}{T}\sum X_t X_t'$ and $\frac{1}{T}\sum X_t u_t$ Satisfy LLN
2) $E(X_t u_t)=0$ $\Rightarrow \hat\beta \xrightarrow{P} \beta$

Th (Asy normality) 1) $\frac{1}{T}\sum X_t X_t'$ satisfy LLN
2) $\frac{1}{T}\sum X_t u_t$ sat. CLT
3) $E(X_t u_t)=0$ $\Rightarrow \sqrt{T}(\hat\beta-\beta)\to N(0, Var(\sqrt{T}\hat\beta))$

$\frac{1}{T}\sum X X' \to \Omega_x$ $\quad \sqrt{T}\frac{1}{T}\sum X_t u_t \to N(0, Var(\frac{1}{\sqrt{T}}\sum X_t u_t)) \Rightarrow$ $\quad \overset{\|}{\Omega_x^{-1} Var(\frac{1}{\sqrt{T}}\sum X u)\,\Omega_x^{-1}}$

Asymptotic Variance Estimators

CASE 1: Cross-section with cond. homoskedasticity.
(X_t, u_t) are iid, $E(u_t^2|X_t)=\sigma^2$

Estimator! $Var(\sqrt{T}\hat\beta) = \Omega_x^{-1}\, \underset{\frac{1}{T}\sum E(u_t^2 x_t x_t')}{\underbrace{Var(\frac{1}{\sqrt{T}}\sum X_t u_t)}}\, \Omega_x^{-1} = \Omega_x^{-1}\sigma^2$ never use!

$$V_\beta = \widehat{Var}(\hat\beta) = \frac{1}{T}\underbrace{\left(\frac{1}{T}\sum \hat u_t^2\right)\left(\frac{1}{T}\sum X_t X_t'\right)^{-1}}_{\text{same}}$$

CASE 2: cross-section with heterosked. $(\sigma^2 \to \Sigma)$

White
$$V_\beta = \widehat{Var}(\hat\beta) = \frac{1}{T}\left(\frac{1}{T}\sum X_t X_t'\right)^{-1}\left(\frac{1}{T}\sum \hat u_t^2 X_t X_t'\right)\left(\frac{1}{T}\sum X_t X_t'\right)^{-1}$$

CASE 3: Time series. (Dependent $\Rightarrow$ stat-erg, or mixing)

$Var(\frac{1}{\sqrt{T}}\sum X_t u_t) = \frac{1}{T}\sum Var(X_t u_t) + cov\ldots =$ Newey-West, (heterot Acorr - HAC) (30 papers etc.)

Var + cov $\Leftarrow$ does not split

choose q ex: (Andrews 91)
$$\widehat{Var}(\hat\beta) = \frac{1}{T}\left(\frac{1}{T}\sum X_t X_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t' \hat u_t^2 + \right.$$
$$\left. + \frac{1}{T}\sum_{j=1}^{q}\underbrace{\left(1-\frac{j}{q+1}\right)}_{\text{weights or kernels}}\sum_{t=j+1}^{T}\left(X_t \hat u_t \hat u_{t-j} X_{t-j}' + X_{t-j}\hat u_{t-j}\hat u_t X_t'\right)\right)\cdot\left(\frac{1}{T}\sum X_t X_t'\right)$$

Rule of Thumb $q=[\sqrt[3]{T}]$

Conditional mean modelling (ARMA)

Building block: white noise ε_t: s.t. $E\varepsilon_t=0$, $V(\varepsilon_t)=\sigma^2$, $Cov(\varepsilon_t,\varepsilon_\tau)=0$ $t\neq\tau$

MA(1) $y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$ $\to$ always cov-stationary.

$\to$ Moments: $Ey_t=\mu$ $\quad Var\, y_t = (1+\theta^2)\sigma^2 = E(y_t-\mu)^2$

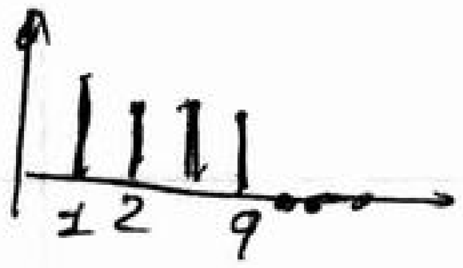
$\to$ Autocov-ces $\gamma_1 = E(\underset{\varepsilon_t+\theta\varepsilon_{t-1}}{y_t-\mu})(\underset{\varepsilon_{t-1}+\theta\varepsilon_{t-2}}{y_{t-1}-\mu}) = \theta\sigma^2$ $\quad \gamma_j=0$ $j>1$

$\to$ autocorr-s
ACgram $\rho_j = \frac{\gamma_j}{\gamma_0} = \left\{\frac{\theta}{1+\theta^2},\ j=1;\ 0,\ j>1\right\}$

Problem of identification: θ and $\frac{1}{\theta}$ both fit.

~~But: it doesn't matter.~~ for forecasting.

Similarly: MA(q) $y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$



MA(∞) $y_t = \mu + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$ $(\theta_0 = 1)$

well defined if $\sum_{j=0}^{\infty} |\theta_j| < \infty$
(and cov-stationary)

AR(1): $y_t = c + \phi y_{t-1} + \varepsilon_t$

$(1-\phi L) y_t = c + \varepsilon_t$ → is cov-stationary if roots of the polinomial are outside the unit circle.

$E(y_t) = \frac{c}{1-\phi}$

$Var(y_t) = \frac{\sigma^2}{1-\phi^2}$ $y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t$ $1 - \phi z = 0$ $z = 1/\phi \Rightarrow \left|\frac{1}{\phi}\right| > 1 \Leftrightarrow |\phi| < 1 \Leftrightarrow$ cov stationary.

(obviously:-))

$\gamma_0 = Var(y_t) = E\left(y_t - \frac{c}{1-\phi}\right)^2 = \phi^2 \gamma_0 + \sigma^2$

$\gamma_j = E(y_t - \mu)(y_{t-j} - \mu) = \frac{\phi^j}{1-\phi^2}\sigma^2 = \phi \gamma_{j-1}$ $\quad = \rho_j = \frac{\gamma_j}{\gamma_0} = \phi^j$

$\phi(y_{t-1} - \mu) - \varepsilon_t \dots$

AR(p): $(1 - \phi_1 L - \dots - \phi_p L^p) y_t = c + \varepsilon_t$

ARMA(p,q): $(1 - \phi_1 L - \dots - \phi_p L^p) y_t = c + (1 + \theta_1 L + \dots + \theta_q L^q)\varepsilon_t$.

(for MA) Invertibility of Lag polynomials:

$|\phi| < 1 \Rightarrow (1 - \phi L)^{-1} = 1 + \phi L + \phi^2 L^2 + \dots$

Thus 1) Cov-stationary AR(p) is always stationary $\Rightarrow$ ($\Leftrightarrow$) MA(∞)

2) same condition for MA(q) $\Rightarrow$ invertible to AR(∞)

Wald Th: Any cov-stationary TS $\Leftrightarrow$ determ. trend + MA(∞)

Lecture 2

Partial Autocorrelation for cov-stationary

The last coef. in a regression of y_t on m of its lags.

$y_t - \mu = \alpha_1 (y_{t-1} - \mu) + \dots + \alpha_m (y_{t-m} - \mu) + u_t$

1) If $y_t \sim AR(p)$ $\alpha_m = 0$ $\forall m > p$.

2) If $y_t \sim MA(q)$ with roots outside unit circle $\Rightarrow$ can invert to AR(∞) $\Rightarrow \alpha_m \xrightarrow{m \to \infty} 0$

Box-Jenkins model identif. strategy.

1) ACF, PACF OLS from $\hat{\alpha}_m$

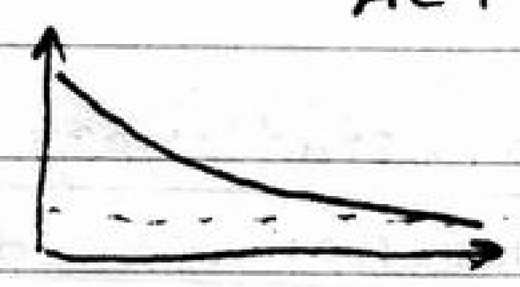
$\frac{1}{T-j}$? $\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0}$ $\quad \hat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} (y_t - \bar{y})(y_{t-j} - \bar{y})$

2) 95% conf intervals

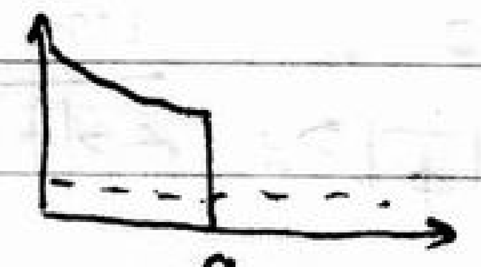
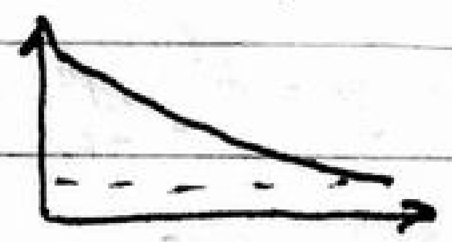
ACF: $\hat{\rho}_j \pm \frac{2}{\sqrt{T}}$ (valid under H_0: y_t - w.noise)

PACF: $\hat{\alpha}_m \pm \frac{2}{\sqrt{T}}$ (valid under H_0: $y_t \sim AR(p), p < m$)

ACF PACF

AR(p)   Intuition:

p

MA(q)  

q

Estimate and Test ARMA models

$$(1-\phi_1 L - \dots - \phi_p L) y_t = c + (1 + \psi_1 L + \dots + \psi_q L^q)\varepsilon_t$$

$\theta = (\phi_1 \dots \phi_p, \psi_1 \dots \psi_q, c, \sigma^2_\varepsilon)$ - estimate by ML

ex) AR(1) $y_t = c + \phi y_{t-1} + \varepsilon_t$.

ML: $f_{y_T \dots y_2 | y_1} = f_{y_T | y_{T-1} \dots y_1} \cdot f_{y_{T-1} | y_{T-2} \dots y_1} \cdots f_{y_2 | y_1}$ | log( )

(a) $\mathcal{L}(\theta) = \sum_{t=2}^{T} \log f_{y_t | \Omega_{t-1}}$ Typically: $Y_t | \Omega_{t-1} \sim N(c + \phi y_{t-1}, \sigma^2)$

MA? $\Rightarrow \mathcal{L}(\theta) = \underbrace{-\frac{T-1}{2}\log 2\pi - \frac{T-1}{2}\log \sigma^2}_{const} - \underbrace{\sum_{t=2}^{T} \frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}}_{(*)}$

Values (c, ϕ) that max $\mathcal{L}(\theta) \Rightarrow$ OLS simply

ML $\to$ consist & asy normal $\forall$ density IF it is the true density only.

if non-normal, use normal (QuasiMLE)

(See White 96, Estimation Inference Specification Analysis)

Results 1) (QMLE) is still consistent $\boxed{A}$: if correctly specifies conditional mean $E(y_t | \Omega_{t-1})$

ex) MA(1) ($\Leftrightarrow AR(\infty)$, cannot regress on infinite num of lags)

Cannot be estimated by OLS.

$\hat{\varepsilon}$? $y_t = \varepsilon_t + \psi \varepsilon_{t-1}$ Normal L-L: $\mathcal{L}(\theta) = k - \frac{T}{2}\log \sigma^2 - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2}$

depends on latent variable $\varepsilon_t \Rightarrow$ eval numerically

a) grid search (few parameters) $-1 \le \psi \le 1$, $0 < \sigma^2 < M$

b) Newton-type methods (e.g. FMINCON.M)

INFERENCE and Testing

AR: use asy distr. of OLS estimator (CASE 3: HAC st. er)

(**Warning**: $\hat{\phi}$ is inconsistent in the regression $y_t = \phi y_{t-1} + u_t$ if u_t is serially correlated. ($E(x_t u_t) = 0$ is violated))

ex: $u_t = \Psi u_{t-1} + \varepsilon_t$ $\quad E(u_t y_{t-1}) = E u_{t-1}^2 \Psi \neq 0$.

not the exact ARMA $\Rightarrow \exists$ ser. correlation $\Rightarrow$ all coefs are shit (Minnesota - Michigan)
(not enough lags!)

Bottomline: HAVE ENOUGH LAGS. (In Doubts, put MA term)

$\Rightarrow$ Use asy. distr. of MLE

$\boxed{\text{Testing}}$ All done using LM test, (though can choose LM, Wald)

- $y_t = x_t' \beta + u_t$ $\quad$ • Often diagnostic tests are: $H_0: \beta_1 = 0$ β_2 unrestricted

Restricted simpler to estimate

(Always underlying assumptions are: cond. homoscedast.)
$\Updownarrow E(u_t^2 | \Omega_{t-1}) = \sigma^2$

1) Est. restricted model imposing $H_0 \Rightarrow \tilde{u}$

2) regress $\tilde{u}$ on $X = (X_1; X_2)$

3) $LM = T \cdot R^2 \leftarrow$ uncentered R^2!

Standard diagnostic test

① Residual serial correlation tests

a) Plot ACF, PACF of $\tilde{u}$

b) Breusch - Godfrey : H_0: no ser corr. in $\tilde{u}$ up to lag p.

$y_t = X_t' \beta + u_t + \delta_1 u_{t-1} + \dots + \delta_p u_{t-p} \Rightarrow H_0: \delta_1 = \dots = \delta_p = 0.$

Run $\tilde{u}_t \sim x_t' \; \tilde{u}_{t-1} \dots \tilde{u}_{t-p}$, $T \cdot R^2 \sim \chi^2_p$.

② Nonlinearity tests.

a) specific nonlinearity. $\rightarrow$ $y_t = x_t' \beta + g(\gamma, x_t) + u_t$ $\quad g(0, ?) = 0$. Same thing $H_0: \gamma = 0$

b) General nonlinearity.

1) $y \sim x_t \rightarrow \tilde{u}$ 2) $\tilde{u}$ on x and $\left.\frac{\partial g}{\partial(\beta, \gamma)}\right|_{\hat{\gamma}=0}$ $\quad$ general $\left.\frac{\partial g}{\partial(\beta,\gamma)}\right|_{\tilde{\beta},\tilde{\gamma}}$

3) $T \cdot R^2 \sim \chi^2_k$ $\quad k = \dim(\gamma)$

Problems: 1) $\frac{\partial g}{\partial \gamma}$? $\nexists$ 2) params not identified under H_0

ex$_1$: threshold AR: $y_t = \phi_1 y_{t-1} \cdot \mathbb{1}_{(y_{t-1} \le c)} + \phi_2 y_{t-2} \mathbb{1}_{(y_{t-1} > c)} + u_t$

ex$_2$: Smooth transition AR

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-1} \frac{1}{1 + e^{-\phi_3 (y_{t-1} - c)}} + \varepsilon_t$$

$\phi_2 = 0 \Rightarrow \phi_3$ not identifiable.

RESET TEST against general non-linearity.

1) ARTIFICIAL REGRESSION
$y \sim x \Rightarrow \tilde{y}$ fitted

2) $y \sim \tilde{y}^2, \tilde{y}, x$, etc. $\tilde{y}^H$
F-test on $\tilde{y}^2 \dots \tilde{y}^H$

Intuition: Taylor expansion

DGP :-)

Relaxing the assumption of Stationarity

ex: GDP: trending mean.

Stock returns: changing variance.

US productivity growth: structural breaks.

Non-linear models: Markov switching

① Trends in mean

RBC (Nelson-Plosser trends → perm. but stochastic)

a) Deterministic: $y_t = c_0 + c_1 t + \psi(L)\varepsilon_t$

$Ey_t = c_0 + c_1 t$ → trend stationary → can subtract

b) Stochastic trends: unit root processes

tricky slippery slopes

$\phi(L) y_t = \varepsilon_t$ (has unit root: $\phi(z) = 0 \ni z = 1$)

Long memory: every shock will be remembered forever

Transform data: 1) estimate $y_t = \alpha + \beta t + u_t$, take u_t.

2) Hodrick-Prescott filter ("band-pass filter")

Decompose: $y_t = s_t + u_t$ (both unobservable)

s_t - trend u_t - stationary component

goal: get smooth s_t

$$\min_{\{s\}} \sum_{t=1}^{T} (y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(s_{t+1} - s_t) - (s_t - s_{t-1}) \right]^2$$

least diff given smoothness, change, penalty for lack of smoothness

$\lambda = 100$ annual

$\lambda = 1600$ quarterly

$\lambda = 14400$ monthly.

3) Take first differences $\Delta y_t = y_t - y_{t-1}$ ex: logs ⇒ growth rates (nominal rep. of real stuff)

4) Second differences (ex: prices)

$\log \frac{P_t}{P_{t-1}} - \log \frac{P_{t-1}}{P_{t-2}}$ since inflation has a unit root.

Warnings (data transf) 1) Δ if AR(1) $|\phi| < 1$

$\Delta y_t = \phi \Delta y_{t-1} + \underbrace{\Delta \varepsilon_t}_{u_t}$ → not a white noise, ⊕ non invertible MA ⇒ bad for IRF

⊕ u_t - is serially correlated ⇒ $\hat{\phi}_{OLS}$ inconsistent

dampens LR

2) HP x to unit-root data

TS: leaves s.r. unaltered

Resulting series is stationary (King-Rebelo 93)

UR: dampens both.

introduces fake (RBC-like) dynamics. (Cogley, Nason 95)

Simulate ⇒ looks better than it really does, due to filter

Structural Change.

- Violation of stationarity: parameters are unstable.

ex) $y_t = c + \phi y_{t-1} + u_t \qquad \theta = (c, \phi, \sigma_u^2)$

Structural Break: at least 1 component changed at some break date.

a) Change in unconditional mean: $\mu = \frac{c}{1-\phi}$ (most relevant case)

b) break in $\phi \Leftrightarrow$ change in persistence.

c) change in volatility σ_u^2 ("great moderation")

Questions: 1) Tests for struct. break at known date.

2) at unknown date.

3) multiple breaks

4) estimate timing.

① Known date (Chow 1960)

$y_t = x_t'\beta + u_t \quad (*)$ $\qquad T_0 \begin{array}{l}\nearrow \beta_R \\ \searrow \beta_{UR}\end{array}$ $\qquad X_B = [x_1 \ldots x_{T_0}]$, $X_A = [x_{T_0+1} \ldots x_T]$.

$$\begin{pmatrix} Y_B \\ Y_A \end{pmatrix} = \underbrace{\begin{pmatrix} X_B & 0 \\ 0 & X_A \end{pmatrix}}_{T \times 2k} \begin{pmatrix} \beta_B \\ \beta_A \end{pmatrix} + \begin{pmatrix} u \\ u \end{pmatrix} \quad (**)$$

makes ass. of homoskedastic er.

$H_0: \beta_B = \beta_A$ (no break)

a) $F = \frac{(u_R'u_R - u_u'u_u)/\cancel{k}}{u_u'u_u / T-2k} \overset{asy}{\sim} \chi^2_k$ or $1/k \sim F(k, T-2k)$

b) use heterosced-robust st. er.

$H_0: R\beta = 0: \quad \underset{k \times 2k}{R} = [I_k, -I_k]$

$$W = (R\hat\beta)'(R \underset{\text{heter-robust}}{\widehat{Var}(\hat\beta)} R')^{-1}(R\hat\beta) \sim \chi^2_k$$

Problems: date must be known

pick from data: endogeneity.

② Unknown date (Andrews 1993) (A+Ploberger 1994)

Try all dates from $(1-\tau)$ to (τ) $\quad \tau \sim 5\%$

$\Rightarrow \sup_{T_0} F_{T_0}$ test $\rightarrow$ worst scenario.

vs critical values.

③ Timing of break

Bai (94): estimate T_0 by OLS along with β.

1) for each T_0 in the interior UR $\Rightarrow$ compute $u_u'u_u = SSR$

$\hat T_0 = \min_T u'u$

Corresponds to supF date only if 1) model linear. + homoscedasticity.

④ Multiple Breaks (Bai-Perron 1998): Sequential 1 break $\Rightarrow$ subsamples etc. consistent though contraversial.

Extensions (Bai 94): confidence intervals for the break date.

The sample is trimmed

End-of-sample breaks more interesting (Andrews 2003) can we distinguish these 2 alternatives

Unit Root Processes

$$(1-\phi_1 L - \ldots - \phi_p L^p)\, y_t = (1+\varphi_1 L + \ldots + \varphi_q L^q)\, \varepsilon_t.$$

Unit root: $\lambda = 1$ (only one) and the rest are outside the unit circle

$\prod_{i=1}^{p} (1-\lambda_i z_i)$

Def | $I(0)$ – cov-stationary $I(1)$ – has unit root.

$I(2) \to$ ex: inflation.

Can you prove that it's useful

Def | ARIMA (p, d, q) if $(1-L)^d y_t$ is stationary, ARMA(p,q)

$d = \{0; 1; 2; [0, \frac{1}{2}]\}$ → ARFIMA (fractional integration (long-memory))

Peter Robinson

Facts:

1) Parameters can be estimated by OLS, but S.E. are different.

2) LLN & CLT Fail.

3) Processes: (Asy depends on exact ass's)

DGP: RW: $y_t = y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid\, N(0, \sigma^2)$, $y_0 = 0$

$y_t = \phi y_{t-1} + u_t \qquad \hat\phi = \dfrac{\sum y_{t-1} y_t}{\sum y_{t-1}^2}$

If $|\phi| < 1$ $\sqrt{T}(\hat\phi - \phi) \xrightarrow{d} N(0, 1-\phi^2)$

If $\phi = 1 \Rightarrow$ $\sqrt{T}(\hat\phi - 1)$ degenerate, scale up, $T(\hat\phi - 1) \xrightarrow{d} DF$

Superconsistent

Def (Brownian motion): $W(\cdot)$ – contin-time stoch process, associates a date $t \in [0, 1]$ with $W(t)$ s.t.

1) $W(0) = 0$

2) $\forall\, 0 \le t_1 \le t_2 \le \ldots < t_k \le 1$ $\quad W(t_j) - W(t_{j-1})$ – indep. norm. r.v.

Functional CLT : let $\varepsilon_t \sim iid(0, \sigma^2) \Rightarrow \forall r \in [0, 1]$ construct.

$$X_T(r) = \frac{1}{T} \sum_{t=1}^{[T \cdot r]} \varepsilon_t, \quad [T \cdot r] = \text{integer part}.$$

$\sqrt{T}\, \dfrac{X_T(\cdot)}{\sigma} \xrightarrow{d} W(\cdot)$ [random functions]

Contin. mapping theorem:

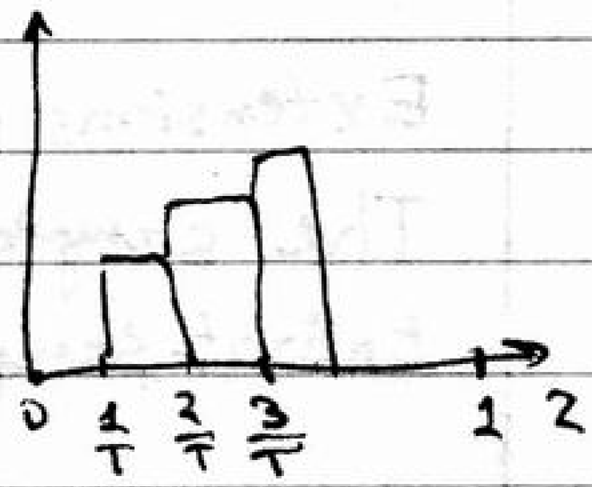
$S_T(\cdot) \xrightarrow{d} S(\cdot)$, $g(\cdot) \subset C^0$, $\Rightarrow g(S_T(\cdot)) \xrightarrow{d} g(S(\cdot))$

Simplest case : RW without drift. $\quad \hat\phi = 1 + \dfrac{\frac{1}{T}\sum y_{t-1}\varepsilon_t}{\frac{1}{T}\sum y_{t-1}^2}$

$$T(\hat\phi - 1) = \frac{\frac{1}{T}\sum y_{t-1}\varepsilon \xrightarrow{d} \frac{\sigma^2}{2}\{W(1)^2 - 1\}}{\frac{1}{T^2}\sum y_{t-1}^2 \xrightarrow{d} \sigma^2 \int^1 W(r)^2 dr} = DF$$

▶ $X_T(r)^2 = \left(\frac{1}{T}\sum_{1}^{Tr} \varepsilon_t\right)^2 = \begin{cases} 0, & 0 \le r < \frac{1}{T} \\ \frac{\varepsilon_1^2}{T^2}, & \frac{1}{T} \le r < \frac{2}{T} \\ \ldots \\ \frac{(\varepsilon_1 + \ldots + \varepsilon_T)^2}{T^2}, & r = 1 \end{cases} = \begin{cases} 0 \\ \frac{y_1^2}{T^2} \\ \\ \frac{y_T^2}{T^2} \end{cases}$

$y_T = \varepsilon_1 + \ldots + \varepsilon_T$



$$\Rightarrow \int_0^1 (X_T(r))^2 dr = \sum_{t=1}^{T} \frac{y_{t-1}^2}{T^3} \Rightarrow \frac{1}{T^2}\sum y_{t-1}^2 = \int_0^1 (\sqrt{T}\, X_T(r))^2 dr \xrightarrow[d]{\;\sigma W(r)\;} \xrightarrow{FCLT} \sigma^2 \int_0^1 W(r)^2 dr.$$

Trend-stationary: $y_t = \delta t + \varepsilon_t \Rightarrow OLS \Rightarrow \hat\delta = \delta + \dfrac{\frac{1}{T^2}\frac{1}{T}\sum t\varepsilon}{\frac{1}{T^3}\sum t^2}$, $\quad \frac{1}{T^3}\sum t^2 = \frac{1}{T^3}\frac{T(T+1)(2T+1)}{6} \to \frac{1}{3}.$

$$\Rightarrow \boxed{T^{3/2}(\hat\delta - \delta) \to N(0, 3\sigma^2)}$$

$\frac{1}{\sqrt{T}}\left[\sum t\varepsilon_t / T\right] \xrightarrow{\text{marting. diff CLT}} N(0, \sigma^2/3)$ $\quad$ $\left[E(y_t^2) = \sigma_t^2 \text{ s.t. } \frac{1}{T}\sum \sigma_t^2 < \infty \to \sigma^2,\ E y_t = \mu \Rightarrow \frac{1}{\sqrt{T}}(\sum y_t - \mu) \xrightarrow{d} N(0, \sigma^2)\right]$

$\frac{1}{T^3}\sum t^2 \to 1/3.$

$E\left(\frac{t}{T}\varepsilon_t \mid \Omega_{t-1}\right) = 0 = \mu$ $\quad$ $\frac{1}{T}\sum \frac{t^2}{T^2}\sigma^2 = \frac{1}{T^3}\sum t^2 \sigma^2 = \frac{\sigma^2}{3}$

$E\left(\frac{t^2\varepsilon_t^2}{T^2} \mid \Omega_{t-1}\right) = \frac{t^2\sigma^2}{T^2}$

But: OLS t-test turns out to be correct.

$$t = \frac{\hat\delta - \delta_0}{\sqrt{s_T^2 / \sum t^2}} \left[s_T^2 = \frac{1}{T-1}\sum (y_t - \hat\delta t)^2\right] = \frac{(\hat\delta - \delta) T^{3/2}}{\sqrt{\underbrace{s_T^2}_{\to \sigma^2} \underbrace{\left(\frac{\sum t^2}{T^3}\right)^{-1}}_{\to (\frac{1}{3})^{-1}}}} \approx \frac{T^{3/2}(\hat\delta - \delta_0)}{\sqrt{3\sigma^2}} \to N(0,1)$$

One of 157 tests: ADF test:

$AR(p)$: $(1 - \phi_1 L - \dots - \phi_p L^p) y_t = \varepsilon_t$, $\varepsilon_t \sim iid\ N(0, \sigma^2)$

$y_t = \rho y_{t-1} + \alpha_1 \Delta y_{t-1} + \dots + \alpha_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (**)$

$\rho = \phi_1 + \dots + \phi_p$ and $\alpha_j = -(\phi_{j+1} + \dots + \phi_p)$, $j = \overline{1, p-1}$

has unit root $\Rightarrow 1 - \phi_1 \cdot 1 - \dots - \phi_p \cdot 1^p = 0 \Rightarrow \rho = 1$.

Test: $H_0: \rho = 1$, $H_1: \rho < 1$. $\Rightarrow$ Estimate $(**)$ by OLS

$$t_{ADF} = \frac{T(\hat\rho - 1)}{1 - \hat\alpha_1 - \dots - \hat\alpha_{p-1}} \quad \text{or} \quad t_{OLS} = \frac{\hat\rho - 1}{s.e.(\hat\rho)}.$$

Elliot-Rothemberg-Stock (Best power)

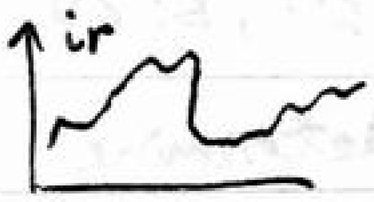
Crucial: need enough lags!

Cases: 1) $y_t = \rho y_{t-1} + \sum \alpha_j \Delta y_{t-j} + \varepsilon_t$, DGP: $\rho = 1$ $\to$ non-standard

2) $y_t = \alpha + \rho y_{t-1} + \sum \alpha_j \Delta y_{t-j} + \varepsilon_t$, DGP: $\alpha = 0$, $\rho = 1$. $\to$ non-standard.

3) ~~$y_t = \alpha + \rho$~~ <same>, DGP: $\alpha \neq 0$, $\rho = 1$. $\to$ Nice, since both asy. normal $\sqrt{T}(\hat\alpha - \alpha)$ and $T^{3/2}(\hat\rho - 1) \sim N(0, ?)$

4) $y_t = \alpha + \delta t + \rho y_{t-1} + \sum \alpha_j \Delta y_{t-j} + \varepsilon_t$, DGP: $\alpha \neq 0$, $\rho = 1$. $\to$ non-standard

Usually choose (4) if  or (2) if 

Nonlinear models of conditional mean

$E(y_t \mid \Omega_t) = g(\underbrace{x_{t-1}}_{k \times 1 \text{ vector}})$

1) Threshold: $TAR(p)$ $\quad$ ex. business cycles

$$y_t = \begin{cases} c_1 + \phi_1 y_{t-1} + u_{1t} & \text{if } w_t \le c \\ c_2 + \phi_2 y_{t-1} + u_{2t} & \text{if } w_t > c \end{cases}$$

$\Updownarrow$

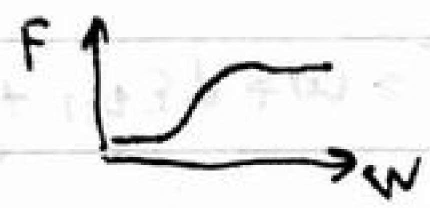
$$y_t = x_{t-1}'\theta_1 + x_{t-1}'(\theta_2 - \theta_1)\mathbb{1}(w_t > c) + u_t$$

 F transition is non-smooth

2) Self-Exciting (SETAR)

same with $w_t = y_{t-d}$

3) Smooth Transition (STAR)

 ex: logit $F(w) = \dfrac{1}{1 + e^{-\gamma(w_t - c)}}$, $\gamma > 0$

Good if c is exog; Poor if c is endog

4) w_t is unobserved

$w_t = s_t$ represents „regimes"

Markov-Switching Models

$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$

$s_t \sim$ 2-state Markov chain, indep. of ε_t

$s_t \in \{1, 2\}$, $Pr(s_t \mid s_{t-1})$ only $= p_{ij}$ ("j" "i")

Can have n states and d-th order M.C.

Estimation:

1) know threshold c $\Rightarrow$ use OLS + Dummy Var's

2) Otherwise $\Rightarrow$ ML

Caution: models tend to choose too

1) many regimes

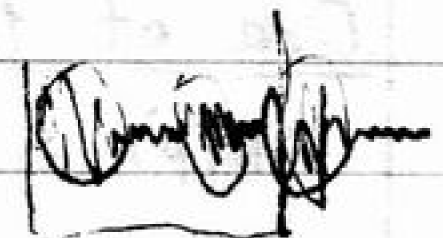
2) Numerically unstable

3) Poor forecasting performance (w.r.t. random walk)

Nonlinear Models of Cond. Variance

$y_t = E(y_t \mid \Omega_{t-1}) + \varepsilon_t, \quad iid(0, \sigma^2)$ $\qquad E(\varepsilon_t^2 \mid \Omega_{t-1}) = \sigma_t^2 = \text{constant}.$

A: All other aspects of distr. do not change over time.

Not true in data: prices, investments, dividends etc, s&P 

Stylized FACTS: 1) Time-varying volatility 2) Volatility clustering.

- General Model: $(*)\ y_t = \mu_t + \underbrace{\sigma_t z_t}_{\varepsilon_t}, \quad z_t \sim iid(0,1)$

cond. skewness $\leftarrow$ i.e. $z_t = \frac{y_t - \mu_t}{\sigma_t}$ iid (cond mean) (cond var) $\to$ location-scale model $\to$ mean, $\to$ var-ce | else const

Parametric Models of σ_t

① ARCH(q): $\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2$ Engle (82) $\quad \alpha_j \geq 0, \omega > 0.$

autoreg. cond heteroskedasticity.

$\nu_t = \varepsilon_t^2 - \sigma_t^2 \quad E(\nu_t \mid \Omega_{t-1}) = 0. \Rightarrow \varepsilon_t^2 = \omega + \sum \alpha_j \varepsilon_{t-j}^2 + \nu_t$ (w.n.)

a) ARCH(q) is cov-stat-nary if $(1 - \alpha_1 z - \dots - \alpha_q z^q) = 0$ has root outside unit circle

b) (a) $\Rightarrow$ $Var\, \varepsilon_t = \frac{\omega}{1 - \alpha_1 - \dots - \alpha_q}$ $\qquad$ c) captures vol. clustering $\varepsilon_t^2 \uparrow \Rightarrow \varepsilon_{t+1}^2 \uparrow$

In practise: a lot of lags are important.

② GARCH(p,q) $\quad \sigma_t^2 = \omega + \sum_1^q \alpha_j \cdot \varepsilon_{t-j}^2 + \sum_1^p \beta_j \sigma_{t-j}^2$

- $\omega > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$.
- $\varepsilon_t^2 = \omega + \alpha(L)\varepsilon_{t-1}^2 + \beta(L)(\varepsilon_{t-1}^2 - \nu_{t-1}) + \nu_t$
$= \omega + (\alpha(L) + \beta(L))\varepsilon_{t-1}^2 + \nu_t - \beta(L)\nu_{t-1}$

$\therefore$ both α & β determine stationarity

ARMA(max(p,q), q) model for ε_t^2

a) cov stat. if $(1 - \alpha(z) - \beta(z)) = 0$ are outside unit circle

b) GARCH(1,1) always enough. (must have $\alpha + \beta < 1$)

c) closer to 1 $\Rightarrow$ higher clustering.

③ Integrated GARCH(p,q): $\alpha(1) + \beta(1) = 1$. : shocks to volatility persistent.

④ EGARCH, TARCH. $-\ \uparrow\downarrow \varepsilon^2$ diff. effects $\Rightarrow$ nonlinearity (asymmetric).

exponential transform

TARCH: $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \underbrace{\lambda}_{>0} D_{t-1} \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad D = \begin{cases} 1, & \varepsilon_{t-1} < 0 \\ 0, & \geq 0. \end{cases}$

threshold

$\lambda > 0 \Rightarrow$ neg. shocks higher

(in mean)

⑤ ARCH-M: trade-off betw. exp. returns and variance. $E(y|\Omega) = \mu + \delta \sigma_t^2$.

INFERENCE in GARCH.

1) Test for ARCH in residuals of regression.

$H_0: E(\varepsilon_t^2 | \Omega_{t-1}) = \sigma^2 \quad H_1: ARCH(q)$ similar to Breusch-Godfrey.

Engle(82) LM test: $\hat{\varepsilon}^2 = c_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \ldots + \alpha_q \hat{\varepsilon}_{t-q}^2 + u_t$ Test: $\alpha_j = 0 \ \forall j$ by $TR^2 \sim \chi_q^2$.

2) Estimation : ML estimation under location-scale assumption.

$y_t = \mu_t + \underbrace{\sigma_t z_t}_{\varepsilon_t}$, $z_t \sim iid(0,1)$. $\Theta = (\underset{\text{of } \mu \& \sigma}{\text{params}}, \nu)$

Let $f_z(z_t, \nu)$ = density vs mean(0) var(1) and nuisance parameter ν

ex: Normal (no ν), Student (ν-deg. of freedom) – fat tails, skewed-t (Bruce Hansen)

$f_y(y_t) = f_z(z)\frac{\partial z}{\partial y} = f_z\left(\frac{y_t - \mu_t}{\sigma_t}\right)\frac{1}{\sigma_t}$

$\mathcal{L}(\Theta) = \sum_{t=1}^T \log f_z\left(\frac{y_t - \mu_t}{\sigma_t}\right) - \frac{1}{2}\sum_{t=1}^T \log \sigma_t^2$

ex: Daily data: $y_t = \sigma_t z_t$, $z_t \sim N(0,1)$, GARCH (1, 1) $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$ deterministic.

$\mathcal{L}(\theta) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum \frac{y_t^2}{\sigma_t^2} - \frac{1}{2}\sum \log \sigma_t^2$

In practice: 1) compute sequence σ_t^2 given $(y_1 \ldots y_T)$ and params (ω, α, β)

$\varepsilon_t = y_t$, so $\sigma_t^2 = \omega + \alpha \underset{\text{"}y_{t-1}^2\text{"}}{\varepsilon_{t-1}^2} + \beta \sigma_{t-1}^2$, initialize $\sigma_1^2 = y_1^2$

2) Write likelihood function | all cond. on y_1. $\mathcal{L} = -\frac{T-1}{2}\log 2\pi - \frac{1}{2}\sum_2^T \frac{y_t^2}{\sigma_t^2} - \frac{1}{2}\sum_2^T \log \sigma_t^2$.

3) $\omega \sim 0.25 \quad \alpha \sim 0.15 \quad \beta = 0.75$ → initial typically.

4) constraints: $\omega > 0, \alpha \geq 0, \beta \geq 0, \alpha + \beta < 1$

5) use "fmincon"

Kevin Sheppard UCSD GARCH TOOL BOX

"Student": $\mathcal{L}(\omega, \alpha, \beta, \nu) = (T-1)\log\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}(\nu-2)^{-1/2}\right] - \frac{\nu+1}{2}\sum_2^T \log\left(1 + \frac{y_t^2}{\sigma_t^2(\nu-2)}\right) - \frac{1}{2}\sum_2^T \sigma_t^2$

$\nu_0 = 10$, $\nu > 2$.

MULTIVARIATE TIME-SERIES MODELS

$\underset{m \times 1}{Y_t}$ is a vector. : VAR(MA) (MA: never :-))

ex:) Bivariate VAR(1)

$y_{1t} = c_1 + a_{11} y_{1t-1} + a_{12} y_{2t-1} + \varepsilon_{1t}$

$y_{2t} = c_2 + a_{21} y_{1t-1} + a_{22} y_{2t-1} + \varepsilon_{2t}$.

"Reduced form": y_1 and y_2 do not have contemporaneous effects (ok for forecasting)

General VARMA(p,q): $Y_t = c + A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \varepsilon_t + B_1 \varepsilon_{t-1} + \ldots + B_q \varepsilon_{t-q}$

$\varepsilon_t \sim N(0, \Omega)$ → 1) contemp. corr. b/w ε is ok
2) no serial corr. in resids.
3) cond. homoscedasticity (no GARCH)

- VMA(q): $A_i = 0$, Always. Stationary, invertible if $|I + B_1 z + \ldots + B_q z^q| = 0$. z^* outside.
- VAR(p): $B_j = 0$, Stationary if $|I - A_1 z - \ldots - A_p z^p| = 0$ z^* outside

"Companion form": stacked lags

$$\underbrace{\begin{bmatrix} Y_t - c \\ \vdots \\ Y_{t-p+1} - c \end{bmatrix}}_{Z_t} = \underbrace{\begin{bmatrix} A_1 & \ldots & A_p \\ I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \;\; 0 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} Y_{t-1} - c \\ \vdots \\ Y_{t-p} - c \end{bmatrix}}_{Z_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{V_t}$$

$E(V_t V_t') = \begin{bmatrix} \Omega & 0 & \ldots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & & \\ 0 & \ldots & & 0 \end{bmatrix}$

Estimation: VAR $\nearrow$ OLS VARMA $\nearrow$ ML

Recursive dimensionality

1) MLE for VAR(p)

$$y_t = c + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t = D' X_t + \varepsilon_t$$

$$D = [c\ A_1 \dots A_p] \quad \underbrace{m \times (1+mp)}_{K} \qquad X_t = \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix} (1+mp)\times 1$$

$$\mathcal{L}(\theta, \Omega) = -\frac{(T-p)}{2} m \log 2\pi - \frac{1}{2}\sum_{t=p+1}^{T}\left[(y_t - D'X_t)'\Omega^{-1}(y_t - D'X_t)\right] + \frac{T-p}{2}\log|\Omega^{-1}|$$

2) OLS equation-by-equation.

get $\hat{D}$ and estimate Ω as:

1) $\hat{\Omega}_{ML} = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$ ($p\times 1$, $1\times p$)

2) $\hat{\Omega}_{OLS} = \frac{1}{T-p-(mp+1)}$ —"— fin sample better.

Baysian: int. restrictions ($m=7, p=4$ macro)

Distr. of params

$\underset{mK\times 1}{\hat{\beta}} = vec(\hat{D})$ $Q = E(X_t X_t')$

$$\sqrt{T-p}\,(\hat{\beta} - \beta) \sim N(0, \Omega \otimes Q^{-1})$$

coefs: $\hat{\beta}_i \sim N\left(\beta_i, \hat{\sigma}_i^2 \left[\sum_{t=p+1}^{T} x_t x_t'\right]^{-1}\right)$

$\hat{\sigma}_i^2 = \frac{1}{T-p}\sum_{p+1}^{T} \hat{\varepsilon}_{it}^2$ or $\frac{1}{T-p-k} \to$ OLS.

I.E. standard OLS. (no White!)

But: need enough lags vs "curse of dimensionality" (1 lag $\Rightarrow m^2$ param)

Lag length Selection: LR test: H_0: p_0 lags vs H_1: $p_1 > p_0$.

Test stat: $(T-p_1)\left[\log|\hat{\Omega}_0| - \log|\hat{\Omega}_1|\right] \sim \chi^2_{m^2(p_1-p_0)}$ (OLS estimates; # restr under H_0)

Granger-Causality

Do past values of x_t help explain y_t.

Original: VAR(p) $\begin{pmatrix} y_t \\ x_t \end{pmatrix} = c + A_1 \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \dots + A_p \begin{pmatrix} y_{t-p} \\ x_{t-p} \end{pmatrix} + \varepsilon$

Def $x_t \Rightarrow y_t$: lags of x_t enter eq's for y_t

x_t does not Granger cause y_t if A_i are lower-triangular.

Implementation: 1) Choose p s.t. $\varepsilon \sim$ iid.

2) univ. regression: $y_t = c + \sum_{i=1}^{p} (\alpha_i y_{t-i} + \beta_i x_{t-i})$

3) F-test H_0: $\beta_1 = \dots = \beta_p = 0$ (No G-C-ty) $F = \frac{T(SSR_0 - SSR_1)}{SSR_1}$ ($= \Sigma \tilde{u}^2$)

But: * Results sensitive to p

* Ommited variables problem. ($z \Rightarrow y$, z corr vs x $\Rightarrow$ can get $x \Rightarrow y$)

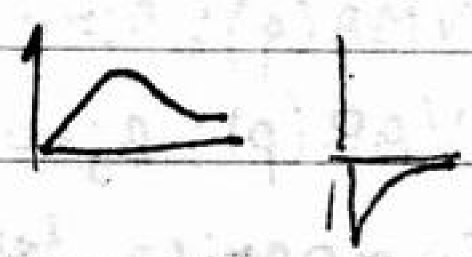
Impulse-response analysis

Response to shocks, causes of variability.

1) Write y_t in MA(∞) form.

$$y_t = \mu + \varepsilon_t + B_1 \varepsilon_{t-1} + \dots + \underbrace{B_p}_{\text{effect of shocks}} \varepsilon_{t-p} + \dots \quad \varepsilon_t \sim iid(0, \Omega)$$

$B_h = \frac{\partial y_{t+h}}{\partial \varepsilon_t'}$ $B_{ij,h}(\varepsilon_{jt} \to y_{i,t+h})$ – IRF



ex) $\lim_{h\to\infty} B_{ijh} = 0 \Leftrightarrow$ stationarity.

not captured

Problem: This does not represent causality. We need: $\frac{\partial E(y_{i,t+h} | y_{1,t}, \Omega_{t-1})}{\partial y_{1,t}}$

dynamic multipliers

In general need ordered quantities

$$\frac{\partial E(y_{i,t+h} \mid y_{1t}, \Omega_{t-1})}{\partial y_{1t}}, \quad \frac{\partial E(\quad \mid y_{1t}, y_{2t}, \Omega_{t-1})}{\partial y_{2t}} \dots \frac{\partial E(y_{i,t+h} \mid y_{1t} \dots y_{mt}\, \Omega_{t-1})}{\partial y_{mt}}.$$

$$\underset{m\times 1}{\frac{\partial E(y_{t+h} \mid y_{1t}\, \Omega_{t-1})}{\partial y_{1t}}} = \underset{m\times m}{B_h}\, \underset{m\times 1}{a_1}$$

$[a_1 \dots a_m] \leftarrow A \rightarrow$ lower-triangular (1s on diagonal)

$\underset{m\times m}{\Omega} = A D A'$ ($D \rightarrow$ diagonal)

$$\dots\dots$$

$$\underset{m\times 1}{\frac{\partial E(y_{t+h} \mid y_{1t} \dots y_{mt}\, \Omega_{t-1})}{\partial y_{mt}}} = \underset{m\times m}{B_h}\, \underset{m\times 1}{a_m}$$

A plot of element i of $B_h a_j$ as a function of h is orthogonalized IRF

it's how you revise your expectation of $y_{t+h,i}$ when learning $y_{j,t}$

But: ordering of $1 \dots m$ matters! for the „new info" concept.

In practise:

1) Estimate VAR(p) by OLS $\quad y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t.$

$\rightarrow$ get $\hat{A}_j,\ \hat{\varepsilon}_t, \Rightarrow \hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$

2) Construct. MA(∞) $\hat{B}_i$ by: $z_t = F z_{t-1} + v_t$ (companion form)

$$\underset{mp\times pm}{F = \begin{bmatrix} A_1 \dots A_p \\ I \ddots 0\, 0 \\ 0 \ \ I\, 0 \end{bmatrix}} \quad \underset{mp\times 1}{z_t = \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}} \quad \underset{mp\times m}{J = \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \end{bmatrix}}$$ Selection matrix

$y_t = J' z_t. \quad v_t = J \varepsilon_t.$

Solve. By „recursive substitution"

$$z_t = F^2 z_{t-2} + F v_{t-1} + v_t = \dots = \sum_{i=0}^{\infty} F^i v_{t-i}$$

$$y_t = J' z_t = \sum_{j=0}^{\infty} J' F^j J \varepsilon_{t-j} \Rightarrow \text{for MA}(\infty):\ \hat{B}_i = J' \hat{F}^i J$$

in practice: $\hat{B}_i = \sum_{j=1}^{i} B_{i-j} \hat{A}_j, \quad i = 1, 2 \dots$ with $B_0 = I$, $A_j = 0\ \forall j > p$
$\Rightarrow$ IRF.

3) Cholesky decomposition. of $\hat{\Omega} = P P'$ (chol(Ω)) $\Rightarrow P = A D^{1/2}$

$$A = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & \ddots & \\ \vdots & \ddots & 0 \\ a_{m1} & \dots & 1 \end{bmatrix} \quad D = \begin{bmatrix} \sqrt{d_{11}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_{mm}} \end{bmatrix}$$

i.e. take diag elements of P.
$\Rightarrow D^{1/2} \Rightarrow A = D^{-\frac{1}{2}} P$

(get rid of units of measurement).

Another way: directly from Cholesky: $\Omega = P P'$ $\quad \underset{m\times m}{B_h}\, \underset{m\times 1}{P_j}$ - effect of one s.dev. in y_j on y_i

s.dev. of $y_t = A^{-1}\varepsilon_t$ increase in y_j affects y_{t+h}.

Forecast Error Variance Decomposition

$$e_{t+h|t} = y_{t+h} - E[y_{t+h} \mid \Omega_t]$$

$$\Updownarrow$$

$$\underbrace{\varepsilon_{t+h} + B_1 \varepsilon_{t+h-1} + \dots B_{h-1} \varepsilon_{t+1}}_{\text{forecast error}} + \underbrace{B_h \varepsilon_t + B_{h+1} \varepsilon_{t-1} + \dots}_{\text{know}}$$

Quantify importance of shocks by the fractions of the variances

$$V(e_{t+h|t}) = E(e \cdot e') =$$

$V(e) = \Omega + B_1 \Omega B_1' + \ldots + B_{h-1} \Omega B_{h-1}'$

Orthogonalized shocks: $u_t = A^{-1} \varepsilon_t$

$E(u' u') = E(A^{-1} \underbrace{\varepsilon_t \varepsilon_t'}_{ADA'} (A^{-1})') = D$

$\Omega = E(\varepsilon\varepsilon') = a_1 a_1' V(u_{1t}) + \ldots + a_m a_m' V(u_{m,t})$

$Var(e) = \sum_{j=1}^{m} \{ Var(u_{jt}) [a_j a_j' + B_1 a_j a_j' B_1' + \ldots + B_{h-1} a_j a_j' B_{h-1}] \}.$

fractions $\frac{Var(u_{jt})[\quad]}{V(e)}$ - measure of importance of shocks

Standard errors for IRF:

$- B_h$ by VAR $\to$ inverted.

Let β be $m \cdot (mp+1) \times 1 \to$ VAR(p) params.

$\sqrt{T-p}(\hat\beta - \beta) \sim N(0, \underbrace{\Omega \otimes Q^{-1}}_{\Omega_\beta})$

MA(∞) $\underset{m^2 \times 1}{\pi_h} = vec(B_h)$ - nonlinear $\sim \beta$ ugly: $Var(\hat\Omega \& \hat Q)$ needed. Δ method: $\sqrt{T_p}(g(\hat\beta) - g(\beta)) \sim N(0, \frac{\partial g}{\partial \beta}' \Omega_\beta \frac{\partial g}{\partial \beta}')$

Bootstrap easier and works better: Kilion (98): CI for IRF (very poorly in finite samples)

a) Only for stationary VAR

b) In practice exponents: biases magnified greatly $\Rightarrow$ Bias correct by bootstrap.

(1) Estimate bias using resampling methods $\to$ based on iid Assumption.

VAR(p): $y_t = \hat c + \hat A_1 y_{t-1} + \ldots + \hat A_p y_{t-p} + \hat u_t$. $\quad \hat\beta = vec(\hat c, \hat A_1 \ldots \hat A_p)$

$\{\hat u_t\}_{t=1}^{T}$ - resample using a random number generator. $N = 1000$ u_t^i

By same equation get $\tilde y_t^i \sim$ new sample $(u_t^i) \to$ reestimate $\hat\beta^i \Rightarrow$ have distribution of $\hat\beta$!

2) $\Rightarrow$ Bias-correct + s.e. $\quad \hat{Bias} = mean(\hat\beta^i) - \hat\beta \quad \Rightarrow \tilde\beta = \hat\beta - \hat{Bias}$

3) $\Rightarrow$ use Bias-corrected $\tilde\beta$ to Bootstrap = 2000 ?

$i = 1..2000$ $\{y_i^*\}_{t=1}^{T} = \tilde c + \sum \tilde A y_{ti}^* + \hat u_t^i$ (Bias corrected; old residual.) $\xrightarrow{\text{estimate}}$ $\tilde y \to$ ~~Bootstrap~~ $\Rightarrow \hat\beta_i^*$

4) Bias-correction II $\quad \tilde\beta^* = \hat\beta^* - \underset{\text{from (2)}}{bias}$ (By new bootstrap. ?)

5) from β each y^* construct MA(∞) $\tilde B_n^* = \sum^{N} \tilde\beta^* \tilde A_j$, $\tilde B_0 = I$

Estimate $\hat\Omega^* = \frac{1}{T-p} \sum_{p=1}^{T} \hat u_t^* \tilde u_t^{*\prime}$ $\quad$ Cholesky $\hat\Omega^* = PP'$

$= ADA' \Rightarrow$ const orthogonalized IRF $\hat C_n^* = \hat B_n^* A$

STRUCTURAL VAR (Adding econ. interpretation $\Rightarrow$ restricting shocks)

$A_0 y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \ (**)$ (Reduced form: RF)

A_0 - non-diagonal $\Rightarrow$ dyn. simultan. eq's model. $A_0^{-1} \cdot | (\quad)$

Need identification restrictions $\phi_i = A_0^{-1} A_i \quad i = \overline{1,p}$

Reduced-form errors: $\varepsilon_t = A_0^{-1} u_t$

1) can be viewed as forecast errors

2) are linear combinations of true structural shocks

The IRF for Reduced form has no economic interpretation.

Struct. parameters: $\overbrace{A_0}^{m^2}, \dots A_p, \Omega_u \longrightarrow (p+1) m^2 + \frac{m(m+1)}{2}$

RF has: $p m^2 + \frac{m(m+1)}{2} \Rightarrow$ need m^2 restrictions

2 standard approaches

1) Set diagonal of A_0 to $1 \Rightarrow m$ restrictions

use econ intuition (theory) to impose $m(m-1)$ restr. (identification)

a) Use zeros sm where? Not plausible economically (80).

b) assume uncorellated shocks: Ω_u -diagonal.

c) restr. on contemporaneous coefficients A_0.

d) restr. on long-run effects $A(1)^{-1} = (A_0 + A_1 + \dots + A_p)^{-1}$

Popular id. restrictions:

1) Recursive identification (Sims-Bernanke)

2) Long-run relationships (Blanchard - Quah)

1: Recursive: a) Ω_u - diagonal $\Rightarrow$ structural shocks uncorrelated

b) A_0 - lower triangular. $\Rightarrow A_0$ captures all contemporaneous correlation

$\frac{m(m-1)}{2}$ $\Rightarrow \frac{m(m-1)}{2}$ restrictions.

that's a "default" (Stata) ID, done by software.

That's the "orthogonalized" IRF: $\frac{\partial y_{t+h}}{\partial \varepsilon_{jt}}$ - effect of innovation in j-th variable $\rightarrow$ reduced form innovation

2: need $\frac{\partial y_{t+h}}{\partial u_{jt}} \rightarrow$ use Cholesky decomposition. $E(\varepsilon\varepsilon') = ADA' \Rightarrow u_t = A^{-1}\varepsilon_t$.

A - lower triangular, 1 along diagonal. A^{-1} = A_0 in SVAR.

assume X_1 not affected by $X_2 \dots X_m$ etc.

$A_0 = A^{-1}$ - by Chol. decomp of Ω_ε $\quad \Omega_u = D$ - diag elements of P.

$A_i = A_0 \phi_i \quad i = 1..p. \quad$ IRF - impulse resp. to econ. shocks.

② Use of LR relationships: • $MA(\infty)$ representation.

SVAR: $A(L)y_t = u_t \Rightarrow SMA(\infty): y_t = \overset{C(L)}{\overset{\shortparallel}{A^{-1}}}(L)u_t$.

• can only use with a growth rate as a component of y_t.

$y_{it} = \log X_{it} - \log X_{it-1} \Rightarrow \log x_{it+h} = y_{it+h} + y_{it+h-1} + \dots + y_{it} + \dots$

$\Rightarrow \dfrac{\partial \log X_{it+h}}{\partial u_{jt}} = \sum_{k=0}^{h} \dfrac{\partial y_{it+h-k}}{\partial u_{jk}} = \sum_{k=0}^{h} \underset{\substack{\shortparallel \\ c_{ij,k}}}{A^{-1}_{ij,k}}$ — Long-run neutrality (ex. money)

Thus $\lim_{h\to\infty} \dfrac{\partial X_{it+h}}{\partial u_{jt}} = \sum_{k=0}^{\infty} c_{ij,k}$ is the element of $C(1) = C_0 + C_1 + \dots$

Long-run neutrality: $\lim_{h\to\infty} \dfrac{X_{i,t+h}}{\partial u_{jt}} = 0$.   e.g. $2=m$ case: shock on X_{2t} does not have effect on level of x_{1t}. $\Leftrightarrow C(1)$ lower triangular.

If $m>2$ long-run neutrality assumes that u_{it} has no long-run effect on x_{jt} for $j<i$. $\frac{m(m-1)}{2}$

⊕ also assume Ω_u is diagonal. $\frac{m(m-1)}{2}$

Estimation of SVAR:

① Indirect LS: Start from RForm: OLS $\hat{\Phi}$ and $\hat{\Omega}_u$

⇒ Solve for $\hat{A}$ analytically.

$A_0 \Omega_u A_0' = \Omega_\varepsilon \qquad A_0^{-1} A_i = \hat{\phi}_i \quad i=1\dots p.$ + identifying assumptions
analytically. if exactly identified

② Full-info maximum likelihood (FIML)

$\mathcal{L}(\phi_1 \dots \phi_p, \Omega_\varepsilon) = -m\frac{(T-p)}{2}\log 2\pi - \frac{(T-p)}{2}\log|\Omega_\varepsilon| - \frac{1}{2}\sum_{t=p+1}^{T} \varepsilon_t' \Omega_\varepsilon^{-1}\varepsilon_t$

Substitute $\varepsilon_t = A_0^{-1}u_t$ and $\Omega_\varepsilon = A_0^{-1}\Omega_u (A_0^{-1})' \Rightarrow$ SVAR likelihood:

$\mathcal{L}(A_0, A_1 \dots A_p, \Omega_u) = -\frac{m(T-p)}{2}\log 2\pi - \frac{(T-p)}{2}\log|A_0^{-1}\Omega_u(A_0^{-1})'| - \frac{1}{2}\sum_{t=p+1}^{T} \varepsilon'[A_0^{-1}\Omega_u A_0^{-1\prime}]^{-1}\varepsilon$

if no restrictions on $A_1 \dots A_p \Rightarrow$ 2 steps

1) regress y_t on $y_{t-1} \dots y_{t-p} \Rightarrow \hat{\varepsilon}_t$

2) substitute $\hat{\varepsilon}$ and impose identifying restrictions on $A_0, \Omega_u \Rightarrow$ ML

③ Have extra (over) identifying restrictions ⇒ use GMM.

Say $r \neq m(m-1)$ ⇒ use them to test the model.

1) Obtain Ω_ε from RF VAR(p)

2) SVAR – ML + impose all $r + m(m-1)$ identification restrictions

3) Derive $\Omega_\varepsilon^R = A_0^{-1}\Omega_u(A_0^{-1})'$ from 2.

4) Compare Ω_ε vs Ω_ε^R (=) ⇒ restr. aren't binding.

test stat $= |\Omega_\varepsilon^R| - |\Omega_\varepsilon| \sim \chi^2(r)$ (≠) ⇒ reject. ⇒ binding.

EXAMPLE: 1) AS + tech shock $\quad A_0 y_t = A_1 y_{t-1} + u_t$. $\quad$ need $2(2-1)=2$ restr.

2) MS + ms shock $\quad$ 1) Ω_u diagonal $\to 1$

$\Rightarrow y_t = \phi y_{t-1} + \varepsilon_t \Rightarrow \hat{\phi}, \hat{\Omega}_\varepsilon$ by OLS. $\quad$ 2) only $M_t \to y_{t+1}$.

ex: ε_{1t} 1 -0.5 ... $\quad \hat{\Omega}_\varepsilon = \begin{pmatrix} 1/2 & 2/5 \\ 2/5 & 1/2 \end{pmatrix} \overset{\text{chol}}{\Rightarrow} P = \begin{pmatrix} 0.707 & 0 \\ 0.566 & 0.424 \end{pmatrix} \Rightarrow \Omega_u = \begin{pmatrix} 0.707^2 & 0 \\ 0 & 0.424^2 \end{pmatrix} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.18 \end{pmatrix}$

1) ε_{2t} 0.5 -1 ...

$\overset{\shortparallel}{A D^{1/2}} \Rightarrow A = \begin{pmatrix} 1 & 0 \\ 0.8 & 1 \end{pmatrix} \Rightarrow A^{-1} = A_0 = \begin{pmatrix} 1 & 0 \\ -0.8 & 1 \end{pmatrix}$

$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -0.8 & 1 \end{pmatrix} \begin{pmatrix} 1 & -0.5 & \dots \\ 0.5 & -1 & \dots \end{pmatrix}$ money / output. $\quad$ if use diff ordering $\Rightarrow$ diff shocks. $\quad A_1 = A_0^{-1}\phi = A\phi$.

2) non-recursive identification. restr. on A_0 + diag Ω_u. What matters is correlation.

$\begin{pmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_2}^2 \end{pmatrix} = \begin{pmatrix} 1 & a_1 \\ a_2 & 1 \end{pmatrix} \begin{pmatrix} .5 & .4 \\ .4 & .5 \end{pmatrix} \begin{pmatrix} 1 & a_1 \\ a_2 & 1 \end{pmatrix}$ 3 eq, 4 unknowns. $\Rightarrow$ any additional restr. gives id.

$\frac{\partial y_{t+h}}{\partial u_{jt}} = B_h a_j \quad B_h$ - coef of SMA(∞) for RF $\quad a_j$ - j-th col. of A_0^{-1}.

HW! **Blanchard - Quah (89)**

decompose output in temp + perm. component $\quad x_{1t}$ - output $\quad x_{2t}$ - unempl.

Let $y_t = \begin{pmatrix} \Delta x_{1t} \\ x_{2t} \end{pmatrix}$ ← supply, ↑ Both, ↙ stationary, demand → $\quad$ SMA(∞): $y_t = C_0 u_t + C_1 u_{t-1} + \dots$

$\Rightarrow \Delta x_{1t}$ - unit root. $\quad$ Ident. assumptions: 1) $\Omega_u = I$.

2) demand shock has no LR eff on output.

1) Recover SMA(∞) by estim RF for VAR(p) $\quad \Rightarrow C_0 + C_1 + \dots$ lower triangular.

$y_t = \varepsilon_t + B\varepsilon_{t-1} + \dots$, $\quad \varepsilon_t = C_0 u_t \Rightarrow \Omega_\varepsilon = C_0 \overset{I}{\Omega_u} C_0' = C_0 C_0'$

$\sum_0^\infty C_i = \left(\sum B_i\right) C_0$ is lower triangular $\Rightarrow \underbrace{(\sum B_i)}_{P} \Omega_\varepsilon \underbrace{(\sum B_i)'}_{P'}$ - known

$B_0 = I$

Practice $\Rightarrow$ Estimate Var(p), RF $\Rightarrow \Omega_\varepsilon$, Invert SMA($\infty$) $\Rightarrow B_i$, Choleski $PP' = (\sum B_i)\Omega_\varepsilon(\sum B_i)'$

$\Rightarrow C_0 = (\sum B_i)^{-1} P \Rightarrow$ SMA(∞): $C_i = B_i C_0 \Rightarrow$ IRF. $C_h = \frac{\partial y_{t+h}}{\partial u_t}$

ex: Chari Kehoe McGrat

① if VAR$_{OLS}$ bad $\Rightarrow (\sum B_i)$ magnifies the error. $\Rightarrow$ CIs large ← very important to choose p correctly.

↑ (p small) ② Sensitive to assumptions about ordering.

and unit roots (Δ stat $\Rightarrow$ MA(∞) ← non-invertible) non identifyable.

Bayesian VAR

- Why? Solves the curse of dimensionality.
  Use bayesian methods to incorporate prior info about params
- Litterman (86) prior; shrink VAR towards a random walk
  VAR(p): $y_t = A(L)\, y_{t-1} + \varepsilon_t$ RW: $\Delta y_t = \varepsilon_t$
- In differences: $y_{it} = \alpha + \gamma_1 y_{it-1} + \ldots + \gamma_p y_{it-p} + X_{t-1}'\beta + \varepsilon_t$
  where $\underset{1\times p(m-1)}{X_{t-1}'} \to$ all other variables' lags

Litterman (Minnesota) prior
1) do not shrink α (="flat prior")
2) shrink γ's towards 0, shrink long lags more than short ones
3) shrink β's more than γ's

for $y_t = \underset{1\times k}{X_t'}\underset{k\times 1}{\beta} + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma^2) \to$ CLR Normal: critical.

Bayesian estimate $y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}$ $X' = \begin{bmatrix} X_1' \\ \vdots \\ X_T' \end{bmatrix}$ if $\beta \sim N(b, V) \to$ prior distr

restricted to prior; how confident you are in your guesses; larger diag. elem $\Rightarrow$ less confident

posterior distr of (β| data)

$\beta \sim N(\beta^*, \sigma^2(X'X + \sigma^2 V^{-1})^{-1})$

$\beta^* = (X'X + \sigma^2 V^{-1})^{-1}(X'y + \sigma^2 V^{-1} b)$

$\oplus$ good for forecasting and values, stable
$\ominus$ no inference

Ex: for Minnesota prior: let $\theta = (\alpha, \beta', \gamma')$, V – diagonal

$Var(\alpha) = 10^8$ $Var(\gamma_j) = \left(\frac{\lambda}{j}\right)^2$, $j = 1, .., p$ $Var(\beta_i) = \left(\omega\lambda \frac{\sigma_y}{\sigma_{x_i}}\right)^2$ $i = 1 \ldots K$

λ, ω – user-defined hyperparameters Typical choices: $\omega = 0.2$ or 0.5
$\lambda = 0.2$ (try diff. for robustness)

for VAR's in differences the prior $\theta = 0$.

$\theta^B = (X'X + \hat{\sigma}^2 V^{-1})^{-1}(X'Y)$ $\quad \underset{(T-p)\times(1+p+(m-1)p)}{X}$ $\quad \underset{(T-p)\times 1}{Y}$

Equivalent to Ridge regression

$\hat{\sigma}$ = s.d. of resid's of y_t on p lags

Kalman Filter: State-space Estimation

s-s. representation $\underset{m\times 1}{y_t}$ = observed variables $\underset{r\times 1}{S_t}$ = unobserved state variables

Observational eq's: $y_t = A'X_t + H'S_t + W_t$

State eq's: $S_t = F S_{t-1} + V_t$

X_t – predetermined variables (e.g. lags of y_t)

$W_t \sim iid(0, \Omega_w)$ $V_t \sim iid(0, \Omega_v)$

intuition: X_t – deterministic S_t – summarizes all relevant past info for predicting y_t.

1) e.g. $\boxed{MA(1)}$ $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\varepsilon_t \sim iid(0, \sigma^2)$

$\Omega_\nu = \begin{pmatrix} 6^2 & 0 \\ 0 & 0 \end{pmatrix}$

$\boxed{MA:}$

State eq: $\begin{pmatrix} \varepsilon_{t+1} \\ \varepsilon_t \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}}_{F} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t+1} \\ 0 \end{pmatrix}$ Obs. eq: $y_t = \underbrace{(1 \;\; \theta)}_{H} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix}$

$X = w = \Omega_w = 0$

Repr. not unique, but all give same ML.

2) $\boxed{\text{Unobserved common factor}}$: Stock & Watson (91) y_t - macro var's, S_t - scalar state of BC.
X_t - idiosyncratic component y_{it} not affected by y_{jt}.

State eq:

$$\begin{pmatrix} S_t \\ X_{1t} \\ \vdots \\ X_{mt} \end{pmatrix} = \begin{pmatrix} \phi_S & & & 0 \\ & \phi_1 & & \\ & & \ddots & \\ 0 & & & \phi_m \end{pmatrix} \begin{pmatrix} S_{t-1} \\ X_{1t-1} \\ \vdots \\ X_{mt-1} \end{pmatrix} + \begin{pmatrix} \nu_{St} \\ \nu_{1t} \\ \vdots \\ \nu_{mt} \end{pmatrix}$$

Obs eq:

$$\begin{pmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} + \begin{pmatrix} \delta_1 & 1 & & 0 \\ \vdots & & \ddots & \\ \delta_m & 0 & & 1 \end{pmatrix} \begin{pmatrix} S_t \\ X_{1t} \\ \vdots \\ X_{mt} \end{pmatrix}$$

δ_i - sensitivity of i-th variable to RBC

Algorithm of estimation: (condit. expectations given data)

observe $(y_1 \ldots y_T)$ and $(X_1 \ldots X_T)$, know $F, \Omega_\nu, A, H, \Omega_w$.

Calculate recursively $\hat{S}_{t+1|t} = E(S_{t+1} | I_t)$. It's MSE: $P_{t+1|t} = E[(S_{t+1} - \hat{S}_{t+1|t})(\;\;)']$

Step 1: Find initial values $\hat{S}_{1|0} = E(S_1)$ $\quad P_{1|0} = Var(S_1)$ (Assumes cov-stationarity of S_t)

$E(S_t) = F E(S_{t-1}) \Rightarrow E(S_t) = 0 = \hat{S}_{1|0}$

$E(S_t S_t') = F E(S_{t-1} S_{t-1}') F' + \Omega_\nu \Rightarrow \Omega_S = F \Omega_S F' + \Omega_\nu \Rightarrow$ take $vec(LHS) = vec(RHS)$

$vec(\Omega_S) = (F \otimes F) vec(\Omega_S) + vec(\Omega_\nu) \quad \Rightarrow vec\, \Omega_S = (I_{r^2} - F \otimes F)^{-1} vec(\Omega_\nu)$

$\Rightarrow \Omega_S$ $r \times r$ from the elements $= P_{1|0}$

Step 2,..T: use 2 eq's:

$V_t = E(Y_t | X_t, \Omega_{t-1}) = H' P_{t|t-1} H + \Omega_w \quad \leftarrow$ denote

1) Contemporaneous est of S_t: $\hat{S}_{t|t} = \hat{S}_{t|t-1} + P_{t|t-1} H V_t^{-1} (y_t - A'x_t - H'\hat{S}_{t|t-1})$

2) Forecasts: $\hat{S}_{t+1|t} = F \hat{S}_{t|t}$

3) Contemporaneous MSE: $P_{t|t} = P_{t|t-1} - P_{t|t-1} \cdot H \cdot V_t^{-1} H' P_{t|t-1}$

4) Forecast MSE: $P_{t+1|t} = F P_{t|t} F' + \Omega_\nu$

ML-estimation: given $\hat{S}_{t|t-1}$ and $P_{t|t-1} \Rightarrow$ calculate: (assuming normality of everything)

$E_t = E(y_t | x_t, \Omega_{t-1}) = A'x_t + H'\hat{S}_{t|t-1}$ $\quad \mathcal{L}(F, \Omega_\nu, A, H, \Omega_w) = \sum_{t=1}^{T} \log f_{y_t | x_t, \Omega_{t-1}}$

$Var(y_t | x_t, \Omega_{t-1}) = V_t \Rightarrow f_{y_t|x_t,\Omega_{t-1}} = (2\pi)^{-\frac{m}{2}} |V_t|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y_t - E_t)' V_t^{-1} (y_t - E_t)\}$.

Initial Values $\to$ guess $\to$ max numerically.

Smoothed estimates of the state variables

need $\{\hat{S}_{t|T}\} \to$ estimates based of whole sample info. $\to$ for interpretation and inference about hist. values

Start from $\hat{S}_{T|T}$ and $P_{T|T}$ $\to$ Backward recursion of 4 eq's

$$\hat{S}_{t|T} = \hat{S}_{t|t} + J_t (\hat{S}_{t+1|T} - \hat{S}_{t+1|t}) \qquad J_t = P_{t|t} F' P_{t+1|t}^{-1}$$

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

Generalization: S-S eq's with random coef's

$$\xi_t = F(X_{t-1})\xi_t + \nu_t \qquad \begin{pmatrix}\nu_t\\ w_t\end{pmatrix}\Big| X_t, \Omega_{t-1} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\Omega_\nu(x) & 0\\ 0 & \Omega_w(x)\end{pmatrix}\right)$$
$$y_t = a(X_t) + H(X_t)'\xi_t + w_t$$

X-deterministic ⇒ coef. deterministic.

Application: Regression with time-varying coefs

$$\underset{k\times 1}{y_t = X_t'\beta_t + w_t} \quad \oplus \quad (\beta_t - \bar\beta) = F(\beta_{t-1} - \bar\beta) + \nu_t \qquad \bar\beta \text{ - s-s. value.}$$
$$\Updownarrow \qquad\qquad \beta_t \sim AR(1)$$
$$y_t = X_t'\bar\beta + X_t'\xi_t + w_t \qquad \xi_t = F\xi_{t-1} + \nu_t \qquad a(x) = X'\beta \quad \Omega_w(X_t) = \sigma_w^2$$
$$H(x) = X_t$$

Implementation: initial values $\theta = (\bar\beta, \sigma_w^2, F, \Omega_\nu)$

$$\xi_{1|0} = 0 \qquad P_{1|0} = (I_{K^2} - F\otimes F)^{-1}\operatorname{vec}(\Omega_\nu)$$
$$\hat\xi_{t+1|t} = F\hat\xi_{t|t-1} + F P_{t|t-1} X_t V_t^{-1}(y_t - X_t'\bar\beta - X_t'\hat\xi_{t|t-1})$$
$$V_t = X_t' P_{t|t-1} X_t + \sigma_w^2 \qquad P_{t+1|t} = F P_{t|t-1} F' - F P_{t|t-1} X_t V_t^{-1} X_t P_{t|t-1} F' + \Omega_\nu$$
$$\mathcal{L}(\theta) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log V_t - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - X_t'\bar\beta - X_t'\xi_{t|t-1})^2}{V_t}$$
$$\beta_t = \hat\xi_{t|T} + \bar\beta \leftarrow \text{smoothed estimates of state variable}$$

Application 2: likelihood of a linearized DSGE model around s-s.

Sargent (89) DSGE → state-space

ex: stochastic neoclass growth model with leisure:

$$\text{RA}: E\sum_{t=1}^{\infty}\beta^{t-1}\frac{[c_t^{\theta}(1-\ell_t)^{1-\theta}]^{1-\tau}}{1-\tau} \qquad \text{PF}: y_t = e^{z_t}k_t^{\alpha}\ell_t^{1-\alpha} \qquad \varepsilon_t \sim N(0,\sigma_\varepsilon^2)\quad |\rho|<1$$
$$k_t = i_{t-1} + (1-\delta)k_{t-1} \qquad z_t = \rho z_{t-1} + \varepsilon_t \;\; (0) \;\; \text{crucial}$$
$$c_t + i_t = y_t$$

Steady-state, linearize, solve for "policy" functions

$c, k, \ell, y \,|\, z=0$ $\qquad \xi_t \leftrightarrows \begin{pmatrix}k_t - k_{ss}\\ z_t\end{pmatrix}$

Policy: $k_t = k_{ss} + a_{11}(k_{t-1} - k_{ss}) + a_{12}(\rho z_{t-1} + \varepsilon_t)$ (1)

$\ell_t = \ell_{ss} + a_{21}(k_t - k_{ss}) + a_{22}z_t$

c 3 3

y 4 4

State eq: (0) + (1)

$$\text{observe}: \tilde y_t = \begin{pmatrix}GDP_t\\ Hous_t\\ INv_t\end{pmatrix} = \begin{pmatrix}y_t + \nu_{1,t}\\ \ell_t + \nu_{2,t}\\ y_t - c_t + \nu_{3,t}\end{pmatrix} \Rightarrow F = \begin{pmatrix}a_{11} & a_{12}\rho\\ 0 & \rho\end{pmatrix} \quad w_t = \begin{pmatrix}a_{12}\varepsilon_t\\ \varepsilon_t\end{pmatrix} \quad \Omega_w = \begin{pmatrix}a_{12}^2\sigma_\varepsilon^2 & a_{12}\sigma_\varepsilon^2\\ a_{12}^2\sigma_\varepsilon^2 & \sigma_\varepsilon^2\end{pmatrix}$$

$X_t = 1$

$$A = \begin{pmatrix}y_{ss}\\ \ell_{ss}\\ y_{ss} - c_{ss}\end{pmatrix} \qquad H' = \begin{bmatrix}a_{41} & a_{42}\\ a_{21} & a_{22}\\ a_{41} - a_{31} & a_{42} - a_{32}\end{bmatrix}$$

Multiple I(1) series

$y_t = \beta x_t + u_t$ – OLS based on assumption u_t – $I(0)$

4 cases 1) $u_t, x_t, y_t \sim I(0)$

2) $y_t \sim I(0)$, $x_t \sim I(1)$, $\beta = 0$ $\to$ OLS correctly indicates $\beta = 0$.

3) $y_t \sim I(1)$, $x_t \sim I(0)$, $\forall \beta$ $u \sim I(1) \to$ OLS not valid

4) $y_t \sim I(1)$, $x_t \sim I(1)$

a) $\exists \beta^*$: $u_t = y_t - \beta x_t \sim I(0) \Rightarrow$ cointegrated

b) $\forall \beta$: $u_t = y_t - \beta x_t \sim I(1) \Rightarrow$ spurious regression.

Newbold, Granger : spurious regression.

Monte Carlo, punchcards. $y_t = y_{t-1} + \varepsilon_{1t}$, $x_t = x_{t-1} + \varepsilon_{2t}$, OLS significant, R^2 high.

why Phillips (86): $T^{-1/2}$. t-stat – func. of Brownian motion. Need to divide by T additionally.

grows with $T^{1/2}$ $\Rightarrow$ likely to reject $H_0: \beta = 0$.

Detecting s.r.: 1) Test for stationarity of x and y.

2) difference data. $\Delta y_t = \beta \Delta x_t + u_t$ $\Rightarrow$ correctly find $\beta = 0$.

VAR with I(1): Sims, Stock, Watson (90) $\Rightarrow$ can use standard t, F-stat for β in front of I(0) var's.

ex: $\begin{vmatrix} y_t \\ x_t \end{vmatrix} = A \begin{vmatrix} y_{t-1} \\ x_{t-1} \end{vmatrix} + B \begin{vmatrix} y_{t-2} \\ x_{t-2} \end{vmatrix} + \begin{vmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{vmatrix}$ 1) $y \sim I(0)$, $x \sim I(1)$

lag-length-test: y_t: $H_0: b_{11} = b_{12} = 0$: $y_t = a_{11} y_{t-1} + b_{11} y_{t-1} - b_{11} \underbrace{\Delta y_{t-1}}_{I(0)} + a_{12} x_{t-1} + b_{12} x_{t-2} + \varepsilon_{1t}$.

F-test valid.

2) if both $x \sim I(1)$, $y \sim I(1)$ $\to$ same thing for Δx_{t-1}, Δy_{t-1} I(0)

Warning: in case of cointegration CANNOT rewrite VAR in differences
misspecified.

Cointegration: $\exists \beta^*$: $\overset{I(1)}{y_t} - \beta^* x_t \sim I(0)$ (or $I(1) \sim y_t$, $\beta y_t \sim I(0)$)

ex: $x_t = x_{t-1} + \varepsilon_{1t}$ $\quad y_t = \gamma x_t + \varepsilon_{2t}$, $\leftarrow$ coint relationship.

Any time you have a long-run eq'm in mind. ex: PIH: $c_t = c_t^{Perm} + c_t^{Trans}$, $c_t^{Perm} = \gamma y_t$ $\quad$ I(0)

if try to rewrite in diff's: $\begin{cases} \Delta x_t = \varepsilon_t & (**) \\ \Delta y_t = \gamma x_t - y_{t-1} + \varepsilon_{2t} = \underbrace{\gamma x_{t-1} - y_{t-1}}_{\text{levels!}} + \varepsilon_{2t} + \gamma \varepsilon_{1t} \end{cases}$ $-\beta \binom{x_{t-1}}{y_{t-1}}$ LR eq'm. $\cancel{\Delta y_{t-1}, \Delta x_{t-1}}$

, not diff-ces.

(**) is called ERROR-CORRECTION MODEL (ECM)

EC term: $z_t = y_t - \gamma x_t = \varepsilon_t \sim I(0)$ – deviation from LR eq'm.

In General: ECM $\begin{cases} \Delta x_t = \beta_1 \Delta x_{t-1} + \gamma_1 \Delta y_{t-1} + \alpha_1 z_{t-1} + \varepsilon_{1t}. \\ \Delta y_t = \underbrace{\beta_2 \Delta x_{t-1} + \gamma_2 \Delta y_{t-1}}_{\text{SR dynamics}} + \underbrace{\alpha_2 z_{t-1}}_{\text{LR dynamics}} + \varepsilon_{2t}. \end{cases}$ ex $0, 0, \begin{bmatrix} 0 & 0 \\ \gamma & -1 \end{bmatrix}$ α_1, α_2

Result 1: A coint. system is a restricted VAR in levels.
Could use VAR in levels, but! inefficient, because does not impose restr.

Result 2: $(x_t\ y_t) \sim CI \Rightarrow$ have ECM form and $|\alpha_1| + |\alpha_2| \neq 0$.

Result 3: — // — $\Rightarrow$ they have a common stochastic trend.

$\begin{cases} x_t = \gamma w_t + \tilde{x}_t \\ y_t = w_t + \tilde{y}_t \end{cases}$ $\Rightarrow$ $\boxed{\text{CI} \Leftrightarrow \text{Common trend} \Leftrightarrow \text{ECM}}$

property $\quad$ intuition $\quad$ DGP.

Tests for CI H_0: no cointegration.

1) β known (from econ. theory).

a) Test comp for $I(1)$

b) Construct $z = \beta' x_t$, test $H_0: z_t \sim I(1)$, reject $\Rightarrow$ CI

2) β unknown: estimate by OLS (super consistent), normalize $\beta = [1, -\gamma_1, -\gamma_m]'$, i.e. $x_1 = \gamma(x_2 \dots x_m]'$

if x_t not CI $\Rightarrow$ residuals of OLS $\sim I(1) \to$ can test. Problem: $\beta_1 = 0$? $\Rightarrow$ miss pec. or $\beta_2 = 1 \Rightarrow \hat\beta$ different.

3) Full info ML for CI. (Johanson 88, 91): start with VECM

$$\Delta X_t = -\underset{m\times h}{\alpha}\underset{h\times m}{\beta'} X_{t-1} - \sum_{j=1}^{p} B_j \Delta X_{t-j} + \varepsilon_t$$

h-cointegrating relationships

ex: $\Delta X_t = -\alpha\beta' X_{t-1} - B_2 \Delta X_{t-1} + \varepsilon_t \sim N(0, \Omega)$

log ML for: $\Delta X_t = A_1 X_{t-1} + A_2 \Delta X_{t-1} + \varepsilon_t$.

$$\mathcal{L}(\Omega, A_1, A_2) = -\frac{Tm}{2}\log(2\pi) - \frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}[\varepsilon_t' \Omega^{-1} \varepsilon_t]$$

$\varepsilon_t = \Delta X_t - A_1 X_{t-1} - A_2 \Delta X_{t-1}$: FIML: max $\mathcal{L}$, s.t. $A_1 = -\alpha\beta'$

Johansen: simple auxiliary regression. FIML $\Leftrightarrow$ algorithm.

STEP1: $\Delta X_t = \hat C_0 + \hat C_1 \Delta X_{t-1} + \hat u_t$ by OLS

$X_{t-1} = \hat d_0 + \hat D_1 \Delta X_{t-1} + \hat v_t$ by OLS

STEP2: „canonical correlations" $\hat\Sigma_{vv} = \frac{1}{T}\sum^{T} \hat v_t \hat v_t'$ $\hat\Sigma_{uu} = \frac{1}{T}\sum \hat u_t \hat u_t'$ $\hat\Sigma_{uv} = \frac{1}{T}\sum \hat u_t \hat v_t'$

Find eigenvalues of $\hat\Sigma_{vv}^{-1} \hat\Sigma_{vu} \hat\Sigma_{uu}^{-1} \hat\Sigma_{uv}$, order as $\hat\lambda_1 > \hat\lambda_2 > \dots > \hat\lambda_m$

STEP3 Johansen (88) FIML, s.t. h restr $\Leftrightarrow \mathcal{L}_h^*$

$$\mathcal{L}_h^* = \frac{-Tm}{2}\log(2\pi) - \frac{Tm}{2} - \frac{T}{2}\log|\hat\Sigma_{uu}| - \frac{T}{2}\sum_{i=1}^{h}\log(1-\hat\lambda_i)$$

Can construct LRatio test for H_0: h c.i. relationships vs ~~H_0: no c.i.r.~~

1) $LR = +2(\mathcal{L}_m^* - \mathcal{L}_h^*) = -T\sum_{i=h+1}^{m}\log(1-\hat\lambda_i)$

2) $H_0: h$ vs $H_1: h+1$ $LR: -T\log(1-\hat\lambda_{h+1})$

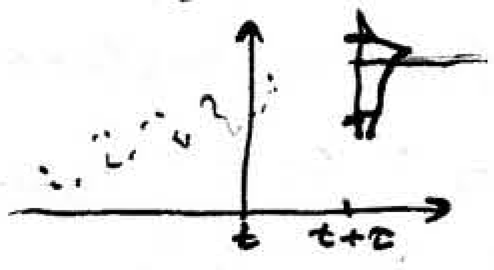
} non-standard critical values

STEP4: knowing h $\Rightarrow$ use h largest eigenvalues $\Rightarrow$

(Hamilton 20.2)   estimate $\hat\beta_1 \dots \hat\beta_h$ - eigenvectors corresp.

Forecasting

$y_{t+\tau} | \Omega_t$, $f_{t+\tau}(y|\Omega_t)$ – objects of interest, τ-horizon.
variable forecast, density forecast, point forecast, interval forecast.
Let x_t – predictors



Loss function: determines a forecast.
which is the „best forecast"

$L(\underset{\text{realization}}{y_{t+\tau}}, \underset{\text{forecast}}{\hat{y}_{t,\tau}})$

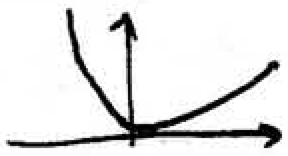
Types of loss:

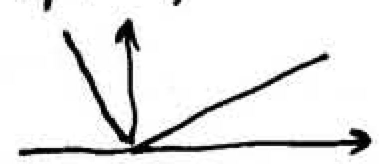
1) $L(\cdot,\cdot)$ = profit, utility.

2) function of forecast error only.

$e_{t,\tau} = y_{t+\tau} - \hat{y}_{t,\tau} \to L(e_{t,\tau}) \to |e|$ a) ; $\to e^2$ b)

c) linex: $L(e) = \exp(\alpha e) - \alpha e - 1$





d) lin-lin: $L(e) = (\alpha - \mathbb{1}(e<0))\,e$
(appr. value-at-risk)

f) direction-of-change.

$L(y_{t+\tau}, \hat{y}_{t,\tau}) = \mathbb{1}\{\text{sign}(y_{t+\tau} - y_t) \neq \text{sign}(\hat{y}_{t,\tau} - y_t)\}$. – appr. for turning points.

Forecast optimality: goal: minimize expected conditional loss

$$\hat{y}^*_{t,\tau} = \min_{f_t} E[L(y_{t+h}, f_t)|\Omega_t].$$

Th: $L = e^2 \Rightarrow \hat{y} = E(y_{t+h}|\Omega_t)$. L = Mean squared forecast error.

Forecasting Time-Series:

1) Suppose cov-stationary $\Rightarrow$ MA(∞) representation. (Wald)

$$y_t = \mu + \theta(L)\varepsilon_t, \quad \varepsilon_t - \text{w.n.}, \quad \theta_0 = 1, \quad \sum_0^\infty |\theta_j| < \infty$$

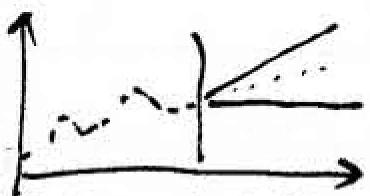
$$y_{t+\tau} = \mu + \underbrace{\varepsilon_{t+\tau} + \theta_1 \varepsilon_{t+\tau-1} + \dots}_{\text{forecast error}} \underbrace{\theta_\tau \varepsilon_t + \theta_{\tau+1}\varepsilon_{t-1} + \dots}_{\text{forecast}}$$

If squared error-loss: $\hat{y}_{t,\tau} = E(y_{t+\tau}|\Omega_t) = \mu + \theta_\tau \varepsilon_t + \dots$

$e_{t,\tau} = y - \hat{y} = \varepsilon_{t+\tau} + \theta_1 \varepsilon_{t+\tau+1} + \dots + \theta_{\tau-1}\varepsilon_{t+1} \to$ serially corr for $\tau > 1$
MA($\tau-1$) process.

MSFE $= E e^2 = (1 + \theta_1^2 + \dots + \theta_{\tau-1}^2)\,\sigma_\varepsilon^2$

$\Rightarrow$ increases with horizon τ.

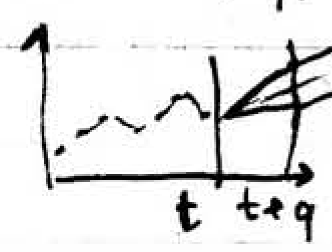


Updating formula (incorp. previous forecasts) : $\hat{y}_{t,\tau} = \hat{y}_{t-1,\tau+1} + \theta_\tau \underbrace{(y_t - \hat{y}_{t-1,1})}_{\text{time-}t\text{ forecast.}}$

error learning. Warning: only valid if data does not change, if there are breaks $\Rightarrow$ better reestimate.

1) if MA(q) $\Rightarrow$ $\hat{y}_{t,\tau} = \begin{cases} \mu + \theta_1\varepsilon_t + \dots + \theta_{\tau-1}\varepsilon_{t+1}, & \tau \le q \\ \mu & \tau > q. \end{cases}$

MSFE $= (1 + \theta_1^2 + \dots + \theta_q^2)\,\sigma_\varepsilon^2$
stops increasing.



2) if AR(1): $y_t = \mu + \phi y_{t-1} + \varepsilon_t \quad |\phi| < 1$

$$y_{t+\tau} = \mu + \phi y_{t+\tau-1} + \varepsilon_{t+\tau} = \underbrace{\mu(1 + \phi + \dots + \phi^{\tau-1}) + \phi^\tau y_t}_{\text{forecast } \hat{y}_{t,\tau}} + \varepsilon_{t+\tau} + \dots + \phi^{\tau-1}\varepsilon_t$$

$\tau \to \infty$, $\hat{y} \to \frac{\mu}{1-\phi}$ uncond mean.

MSFE $= (1 + \dots + \phi^{2(\tau-1)})\,\sigma_\varepsilon^2$, $\uparrow \tau \to \frac{\sigma_\varepsilon^2}{1-\phi^2}$ uncond.

In practice: Available sample $(y_1 \ldots y_T)$ $\underbrace{1 \longrightarrow T}_{est} \xrightarrow{forec} T+\tau$

1) Assume $AR(p)$: $y_t \sim const, y_{t-1} \ldots y_{t-p}$: OLS $\Rightarrow \hat{\mu}, \hat{\phi}_1, \ldots \hat{\phi}_p$

Construct: $\hat{y}_{t,\tau} = \hat{\mu} + \hat{\phi}_1 \hat{y}_{t,\tau-1} + \ldots + \hat{\phi}_p \hat{y}_{t,\tau-p} \Rightarrow$ Iterate

2) Projection method (most widely used)

$y_{t+\tau} = x_t' \beta + u_{t+\tau}$ by OLS: $y_t \sim x_{t-\tau} \Rightarrow \hat{y}_{T,\tau} = x_T' \hat{\beta}_T$

$\tau = 1 \Rightarrow 1 \equiv 2$ $\quad \tau > 1$: $AR(1) \Rightarrow$ 1) $\hat{y} = \hat{\phi}^{\tau} y_t$ $\quad$ Theory: $\hat{\alpha} = \hat{\phi}^{\tau}$

2) $\hat{y} = \hat{\alpha} y_t$ $\quad$ Practice: (1) magnifies errors

Sources Of Forecast Errors (Quadratic Loss)

Forecast model: $y_{t+\tau} = x_t' \beta + u_{t+\tau}$ $\quad \hat{y}_{t,\tau} = x_t' \hat{\beta}$

DGP: $y_{t+\tau} = E(y_{t+\tau} | \Omega_t) + \varepsilon_{t+\tau}$

$$e_{t,\tau} = y_{t+\tau} - \hat{y}_{t,\tau} = \underbrace{E(y_{t+\tau}|\Omega_t) - x_t'\beta}_{\text{specification error}} + \underbrace{x_t'(\beta - \hat{\beta})}_{\text{estimation error}} + \underbrace{\varepsilon_{t+\tau}}_{\text{unpredictable component}}$$

Only first two can be changed, there is a tradeoff.

Model selection criteria:

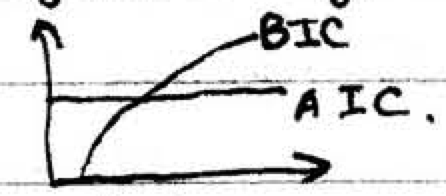
Have a number of competing models $\Rightarrow$ use info criterion.

K - # parameters $\quad$ m - sample size. $\quad$ select best, controlling for estimation error.

$$IC = -\frac{\mathcal{L}(k)}{m} + \frac{K}{m} f(m) \quad \rightarrow \text{average likelihood vs penalty}$$

ex: AIC (Akaike) $f(m) = 2$ $\quad$ BIC (Schwartz's Bayesian) $f(m) = \log m$

$\downarrow$ selects more lags $\quad$ $\downarrow$ less parsimonious



Consistency of IC: P(will select the DGP as $m \to \infty$) = 1

(for the class of models with DGP truly inside)

Nishi(88): Need: 1) $\lim_{m\to\infty} \frac{f(m)}{m} \to 0$ $\quad$ 2) $\lim_{m\to\infty} \frac{f(m)}{\log\log m} \to \infty$ $\quad \Rightarrow$ AIC is not consistent

When forecasts fail + remedies

1) Specification + estimation errors

a) use IC $\quad$ b) shrinkage estimation $\quad$ c) forecast estimation.

2) Overfitting: model explains both noise and signal in the data.

$\Rightarrow$ evaluate rather on out-of-sample, than on in-sample perf.

3) Trends: deterministic or stochastic?

$\rightarrow$ unit root pretests

4) Structural breaks

a) Time-varying parameter models $\quad$ b) Discard old data (rolling window)

c) discounted least squares.

Forecast evaluation.

1) Are they optimal given loss func (absolute)

2) Best out of a set of alternatives (relative)

Economic Questions asked in an evaluation framework:

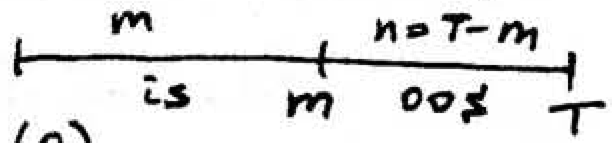
1) Are expectations rational? (mrkts efficient, returns predictable)
stock forecasts optimal

2) Do interest rates forecast inflation?
(compare AR-forecasts vs AR-forecasts + IR) → nested model comparisons

Out-of-sample evaluation:

In-sample: use (1…T) all observations $\not\Rightarrow$ good fit out-of-sample.

OOS: split data into in-s. and oos portions:

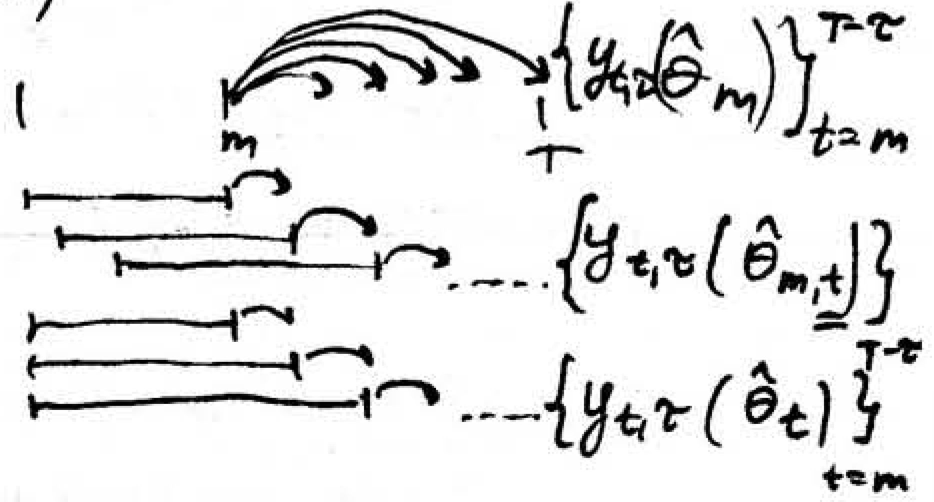


Forecasting schemes: forecasts: $\hat{y}_{t,\tau}(\theta)$

1) fixed. fix m and $f(\hat\theta)$ same for all $\tau \in [1,n]$

2) rolling reestimate [1,m], [2,m+1] … for all τ

3) recursive reestimate [1,m], [1,m+1] …



Compare sequence of forecasts to realizations and obtain forecast errors $\{e_{t,\tau}\}_{t=m}^{T-\tau}$ ⇒ Evaluate using loss function $L(\cdot)$

Notice: must account for ~~evaluation~~ estimation uncertainty.

EVALUATION OF POINT FORECASTS

1) Absolute: optimal w.r.t. $L(\cdot)$ Properties:

$\hat{y}_{t,\tau}$ - optimal, if $\hat{y}_{t,\tau} = \arg\min E[L(y_{t+\tau}, f)|\Omega_t)$ L' - "generalized for. er." for $L(\cdot)=x^2 \to L'=e$

⇒ 1) Unbiased $E[L'(y,\hat{y})]=0$ 2) uncorrelated with Ω_t: $E[L'(y,\hat{y})|\Omega_t]=0$.

ex: $L=e^2$ ⇒ popular „rationality test" Mincer-Zarnowitz (69) regression.

Out of sample. $H_0: \alpha_0 = \alpha_1 = 0$, Wald, $e_{t,\tau} = \alpha_0 + \alpha_1 \hat{y}_{t,\tau} + u_{t+\tau}$
use HAC(τ-1) $\sim \Omega_t$ proxy. MA(τ-1)

Forecast unbiasedness test: $e_{t,\tau} = \alpha + u_{t,\tau}$ $H_0: \alpha=0$. HAC(τ-1), t-test

Problem: ignores estimation errors: Solution: West (96) McCracken (00)
(not used much) propose corrections, based on estim. errors

test of sign predictability (if only interest in sign of change)

$L(y,\hat{y}) = \mathbb{1}\{\text{sign}(y_{t+\tau}) \neq \text{sign}(\hat{y}_{t,\tau})\}$.

Construct dummies $\mathbb{1}(\hat{y} \geq 0) = \alpha_0 + \alpha_1 \cdot \mathbb{1}(y_{t+\tau} \geq 0) + \text{error}$. : $H_0: \alpha_1 = 0$.

Pesaran, Timmermann (94) - extend to case of multiple choice. (not very good)

<u>Comparing accuracy of multiple forecasts</u>

- many suboptimal forecasts → select best w.r.t. $L(\cdot)$ say $\hat{y}^{(1)}, \hat{y}^{(2)}$

Test if they have equal expected loss:

$$H_0: E\left[L(y,\hat{y}^{(1)}) - L(y,\hat{y}^{(2)})\right] = E\Delta L = 0$$

- Estimate $EL^{(i)}$ by average OOS loss $\bar{L}^{(i)} = \frac{1}{m}\sum_{t=m}^{T-\tau} L^{(i)}_{t+\tau}$

ex: quadratic: MSFE, absolute: MAFE → choose min $\bar{L}$

Diebold, Mariano (95): $H_0: E\Delta L_{t+\tau} = 0$, assume satisfies CLT

$\sqrt{n}(\overline{\Delta L} - E\overline{\Delta L}) \to N(0, \sigma^2)$ $\quad t^{DM} = \frac{\sqrt{n}\,\overline{\Delta L}}{\hat{\sigma}} \to N(0,1)$, $\hat{\sigma} \to HAC(\tau-1)$ rule of thumb.

Reject ⇒ choose lower $\bar{L}$

<u>Problem</u>: estimation error West (96)! Good if rolling window White, Giaco (06) and no nested!

<u>recursive</u> ⇒ (West 96) $t^{west} = \frac{\sqrt{n}\,\overline{\Delta L}}{\sqrt{\hat{\sigma}^2 + \text{terms due to est error}}}$ → difficult to compute → should not be nested.

<u>More than 2 models</u>

White (00) Reality check for data snooping: Does any model beat the benchmark?

l models, 1 benchmark, loss $L_{t+\tau}$

Compute: $L^{(0)}, L^{(1)}, \ldots L^{(l)} \Rightarrow \Delta L^{(1)} \ldots \Delta L^{(l)}$ w.r.t. (0) $\quad H_0$: no model beats the benchmark

$H_a: \max_{k=1,\ldots,l} E[\Delta L^{(k)}_{t+\tau}] = 0$

Test stat: $\max_{k=1\ldots l} \sqrt{n}\,\overline{\Delta L}^{(k)}$ ($\overline{\Delta L}$ = OOS avg loss). Derive p-values by bootstrap.

$\begin{bmatrix} \overline{\Delta L}^{(1)} \\ \overline{\Delta L}^{(l)} \end{bmatrix}$ - resample by "block" bootstrap for possible "stationary" serial correl in $\Delta L_{t+\tau}$

Construct pseudo-samples

indep. $\{\Delta L^{(1)}_{t+\tau}\} \ldots \{\Delta L^{(l)}_{t+\tau}\}$

$\downarrow$ $m \times 1$ $\quad \overline{\Delta L}^{(1)}_{boot}$ $\ldots$ $\overline{\Delta L}^{m \times (l)}_{boot}$.

$\downarrow$ t-statistic$^{(1)}_j$ $\ldots$ t-stat$^{(l)}_j$] take test stat $= \max \sqrt{n}\left(\overline{\Delta L}^{(k)}_{boot} - \overline{\Delta L}^{(k)}_{original}\right)$ repeat $j = 1 \ldots J$ ⇒

⇒ compute $(1-\alpha)$ quantile of bootstrap distr.

if original t-stat > bootstrap quantile ⇒ reject H_0.

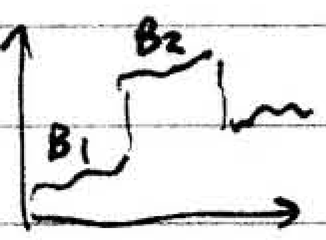
<u>Note on bootstrap for time-series</u>: iid bootstrap destroys serial corr's in data.

Bootstraps from original ⇐

- block bootstrap (Künsch, 89) → resample blocks of data with ~~replacing~~ replacement, link together.

$\underbrace{X_1\ X_2 \ldots X_B}_{B_1}\ X_{B+1} \ldots X_T$

B_2, $\ldots$, B_q $\quad q = T - b + 1$



Problem: may look like

- Sample randomly from $B_1 \ldots B_q$
- concat $B_{j_1} \ldots B_j \ldots B_{j_T}$ (same length T)

Sol: Politis, Romano (95) Stationary bootstrap (random length) ⇒ preserves properties

hypergeometric

15th - deadline 16th