

Variáveis Explanatórias e Intervenções em ME

$$X_t = (x_{1t} \ x_{2t} \ \dots \ x_{kt}) \sim k \times 1$$

$$\gamma = (\gamma_1 \ \gamma_2 \ \dots \ \gamma_k) \sim k \times 1$$

$$\text{Ex: } \begin{cases} y_t = \mu_0 + (\gamma' x_t) + \epsilon_t & \epsilon_t \sim N(0, \sigma^2) \\ \mu_{t+1} = \mu_t + \beta_t + \eta_t \\ \beta_{t+1} = \beta_t + \zeta_t \end{cases}$$

duas possibilidades:

i) Escreve o modelo na forma

↳ por MV ↳ parâmetros fixos associados
 secção 6: $y_t = z_t^* x_t + c_t + \epsilon_t$ onde $c_t = \gamma' x_t$

$$v_t = y_t - \hat{y}_{t|t-1} = y_t - (z_t^* \alpha_t + c_t)$$

↳ previsões | passo a frente

$$v_t = v_t(\gamma, r) \Rightarrow \ell(\gamma, r) \propto -\frac{1}{2} \sum (v_t | F_t)^2$$

↳ contém r

maximiza em relações a γ .

obs: é possível
concentrar a veros
em relações a γ e
resolver por
método generalizado.

ii) Incorpora γ no vetor de estado

↳ dentro do vetor de estado

$$y_t = \mu_t + \gamma' x_t + \epsilon_t$$

Neste caso,

estaremos implicitamente
considerando γ invariante
no tempo: $\gamma_{t+1} = \gamma_t$

$$y_t = \begin{pmatrix} 1 & 0 & x_t \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_t \end{pmatrix} + \epsilon_t$$

$k \times 1$: incluimos k componentes

a mais.

Em $t=T \Rightarrow \gamma_T$: é esta que ililibra
nossos pesos.

$$\begin{pmatrix} \mu_{t+1} \\ \beta_{t+1} \\ \delta_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \\ \delta_t \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix}$$

$\tilde{\gamma}_{t+1} = \tilde{\gamma}_t \Rightarrow$ não há variações em $\tilde{\gamma}_t$ (poderia haver!)

Ex:

precisa dizer

$$\tilde{\gamma}_{t+1} = \phi \tilde{\gamma}_t + w_t$$

$$\begin{pmatrix} \delta_{1,t+1} \\ \delta_{2,t+1} \end{pmatrix} = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix} \begin{pmatrix} \delta_{1,t} \\ \delta_{2,t} \end{pmatrix} + \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix} \leftarrow \begin{array}{l} \text{como varia no} \\ \text{tempo (ex.: de evolução)} \\ \text{do coeficiente, estrutura)} \end{array} \rightarrow \text{VAR (1)}, w_t \sim N(0, Q_w)$$

cheia
(covariâncias
no círculo)

Pode-se mostrar que as estimativas de $\tilde{\gamma}$ obtidas via Fk^{atualizada} são equivalentes às de MO recursivas



$$\tilde{\gamma}_t = \tilde{\gamma}_{t-1} + ()$$

Ex: Se quisermos extrair esse princípio as componentes ^{NO} inerentes a uma série p/ buscar explicações econômicas / físicas, mas devemos colocar variáveis explicativas



concorrência com as componentes, que nos reas + as mesmas

(ex: análise de venda de cerveja,

incluindo temperatura como var. explicativa)

↳ comp. sazonal n'teria n'uma explicação que dize

→ Pode-se 1º fazer o procedimento q var. explicativa

Se por exemplo vêmos ainda n'el RB, tem correlações

⇒ pode incluir variável explicativa

Ainda assim, o significado da sazonalidade nas

será n'as o mesmo q o verificado ante. da variável explicativa.

Podemos querer separar por exemplo a influência de fatores
na sazonalidade:

$$(i) y_t = \mu_t + \gamma_t + \varepsilon_t$$

γ_t : raz. total \Rightarrow temp + feriados etc.

$$(ii) y_t = \mu_t + \gamma_t + \delta x_t + \varepsilon_t$$

γ_t : temperatura

δx_t (expansão temp.) \Rightarrow feriado

Obs 2: Essa variável explicativa pode ser uma dummy
ou ainda, intervenções

- Por exemplo, faz modelagem e análise de resíduos
ainda apresenta problema, como por exemplo, por
conta de outliers

\Rightarrow Modela intervenções a posteriori

O que é atípico não pode ser decidido a priori. É
preciso modelar e verificar.

Se o modelo dinâmico, só se pode identificar o
que é atípico a posteriori. Não algo que a princípio
antes de rodar o modelo pareça estranho, pode ser
absorvido pelo modelo.

Por isso, esta análise do que é atípico só pode ser
feita a posteriori, olhando o comportamento dos resíduos.

Exs. de intervenções:

$$i) \text{ Pulso } x_t = \begin{cases} 1 & t = t^* \\ 0 & t \neq t^* \end{cases}$$

ii) Escada
(step function)

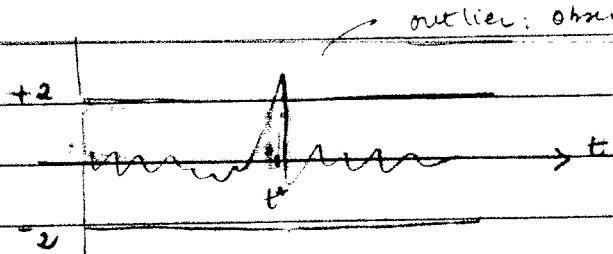
t

$$x_t = \begin{cases} 0 & t < t^* \\ 1 & t \geq t^* \end{cases}$$

ilibra

Qual tipo de funções usar para variações dependerá da análise de resíduos na novacás.

Ex1: $\hat{V}_t^* = v_t \sim N(0,1)$
 $f_x^{''2}$



outlier: observação atípica

), neste caso, coloca variável explicativa
como função pulso.

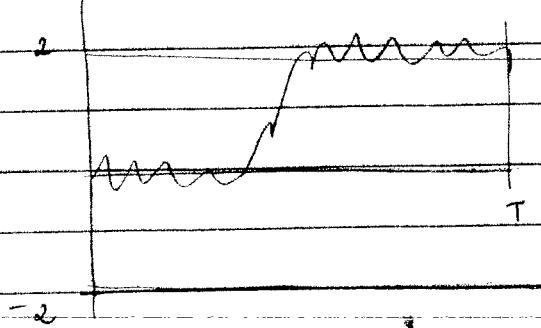
Como em t^* ocorreu algo atípico, muito para
do padrão, modela como dummy p/ melhorar
o ajuste do modelo no sample.

Este ponto atípico pode estar comprometendo
o modelo no que se refere a heterocedasticidade,
normalidade, outros fatores.

Poderia estar condannando o modelo por conta
de uma observação.

Com intervenções \Rightarrow forma de acomodar obs. ao modelo.

Ex2:



\Rightarrow usa função escada.

Há efetivamente uma
mudança de nível
(que pode ser vista
a um período).

Há outras funções elaboradas.

Mora

Importante \Rightarrow distingui entre efeito transitório e permanente.

6

)

após apena

há uma

observação nesse

mudança

período t.

de estrutura.

Não sempre é possível facilmente verificar efeito transitório permanente na análise dos resíduos das regressões.

Uma forma é olhar se resíduos no nível das variáveis de estado.

↓

Só nos ME podemos fazer esta distinção, tratando resíduos no nível das variáveis de estado

(No modelo, um efeito transitório em η_t é permanente em y_t , devido a estrutura AR de μ_t)

$$\mu_t = \mu_0 + \varepsilon \eta_t$$

—+—

Diagnósticos

$$\text{Modelo: } y_t = \beta_t x_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_e^2)$$

$$x_{t+1} = T_t x_t + R_t \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

Hipóteses: i) linearidade das componentes
aditividade

x combinam de
forma linear
aditiva
A perman
ge

ii) $\varepsilon_t \rightarrow$ i) Normal

iii) homocedástico

iv) descorrelatado $\rightarrow E[\varepsilon_t, \varepsilon_{t-s}] = 0, \forall s \neq 0$

Método de ajuste: R^2 , MAPE, RMSE etc

(adequacia)

2014

Entonces el error es menor

$$\hat{v}_t = \frac{y_t - f_t}{\hat{\sigma}_t}$$

Y el error es menor que el error de predicción

Porque el error es menor

Porque el error es menor

Entonces el error es menor que el error de predicción

$$E_t = y_t - f_t + \hat{v}_t \sim N(0, \hat{\sigma}_t^2)$$

Entonces el error es menor que el error de predicción

Entonces el error es menor que el error de predicción

Entonces el error es menor que el error de predicción

$$E_t = y_t - 2f_t + \hat{v}_t \Rightarrow E_t \text{ más cercano}$$

$$\hat{v}_t = y_t - 2f_{t+1} \sim N(0, \hat{\sigma}_t^2)$$

Sabemos que $E(v_t) = 0$

$$E(v_t^2) = f_t$$

$$\Rightarrow \text{varianza redondeada. } \tilde{v}_t = \frac{v_t}{\hat{\sigma}_t} \sim N(0, 1)$$

Entonces el error es menor

Entonces el error es menor

Uma das diagnósticas é verificar se $\text{var}(\hat{V}) \approx 1$

$\text{var}(\hat{V}) \approx 1 \Rightarrow \text{Bom!}$

→ Normalidade:

Quais procedimentos podemos usar para verificar?

Histograma

Q-Q plot \circledast

testes estatísticos: $\left. \begin{array}{l} \text{Jarque-Bera (simetria e} \\ \text{curtose)} \end{array} \right\}$

A Darling

$$JB = \frac{n}{6} (\hat{s}^2 - 1) + \frac{n}{24} (\hat{k}^2 - 3) \sim \chi^2(2)$$

(estudar!)

$AD = \int (F_n - F_0)^2 df$ (relações entre densidade empírica

$F_0(1-F_0)$ e teórica)

\Rightarrow tem mais

sob H_0 ,

potencial/poder

do que JB

\circledast Q-Q plot

V_t

i. Dispersão dos pontos em volta de 45° com

alguma variabilidade estatística.

$N(0, 1)$

→ Homocedasticidade: variância do erro cte.

Pode ser feita de 2 formas

→ No Stata, busca um tipo específico de homocedasticidade

obs: pf x n é tamanha \Rightarrow usa-se m = 8 } pega tb diagonalidade
" " mensal \Rightarrow usa-se m = 15

(
Na FAC, olha tb pontual

r_j = autocorrelações das inovações

- - -

pf saber se há tb efeito AR(1), outra possibilidade é
trabalhar com estatística Ljung-Box $M v_t^2$

(b) teste Q para $v_t^2 \Rightarrow$ investigar efeito AR(1) em v_t

(
FAC pf $v_t^2 \Rightarrow v_t^2 = \alpha + \beta v_{t-1}^2 + \epsilon_t$
 $\text{com } (v_t^2, v_{t-1}^2) \neq 0.$

- - -

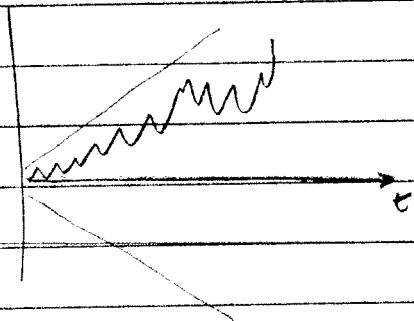
Outros tipos de teste

- Soma Acumulada das inovações²

↳ se modelo bem especificado,

fica dentro de uma banda

Se cai para, pode ser sinal de
heterocedasticidade



álibra

Preditores Auxiliares

e Inovações: $v_t = \hat{\epsilon}_{t|t-1} = y_t - \hat{y}_{t|t-1}$

• Resíduos auxiliares: $\hat{\epsilon}_t = E[\epsilon_t | Y_n]$

(cap. 4, sec. 4.4 A)

parece ✓
com inovações,

mas inovação
é baseada
até info de
 $t-1$.

R.Aux. é
concorrente
inovações
suavizada.

↳ valor esperado suavizado.

mostra que: $\hat{\epsilon}_t = H_t \mu_t, t = n, \dots, 1$

onde $\mu_t = F_t' v_t - K_t' r_t$ → do algoritmo de suavização
(calculado de forma recursiva)

$H_t = E[\epsilon_t \epsilon_t']$ → no caso univariado $= \sigma_{\epsilon}^2$

$\Rightarrow \text{var}(\hat{\epsilon}_t | Y_n) = H_t - H_t D_t H_t$

$D_t = F_t' + K_t' + N_t K_t$

↳ t do algoritmo de suavização.

Temos também que:

$\hat{\eta}_t = E[\eta_t | Y_n] \Rightarrow \alpha_{t+1} = T_t x_t + R_t \eta_t$
 $= Q_t R_t' r_t$ → estimativa suavizada de η_t (do nível do
nível da variável de estado).

$Q_t = E[\eta_t \eta_t']$

geralmente é constante $\text{var}[\hat{\eta}_t | Y_n] = Q_t - Q_t R_t' N_t R_t Q_t$

Além do mais: $\hat{\epsilon}_t$ e $\hat{\eta}_t$ são sair descorrelacionados, nem descorrelacionados entre si.

⇒ Resíduos auxiliares podem ser facilmente calculados a partir do algoritmo de suavização.

()

()

08/06/11.

faz a mesma
mas é mais formalizado
com o teste estatístico.
Mas só bem documentada em
livros e artigos.

Resíduos Auxiliares como ferramenta de diagnóstico

constroem
2 tipos de
resíduos, além
dos inovacões,
específicos de m-

a partir do
algoritmo de
suavizações

calculados a
cada instante

obs: Nas

outras M

componentes de

pt ou da

sazonalidade.

que principios,

ordem de tura,

as normalmente

modificando

nível esta o novo

ada a pt.

$$\hat{e}_t = E[\hat{e}_t | y]$$

$$\hat{\eta}_t = E[\hat{\eta}_t | y]$$

chocque da
componente pt.

$$y = (y_1, y_2, \dots, y_T) \quad t=1, 2, \dots, T$$

também usamos a notação

$$Y_T = (y_T, y_{T-1}, \dots, y_1) \equiv y$$

Nas conumbras
características
dos resíduos
de inovações

estimativas suavizadas

dos erros:

(1) são intra-correlacionadas (pois las

suavizadas
tem comp.
de info em
comum)

calcular la

Nas traípos

mas pt diapeis
ticas (mas per-

mentas e nas ua-

dos em testes
estatisticos)

Estratégia: Usar $\hat{\eta}_t$ e \hat{e}_t para distinguir entre choques transientes
e choques permanentes em y_T , de forma a me-

lhor caracterizar os analises de intervenções.

(descreve - detecta de patrões anomais nos resíduos, observações
anómalas)

$$\text{Suponha modelo nível local } y_t = \mu_t + \epsilon_t$$

(exemplo simples)

$$\mu_t = \mu_{t-1} + \eta_t \Rightarrow \mu_t = \mu_1 + \sum_{i=1}^t \eta_i$$

choque

pt ajuste

a posteriori

depois

de obter

var

componente

mento segun-

do o modelo.

pode ter ocorrido

uma situaçao

que o modelo

não seja capaz

de detectar

Dois tipos de choques → em ϵ_t : transiente

afetará y_t apenas em t

→ outlier: tratado via funções peso

$$D_t = \begin{cases} 1 & t=t^* \\ 0 & t \neq t^* \end{cases} \Rightarrow y_t = \mu_t + \delta D_t + \epsilon_t$$

→ em η_t : Dada recursividade de μ_t , em
virtude de sua natureza markovian.

Nas é uma permanente
eterno. Pode se com-
pensando com η_t subse-
quentes. Dependendo do
nível do choque, pode
ser sendo carregado.

e raiz unitária, um choque em η_j
 $1 \leq j \leq t$ será incorporado a μ_t de forma
permanente e, assim, afeta y_t também
de forma permanente.

→ Poda pt forma, calculando \hat{e}_t e $\hat{\eta}_t$. Detectar períodos onde $\hat{e}_t > k_1 \hat{\sigma}_{\hat{e}_t}$ e $\hat{\eta}_t > k_2 \hat{\sigma}_{\hat{\eta}_t}$
se pt que \hat{e}_t é coloca função peso e estima parâmetro δ . "Ajuda" externa ao modelo.
Isto só ajuda pt melhor estatísticas do modelo em si.

(*) Se detecta ocorrências de outliers.
→ Esta é a forma padrão que o stamp trata intervenções. Itál formas + sofisticadas de tratar.

Ex. Supor que em $t = t^*$ há um choque em η_{t^*} . \rightarrow informalmente
Patamar y_{t^*} é "grande" = $K\eta_t$, $K = 2, 2.5, 3$ etc.

Assim, para $t > t^*$, μ_t contraria este choque e, assim,

$$\begin{cases} y_t = \mu_t + \eta_t^* + \eta_{t-1} + \dots + \eta_1, & t = t^* \\ y_t = \mu_t + \eta_t + \eta_{t-1} + \dots + \eta_1, & t > t^* \end{cases}$$

Ou seja, se $t \geq t^* \Rightarrow y_t$ mudará de nível de forma permanente.

Assim, ao detectarmos um "outlier" em η_t , podemos tratá-lo como tal em η_t :

(a) via funções pulso

(b) como uma função escada $\begin{cases} 1 & t \geq t^* \text{ na equação de } y_t \\ 0 & t < t^* \end{cases}$

$$(a) y_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \delta D_t + \eta_t$$

$$D_t = \begin{cases} 1 & t = t^* \\ 0 & t > t^* \end{cases}$$

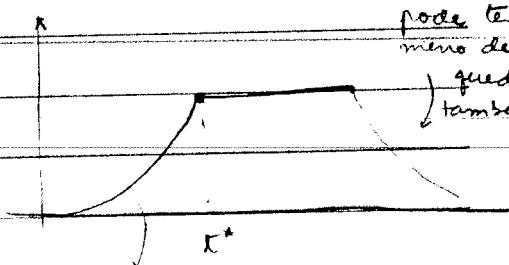
$$(b) y_t = \mu_t + \delta D_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \eta_t$$

$$D_t = \begin{cases} 1 & t \geq t^* \\ 0 & t < t^* \end{cases}$$

6

Outra alternativa:



Obs: dist. dos

resíduos auxiliares

também é assintó-

ticamente normal.

Cuidado! Teste de Norma

Lída de Jarque - Boa padrões

presupõe iid.

No caso dos resíduos auxiliares,

nas fas iid → perita usar teste

Jarque - Boa corrigido.

pode haver uma cota inicial

Ex: Lei de cinto de Esperança introduzida, Análise do n.º de acidentados no trânsito (letalidade)

$$\text{dy 2: } y_t = \phi y_{t-1} + \delta D_t + \epsilon_t$$

↳ pular em $t=t^*$

Mais formas de estimar

Dividir

→ MV const $\Rightarrow m_0 \Rightarrow \mu_{\epsilon,t} = 0$

→ MV incord $\Rightarrow y_0 \sim N(0, \frac{\sigma^2}{1-\phi^2}) \Rightarrow \mu_0 \approx 0$ (nas raízes)

Exemplo: série anual

1871 - 1970

Vazão do Nilo em 10^8 m^3

Importante conhecer a série pf entender questões dos outliers.

1899 \Rightarrow construção da barragem de Assuan \Rightarrow seca permanente (*1)

1913 \Rightarrow obs. atípica \Rightarrow transiente (*2)

Estratégia: (*1) em $\hat{\eta}_t$

(*2) em $\hat{\epsilon}_t$

→ Olha a série: Data Plot

→ Model

models

basic components: level \rightarrow stochastic

nível local

Estimation method: MV

→ Test menu

Auxiliary residual graphics

More written outputs \Rightarrow write absolute values

exceeding the value of [3]

Irregular: $\hat{\epsilon}_t$

level: $\hat{\eta}_t$

tilbra

Vieira > Results

Values larger \Rightarrow p value alto?
(prob) \Rightarrow rejeitar hipótese nula

Select interventions

Início de 1913

level \Rightarrow 1899

Tat $\xrightarrow{\text{menor}}$ component graphic

Nível praticamente cte. (linhareta) a menos da intervenções mudando o patamar.

$$\text{Provavelmente } \sigma^2_{\eta} \approx 0 \quad (\mu_t = \mu_{t-1} + \eta_t)$$

No Manual: p estimar models

e' colocado um medium cycle (10 anos)

Obs: Ter a possibilidade de intervenções transitórias e permanentes é inerente a noi.

Em modelo ARIMA, p.ex., tem uma fonte apenas de choque.

Modelos Estruturais Multivariados

- Generalizações de modelos estruturais univariados.

Mais complexos em relação a estimativas (f_k e m_v), identificacões paramétrica, interpretacões e diagnósticos.

$$\text{univariado} \Rightarrow y_{it} = \mu_{it} + \gamma_{it} + \varepsilon_{it}$$

↳ diagnósticos

$$V_{it} = \frac{y_{it} - \bar{y}_{1:t-1}}{F_{it}^{1/2}}$$

(testes JB, B(jung, Ptc.)

$$\text{multivariado} \Rightarrow \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_t \\ \gamma_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$\Rightarrow V_t = \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix} \sim N(0, F_t)$$

tests pp saber se das

normais bivariadas.

(se fôs N bi \Rightarrow cada marginal

mai N univariado, mas testar

Normal univariado mais suficiente H

verificar se vale Normal bivariada)

pode resíduos.

Tels de diagnósticos precisam ser bi, tri, ..., multivariados.

= Teste bivariado para normalidade

Auto correlações - o inovações inter e intra
etc. descorrelatados
(testes conjuntos)

Motivações:

flexibilidade → estimar componentes comuns entre ST's.

→ melhorar influência e previsões / simulações.

$$z_t = f(y_{1t}, y_{2t})$$

t → fatores de risco ext: z_t : lucro Pt de

resultado e

y_{1t} : preço petróleo

Variável comum

y_{2t} : " gasolina.

(ex: dados financeiros)

Se tratar y_{1t} e y_{2t} de forma separada, perde possibilidade de ter influências / ligações diretas entre os 2.

É indicado por razões econômicas, financeiras o tratamento de fatores de risco de forma multivariada.

de forma mais geral, pode ser complexo de se tornar exato (muito acadêmico) mas na prática indicam-se alguns ajustamentos.

1) Modelo SURSE: seemingly unrelated TS equations

$$y_t = \mu_t + \epsilon_t \quad \rightarrow p \times 1$$

$\rightarrow p \times p$

$$\mu_{t+1} = \mu_t + \eta_t \quad \rightarrow p \times 1$$

$\rightarrow p \times p$

Se $p=5 \Rightarrow 15$ parâmetros de Σ_ϵ e

" " de Σ_η

obs: N° de parâmetros é $n + \binom{n}{2}$

/ \

Variâncias covariâncias

última

Então, se não tivermos nenhuma restrição, estariamos falando de 30 parâmetros.

Para $p=2$

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

y_t não depende de μ_t

y_t " " de μ_t

→ mas x comunicam explicitamente.

$$\begin{pmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} \end{pmatrix} = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

A comunicação vem das covariâncias. Se fizer teste de hipótese e covariâncias

$$\epsilon_t \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\epsilon_1}^2 & \sigma_{\epsilon_1 \epsilon_2} \\ \sigma_{\epsilon_1 \epsilon_2} & \sigma_{\epsilon_2}^2 \end{pmatrix} \right)$$

estatisticamente nulas

→ devem tratar séries separadamente

$$\eta_t \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_1 \eta_2} \\ \sigma_{\eta_1 \eta_2} & \sigma_{\eta_2}^2 \end{pmatrix} \right)$$

⇒ Nesta forma geral, o SUTSE é de difícil estimacão.

— " —

Queremos um modelo SUTSE + simples, cf menos parâmetros.

Definições: Modelos homogêneos

= Processo estocástico de 2ª ordem, p-variado, é dito homogêneo se todas as combinações lineares de suas componentes possuem as mesmas propriedades estocásticas de 2ª ordem.

Em outras palavras, $y_t = (y_{1t}, y_{2t}, \dots, y_{pt})$ é homogêneo se $H \in \mathbb{R}^p$, o processo univariado estacionário $z_t = (A^d)(x^t)y_t$ possui FAC independente de x .

pode ser

que y mas xjam estacionários (xjam necessários de purificar)

como a FAC é a unif. H

qualquer combinação, mas pode depender de x .

$$\text{Ex: } z_t = 1y_{1t} + 2y_{2t}$$

$$0,5y_{1t} + 0,7y_{2t}$$

Y FAC não depende de x

Lembrar que é difícil de mostrar pq g p = 2.

Proposição 1: Se y_t segue um SUTSE de nível local, então y_t será homogêneo SSS

$$, \text{ se } p=5 \left\{ \begin{array}{l} \text{SUTSE Geral: } \#4=3 \\ \text{SUTSE homog: } \#4=1 \end{array} \right.$$

$$\Sigma_{\eta} = q \Sigma_{\varepsilon}, q \in \mathbb{R}^+$$

$$\text{WUTSE homog: } \#4=1 \quad (15+q)$$

- $\mu_t, \delta_t, \gamma_t$ etc.

Proposição 2: se y_t segue um SUTSE com j tipos de componentes ortogonais, então y_t será homogêneo SSS

$$\Sigma_{\eta} = \begin{pmatrix} \Sigma_{\eta_1} & & & & \\ & \ddots & & & \\ & & \Sigma_{\eta_2} & & \\ & & & \ddots & \\ & & & & \Sigma_{\eta_j} \end{pmatrix} = \begin{pmatrix} q_1 \Sigma_{\varepsilon_{11}} & & & & \\ & \ddots & & & \\ & & q_2 \Sigma_{\varepsilon_{22}} & & \\ & & & \ddots & \\ & & & & q_j \Sigma_{\varepsilon_{jj}} \end{pmatrix}$$

como processo
de componentes
ortogonais
so há comunicação (?)

1 - Tendência
2 - Razionalidade onde $\Sigma_{\varepsilon} = h \Sigma_{\varepsilon^*}$

DIVIDA
cada Σ_{η_i} é que é
um bloco por componentes
ortogonais e vai de 1 a p
entre si mesmos e não a j

Propriedades Empíricas de PH's. (processos homogêneos)

- Pode haver teorias "a priori" que justifiquem a adoção de estrutura de proporções entre as covariâncias
- As regras de Kalman podem ser implementadas para cada uma das p. equações separadamente (ver Harvey)
(livro azul, + grosso.)
- Há menos risco de não identificacões paramétricas pela baixa dimensão de Ψ .
→ processo homogêneo é sempre identificável.
- Estimação por MV é consideravelmente simplificada.

Tudo pode ser estudado se nenhuma complicação

MML, q razoabilidade, d'ciclo etc.

22/06/11

seja levado em conta

MNL - multivariado

i) SUTSE - Irrestrito - Modelos + que não diz respeito a parâmetros.

Muitos parâmetros p/ estimar.

Problemas de convergência

e estimações.

uma multiplicação e'

usar modelo trinomial

$$\Sigma_{\eta} = g \Sigma_{\epsilon}$$

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_{\epsilon})$$

$\downarrow p \times 1$

$\downarrow p \times p$

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \Sigma_{\eta})$$

$\downarrow p \times p$

$$\Sigma_{\eta} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{K \times p \times K}$$

ii) Parametrizações do SUTSE - Na aula passada, avaliamos $H = p = 2$.

Poderemos sempre redenomir $\alpha_t^+ = \beta x_t$

A definição do vetor de estado α_t

é totalmente arbitrária. Podemos

sempre ter uma outra

parametrização que leve a

uma outra interpretação, mas

com mantendo propriedade e

verossimilhança.

Ver desenho
de vetores
no espaço
e aula
passada

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} I_K & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} \mu_t^+ \\ \bar{\mu}_t^- \end{pmatrix}$$

$(p-k) \times 1$

$$\begin{pmatrix} \mu_t^+ \\ \bar{\mu}_t^- \end{pmatrix} = \begin{pmatrix} \mu_{t-1}^+ \\ \bar{\mu}_{t-1}^- \end{pmatrix} + \begin{pmatrix} \eta_t^+ \\ \bar{\eta}_t^- \end{pmatrix}$$

$$\downarrow (p-k) \times 1 \qquad \qquad \qquad N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{\eta}^+ & 0 \\ 0 & \bar{\Sigma}_{\eta}^- \end{pmatrix} \right)$$

Parametrização
não p/ distingui a
questões de componentes
comuns.

$$\text{onde } \Sigma_{\eta}^+ = \Sigma_{11} \quad e \quad T = \Sigma_{21} \Sigma_{11}^{-1}$$

iii) Modelos com níveis comuns - agora: determinar caso existência
de comp. comuns (neste caso do MNL, estamos buscando no m.

se posto $(\Sigma_{\eta}) = k < p \Rightarrow \mu_t^+$ contém k níveis comuns

sendo $\Sigma_{\eta}^+ p.d. \quad e \quad \bar{\Sigma}_{\eta}^- = 0$, entao pode-se escrever (II) como:

casos importantes

p/ dizer

quantas

componentes L.I.

tem independentes.

Se nosso posto cheio \Rightarrow não há

$$\bar{\mu}_t = \bar{\mu}_{t-1} \Rightarrow$$
 etc. determinística = $\bar{\mu}$

tilibra

$$y_{1t} = \mu_t^+ + \epsilon_{1t}$$

$$(p-k) \times 1$$

$$y_{2t} = \Pi \mu_t^+ + \bar{\mu} + \epsilon_{2t}$$

Forma simplificada do MEC onde

1ª k componentes são direcio-

nadas por μ_t^+ e outras ($p-k$)

são c.c de μ_t^+

K

$$\mu_t^+ = \mu_{t-1}^+ + \eta_t^+, \quad \eta_t^+ \sim N(0, \Sigma_\eta^+) \quad \rightarrow \text{apenas } k \text{ componentes evoluem}\newline \text{segundo processo linear local}$$

Assim, na prática, saber se um determinado ME possui

"componentes comuns" pode ser inferido observando-se o posto

linear da matriz covariânia dos choques dessa componente (no caso
especial de $p=2$ ($\Rightarrow \rho_{\eta_1, \eta_2} = 1$))

Na prática, partindo de um SUTSE genérico,
queremos ver se há comp. comuns (**)

(Outra p/ posto Σ_η)

Nyblom & Harvey (2001) desenvolvem um teste p/ componentes comuns
dentro do arcabouço de ME. \rightarrow Testes + formais p/ verificar se
posto Σ_η é $< p$.

-" -

Modelos Rotacionados

que possuem k fatores comuns

Muitas vezes, o modelo tipo (IIa) não é de fácil interpretação.

Entretanto, como as componentes NO não são únicas, podemos
procurar um outro conjunto de componentes comuns que a-
presentem melhor interpretação.

Considere as equações do modelo (IIa)

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} I_k \\ \Pi \end{pmatrix} \mu_t^+ + \epsilon_t$$

$$\therefore \mu_t^+ = \mu_{t-1}^+ + \eta_t^+ \quad (*)$$

para obter uma matriz de var
cova diagonal.

Seja $\mu_t^* = (\Sigma_{\eta}^+)^{-1/2} \mu_t^+$ a "nova componente"

$$\Rightarrow \mu_t^* = (\Sigma_{\eta}^+)^{-1/2} \mu_t^+$$

de (*):

$$(\Sigma_{\eta}^+)^{-1/2} \mu_t^+ = (\Sigma_{\eta}^+)^{-1/2} \mu_{t-1}^+ + (\Sigma_{\eta}^+)^{-1/2} \eta_t^+$$

$$\Rightarrow \mu_t^* = \mu_{t-1}^* + \eta_t^* \text{ onde } E[\eta_t^* \eta_t^{*'}] = I_K$$

$$\mu_t^* = \mu_{t-1}^* + \eta_t^* \quad (\text{Provar!})$$

Por outro lado:

$$y_t = \begin{pmatrix} I_K \\ \pi \end{pmatrix} \mu_t^* + \begin{pmatrix} 0 \\ \bar{\mu} \end{pmatrix} + \epsilon_t$$

$$= \begin{pmatrix} I_K \\ \pi \end{pmatrix} \Sigma_{\eta}^{1/2} \mu_t^* + \begin{pmatrix} 0 \\ \bar{\mu} \end{pmatrix} + \epsilon_t \quad (\text{a})$$

Seja $\pi^+ = (I, \pi')$

$$\textcircled{1} = \pi^+ \Sigma_{\eta}^{-1/2}$$

$$\mu_0 = (0, \bar{\mu})'$$

Então (a) pode ser resrito como:

$$\left\{ \begin{array}{l} y_t = \textcircled{1} \mu_t^* + \mu_0 + \epsilon_t \quad \epsilon_t \sim N(0, \Sigma_{\epsilon}) \\ \mu_t^* = \mu_{t-1}^* + \eta_t^* \quad \eta_t^* \sim N(0, I_K) \end{array} \right.$$

E o ponto de partida só conseguimos a fazer rotacionar (a partir
da obtenção das componentes ortogonais)

Nessa nova parametrização (notas), os fatores comuns nas variâncias unitárias.

Se $(\Sigma_{\eta}^+)^{1/2}$ é triangular superior, entao θ é da forma $\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$ também (e só isso!).

Ex: $\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \mu_{3,t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}$

$\Rightarrow \theta_{ij} = 0$ para $j > i$

obs: CORREÇÕES nas NOTAS:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} \theta_{11}, 0 \\ \theta_{21}, \theta_{22} \\ \theta_{31}, \theta_{32} \end{pmatrix} \begin{pmatrix} \mu_{1t}^+ \\ \mu_{2t}^+ \\ \mu_{3t}^+ \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}$$

Pg 15: $\mu_t^+ = \mu_{t-1}^+ + \eta_t^+$

$$\Rightarrow y_{1t} = \mu_{1t}^+ + \epsilon_{1t} \quad \text{poderia ter WT}$$

$$y_{2t} = \theta_{21} \mu_{1t}^+ + \theta_{22} \mu_{2t}^+ + \epsilon_{2t}$$

$$y_{3t} = \theta_{31} \mu_{1t}^+ + \theta_{32} \mu_{2t}^+ + \epsilon_{3t} \quad \text{poderia ter direc.}$$

até $\mu_t^+ = \mu_{t-1}^+ + \eta_t^+$

— II —

(**) Na prática, partindo de um WTSF comum, se $p=2$

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \Rightarrow \hat{\rho} = \frac{\sigma_{12}}{\sigma_1 \sigma_2} \quad \text{se } \hat{\rho} \approx 1 \Rightarrow \text{não uma comp. comum.}$$

(obs: se $p=2$ podemos ter até uma comp. comum.

se p genérico \Rightarrow pode ter k comp. comuns).

Se tivermos modelando simultaneamente 3 séries:

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} \quad \begin{array}{l} \text{petróleo} \\ \text{gasolina} \\ \text{gas.} \end{array} \quad \begin{array}{l} \text{quando há comp. comum} \\ \Rightarrow \text{há interações} \\ \text{(movimentos nos preços das parcerias)} \\ \text{Transfere os petróleo pf derivados}) \end{array}$$

Obs: p-valor

espectro RB \Rightarrow cte.

MADF: p_f para da amostra

\Rightarrow deixar 6 obs. para da amostra

— — —

Exemplo:

Dentro do arco banco ME

- Tornando 2 séries, como avaliar se cointegrar
- Seja usar FOTSE, tem tendência e verifica correlação entre 2 séries.

usando FOTSE, outras matrizes de variancia do p.

— — —

Ex. 4: com models

Gerar série c/ 300 obs por MC (série artificial)

Como sabe valor verdadeiro, usa como valor inicial algo perto do valor verdadeiro.

$$\text{Ex: } \sigma_\theta^2 = 10^{-3}$$

$$\text{com } \theta = 0,01$$

$$\sigma_\eta^2 = 0,01$$

$$y(t) = 0,1$$

$$\sigma_\zeta^2 = 0,0001$$

$$0,1$$

$$\sigma_w^2 = 0,2$$

$$0,1$$

outros
coisa,
seja
série com
tend. e sazonal
idade
determinística

Optimizar as opções.

Mas valores EIR

\Rightarrow usa log-verosimil: $\ell = l + \exp((\eta))$

(a cada iteração)

$\underbrace{\ell}_{\text{EIR}}$ equilíbrio

$$\sigma_\eta$$

Outra coisa: calcular a variação infinitesimal $\Rightarrow \frac{t}{n}$
quando passar p/
a ordem (as
acabar iteradas) (dividir por n de obs)

É essa parte p/ a sér. real.

→ Outro ponto: concentre-se na infinitesimal

MLN SUTSE homogêneo

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon)$$

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta)$$

preciso fer
estimadas.

$$\Sigma_\eta = q \Sigma_\varepsilon \Leftrightarrow \begin{pmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_1 \eta_2} \\ \sigma_{\eta_1 \eta_2} & \sigma_{\eta_2}^2 \end{pmatrix} = q \begin{pmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_1 \varepsilon_2} \\ \sigma_{\varepsilon_1 \varepsilon_2} & \sigma_{\varepsilon_2}^2 \end{pmatrix}$$

outras
por construções
seus automaticamente
obtidas.

~~$$p=2$$~~
~~$$y_{it} = \mu_{it} + \varepsilon_{it}$$~~

$$\Rightarrow \Delta y_{it} = \eta_{i,t-1} + \Delta \varepsilon_{it}, \quad i=1,2$$

$$\mu_{it+1} = \mu_{it} + \eta_t$$

~ MA(1) (I)

$$\text{Assume: } z_t = w_t + \theta w_{t-1}, \quad w_t \sim N(0, \sigma_w^2)$$

$$E[z_t] = 0$$

Nas fazy.
verificando.

$$E[z_t^2] = (1+\theta^2) \sigma_w^2 \Rightarrow p_z(1) = E[z_t z_{t-1}] = 0$$

var(z_t) $(1+\theta^2)$

Reveria
olhar MA bivariado. $E[z_t z_{t-1}] = \theta \sigma_w^2$ (II)

$$\text{De (I): } E[\Delta y_{it}] = 0$$

$$E[\Delta y_{it}^2] = \sigma_{\eta_i}^2 + 2\sigma_{\varepsilon_i}^2 \Rightarrow P_{\Delta y_{it}}(1) = -\frac{\sigma_{\varepsilon_i}^2}{(\sigma_{\eta_i}^2 + 2\sigma_{\varepsilon_i}^2)} \quad \text{FAC das}$$

$$E[\Delta y_{it}, \Delta y_{i,t-1}] = -\sigma_{\varepsilon_i}^2 \quad \text{primeira reduzida de cada uma das equaçõeas col do SUTSE}$$

$$\text{mas por homogeneidade: } \sigma_{\eta_i}^2 = (q)\sigma_{\varepsilon_i}^2, \quad i=1,2$$

$$\Rightarrow P_{\Delta y_{it}} = -\frac{\sigma_{\varepsilon_i}^2}{q\sigma_{\varepsilon_i}^2 + 2\sigma_{\varepsilon_i}^2} = -\frac{1}{q+2} \Rightarrow \text{independe de } i \quad (i=1,2)$$

(é a razão p1 yit, yit. Se tornarmos t6

C.L. de x's. \Rightarrow nas. máx alterado).

$$V_t = \alpha_1 \Delta y_{1t} + \alpha_2 \Delta y_{2t} \Rightarrow \text{FAC: } p(1) \text{ também ilibra}$$

$\frac{1}{MA(1)} \quad \frac{1}{MA(1)}$

V_t independe de α_1 e α_2 .

cada um dos MA
tem a mesma dinâmica

Portanto, na forma reduzida do SUTCE homogêneo, as séries
possuem a mesma dinâmica

→ trata-se de um modelo bem restritivo.

$$\text{Estimacões} \Rightarrow \Psi = (\sigma_{\epsilon_1}^2, \sigma_{\epsilon_2}^2, \sigma_{\epsilon_{t-1}}^2, q), \quad \tilde{y}_{t-1} = (\tilde{y}_{1,t-1}, \tilde{y}_{2,t-1})$$

$$l(\Psi) = \log L(\Psi) = \sum_{t=1}^n \log p(y_t, y_{t-1} | \tilde{y}_{t-1}) \quad \begin{array}{l} \text{funções densidade} \\ \text{conjunta bivariada.} \end{array}$$

$$= \sum_{t=1}^n \log \left[\frac{1}{(2\pi)^{|F_t|/2}} \exp \left\{ -\frac{1}{2} \tilde{v}_t^T F_t^{-1} \tilde{v}_t \right\} \right]$$

Usando inicializações
dipos exatas.

$$= -n \log 2\pi - \frac{1}{2} \sum_{t=1}^d w_t - \frac{1}{2} \sum_{t=d+1}^n \log |F_t| - \frac{1}{2} \sum_{t=d+1}^n \tilde{v}_t^T F_t^{-1} \tilde{v}_t$$

Se usarmos bfg
kappa, usas

mas fast afetados pelo inicio exata.

um termo $\sum w_t$.

$$w_t = \begin{cases} \log |F_{0,t}| & , f_{0,t} \text{ é pos. definida.} \\ \log |F_{0,t}| + \gamma_p^{(0)} f_{0,t}^{-1} v_t^{(0)}, & f_{0,t} = 0 \end{cases}$$

~~Divide para a ser par ($t=1$)~~
~~ou $t=d$?~~ (quanta componentes fas mas stoc.)

Obs: Modelo SUTCE pode ser generalizado se o caso q + componentes

$$y_t = \mu_t + \epsilon_t$$

$$y_{it} = \mu_{it} + \epsilon_{it}$$

covariâncias entre

Variâncias de tendências

Aula passada: modelo linear SUTSE $\Rightarrow y_t = \mu_t + \varepsilon_t$ $y (p \times 1)$ multivariado

Cada y_{it} depende de sua componente e de seu erro, basta como as componentes.

A dependência vem da estrutura das matrizes de covariâncias (E_t, Σ_t)

Modelos SUTSE (contínuas)

Precisamos calcular algumas hipóteses: p dimensões n de parâmetros a ser estimado. Ex: Homogeneidade ($Z_1 = g Z_0$)

É possível generalizar a estrutura SUTSE de forma a incorporar outros componentes, e.g., linear estocástica, sazonalidade e ciclo.

Vamos supor ortogonalidade entre componentes de interpretações diferentes ($\mu_t \perp \delta_t$, $\mu_t \perp \eta_t$ etc.)

considere $p=2$ (p modelo de nível e ciclo, onde nível segue modelo linear estocástico)

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

3 parâmetros

$\hookrightarrow \varepsilon_t \sim N(0, \Sigma_\varepsilon)$

$$\begin{pmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} \end{pmatrix} = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} + \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

matrizes rasas
cheias

$\hookrightarrow \eta_t \sim N(0, \Sigma_\eta)$

3 par.

$$\begin{pmatrix} \beta_{1,t+1} \\ \beta_{2,t+1} \end{pmatrix} = \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \end{pmatrix} + \begin{pmatrix} \xi_{1t} \\ \xi_{2t} \end{pmatrix}, \quad \xi_t \sim N(0, \Sigma_\xi)$$

3 par.

$$\begin{pmatrix} \eta_{1,t+1} \\ \eta_{2,t+1} \\ \eta_{1,t+1}^* \\ \eta_{2,t+1}^* \end{pmatrix} = \begin{bmatrix} 0 < p < 1 \\ P & \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \otimes I_2 \end{bmatrix} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{1,t}^* \\ \eta_{2,t}^* \end{pmatrix} + \begin{pmatrix} K_{1t} \\ K_{2t} \\ K_{1t}^* \\ K_{2t}^* \end{pmatrix}$$

$\lambda_c = \frac{2\pi}{T_c}$ (a razão entre 2 séries)

$$K_t \sim N(0, \Sigma_K)$$

3 par.

2x2

$$\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix} \sim 4 \times 4$$

$$(\text{pois } \sigma_k^2 = \sigma_{k*}^2)$$

é ilírica

Obs: cada matriz Σ apresenta 3 parâmetros a serem estimados.

(11) Colocando o modelo de tendências linear stocástica e ciclo

SUTSE na forma EE:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}_{2 \times 8} \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \\ \beta_{1t} \\ \beta_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

importante a ordem
de entrada do vetor
de estado.
(se muda, encontra-se
uma nova forma EE)

$$\begin{pmatrix} \Psi_{1t} \\ \Psi_{2t} \\ \Psi_{1t}^* \\ \Psi_{2t}^* \end{pmatrix}_{8 \times 1}$$

$$\begin{array}{c|c|c|c|c|c} \text{T} & & & \text{R} & & \\ \hline \mu_{1,t+1} & 1 & 0 & 1 & 0 & \mu_{1t} & 1 \\ \mu_{2,t+1} & 0 & 1 & 0 & 1 & \mu_{2t} & 1 \\ \beta_{1,t+1} & 0 & 0 & 1 & 0 & \beta_{1t} & 1 \\ \beta_{2,t+1} & 0 & 0 & 0 & 1 & \beta_{2t} & 1 \\ \Psi_{1,t+1} & & a & 0 & b & 0 & \Psi_{1t} & 0 \\ \Psi_{2,t+1} & 0 & & & a & 0 & \Psi_{2t} & 1 \\ \Psi_{1,t+1}^* & \sim & c & 0 & a & 0 & \Psi_{1t}^* & 1 \\ \Psi_{2,t+1}^* & & 0 & c & 0 & a & \Psi_{2t}^* & 1 \end{array}$$

$8 \times 1 \qquad 8 \times 8 \qquad \downarrow \qquad 8 \times 8$

$$a = p \cos \lambda_c$$

$$b = p \sin \lambda_c \Rightarrow c = -b$$

$$c = -p \sin \lambda_c$$

Considere as matrizes do sistema do modelo univariado equivalente:

(É interessante ver uma outra forma de escrever a forma EE.

Hibrida

baseada nas matrizes do modelo univariado original.

equivalentes p/ o modelo univariado /
com as r^{as} componentes. ($y_t = \mu_t + \eta_t + \varepsilon_t$)

$$Z_0 = (1 \ 0 \ 1 \ 0) \quad T_0 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & a \end{pmatrix} \quad R_0 = I_4$$

Então, segue que:

$$- Z_0 \otimes I_2 = (1 \ 0 \ 1 \ 0) \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

dimensões p
(do vetor y_t)

$= 2$ (matriz 2 multivariada futurE)

$$- T_0 \otimes I_2 = T \text{ (movar!)}$$

$$- R = R_0 \otimes I_p \equiv I_8$$

E assim podemos escrever, de uma forma geral,

H SUTSE como: (qualquer que seja a combinação de componentes do SUTSE)

$$\left\{ \begin{array}{l} y_t = (Z_0 \otimes I_p) \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(\underline{\theta}, \Sigma_\varepsilon) \\ \alpha_{t+1} = (T_0 \otimes I_p) \alpha_t + (R_0 \otimes I_p) \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta) \end{array} \right. \Rightarrow \text{Notações} \\ + \text{conota}$$

$$\alpha_{t+1} = (T_0 \otimes I_p) \alpha_t + (R_0 \otimes I_p) \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta)$$

Dividir:
matrix é
dividida
entre
componentes?

Se $A \sim m \times n$

$B \sim p \times q$

$$A \otimes B \hat{=} \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{pmatrix} \sim mp \times rq$$

tilibra

Fatores comuns e cointegradas - (aposta)

i) Modelos de nível local

Inicialmente considere o SUTSE bivariado.

$$(I) \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon) \quad \text{dependência vira do fato que matrizes das chias (nunca comp. de nua a séri e transferida pra outra)}$$

$$\begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \quad \eta_t \sim N(0, \Sigma_\eta)$$

onde $\Sigma_\eta = \begin{pmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_1 \eta_2} \\ \sigma_{\eta_1 \eta_2} & \sigma_{\eta_2}^2 \end{pmatrix}$, $\rho_\eta = \sigma_{\eta_1 \eta_2} / \sigma_{\eta_1} \sigma_{\eta_2}$; $\sigma_{\eta_1 \eta_2} = \rho_\eta \sigma_{\eta_1} \sigma_{\eta_2}$

O modelo (I) pode ser resrito como:

$$(II) \begin{aligned} y_{1t} &= \mu_{1t} + \varepsilon_{1t} \\ y_{2t} &= \bar{\mu}_t + \bar{\varepsilon}_{2t} + \varepsilon_{2t} \\ \mu_{1t} &= \mu_{1,t-1} + \eta_{1t} \\ \bar{\mu}_t &= \bar{\mu}_{t-1} + \bar{\eta}_t \end{aligned} \rightarrow \text{sistema (II) é reparametrizado}$$

do (I) mas preserva invariante algumas coisas importantes ②

onde $\bar{\mu} = \rho \frac{\sigma_{\eta_2}}{\sigma_{\eta_1}}$, var $\begin{pmatrix} \eta_{1t} \\ \bar{\eta}_t \end{pmatrix} = \begin{pmatrix} \sigma_{\eta_1}^2 & 0 \\ 0 & (1-\rho^2) \sigma_{\eta_2}^2 \end{pmatrix}$ (nas chias)

Aqui, a matriz é diagonalizável!

Os sistemas (I) e (II) são equivalentes pois: \Rightarrow mostras (i) e (ii)

e estas

② (i) a estrutura de dependência dos y's permanece inalterada

de prova

卷之三

bivalviada

(ii) a previsão dos y_i 's e a função de verossimilhança permanecem inalteradas.

(x calcularmos forma reduzida, F.A.C etc. \Rightarrow tudo será igual).

Essencialmente, o sistema (II) é uma nova parametrização do sistema (I), obtida por uma transformação do tipo:

$$\alpha_t = B \alpha_t^*, \quad \det(B) \neq 0.$$

(transformações nas regras do voto de estado)

Através de una

nova parâmetrização
para a identificação
de componentes
comuns.

$$\alpha_t = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \pi & 1 \end{pmatrix} \begin{pmatrix} \mu_{1t} \\ \bar{\mu}_t \end{pmatrix}$$

↓ ↓ ↓

(I) B (II)

$$(\det B = 1 \neq 0)$$

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \pi & 1 \end{pmatrix} \begin{pmatrix} \mu_{1t} \\ \bar{\mu}_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$\text{obs: } \Delta (\text{I}) \mu_{2t} = \mu_{2,t-1} + \eta_{2t}$$

→ Nas notas, assim está definido η_F .

$$\text{mas } |\bar{\eta}_E = \eta_{2E} - \pi \eta_{1E}| \quad \therefore \eta_{2E} = \bar{\eta}_E + \pi \eta_{1E}$$

Ino p. que
Ostenhamas
Var (Niv) + diapause

$$\text{Logo: } \mu_{2,t} = \mu_{2,t-1} + \bar{\eta}_t + \pi \eta_{jt}$$

$$\text{entonces } B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\therefore \mu_{2t} = \mu_{2,t-1} + \bar{\eta}_t + \pi(\mu_{3t} - \mu_{1,t-1})$$

$$\therefore \mu_{2,t} - \pi \mu_{1,t} = (\mu_{2,t-1} - \pi \mu_{1,t-1}) + \bar{\eta}_t$$

$$\therefore \bar{\mu}_t = \bar{\mu}_{t-1} + \bar{\eta}_t$$

$$\bar{\mu}_t = \mu_{2,t} - \pi \mu_{1,t} \quad \therefore \quad \mu_{2,t} = \bar{\mu}_t + \pi \mu_{1,t}$$

usando as linhas componentes

dos 2 sistemas (I) e (II)

$$\therefore \mu_{2,t} = (\pi \rightarrow) \begin{pmatrix} \mu_{1,t} \\ \bar{\mu}_t \end{pmatrix}$$

$$\text{Logo: } \begin{pmatrix} \mu_{1,t} \\ \bar{\mu}_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \pi & 1 \end{pmatrix} \begin{pmatrix} \mu_{1,t} \\ \bar{\mu}_t \end{pmatrix}$$

dm:

$$\text{Do sistema (I): } \underline{\mu}_t = \underline{\mu}_{t-1} + \underline{\eta}_t$$

$$\text{Da transformada: } \underline{\mu}_t = B \underline{\mu}_t^* \quad \therefore \quad \underline{\mu}_t^* = B^{-1} \underline{\mu}_t$$

$$\Rightarrow \underbrace{B^{-1} \underline{\mu}_t^*}_{\underline{\mu}_t^*} = B^{-1} \underline{\mu}_{t-1} + B^{-1} \underline{\eta}_t$$

$$\underline{\mu}_t^* = \underline{\mu}_{t-1} + \underline{\eta}_t^*$$

$$\text{neste caso: } B^{-1} = \begin{pmatrix} 1 & -\pi \\ 0 & 1 \end{pmatrix}$$

$$\text{Com esse } B^{-1} \Rightarrow \text{obtemos } \underline{\eta}_t^* = \underline{\eta}_{2,t} - \pi \underline{\eta}_{1,t}$$

\Rightarrow A parametrização (II) nos permite investigar

(copia)

(III)

convergências entre os choques das 2 tendências.

(pg 10 das notas)

Se $\rho = \pm 1 \Rightarrow$ existiria uma componente comum necessariamente.

• Se $\rho_\eta = \pm 1 \Rightarrow \text{var}(\bar{\eta}_t) = (1 - \rho_\eta^2) \sigma_{\eta_t}^2$

(cond. suficiente p/ que só haja comp.

$\therefore \text{var}(\bar{\eta}_t) = 0$

de tendência comum

$\hookrightarrow \bar{\mu}_t$ será constante

$\Rightarrow \bar{\mu}_t = \bar{\mu}_{t-1} = \bar{\mu} + \nu_t$

E, assim, temos apenas 1 componente comum, pois:

$$y_{1t} = \mu_t^+ + \varepsilon_{1t}$$

$$y_{2t} = \pi \mu_t^+ + \bar{\mu} + \varepsilon_{2t}$$

$$\mu_t^+ = \mu_{t-1}^+ + \eta_{t-1}^+, \quad \eta_t^+ \sim N(0, \sigma_{\eta_t}^2)$$

\Rightarrow existe componente comum em y_{1t}, y_{2t}

• Se $\pi = 1$ (dado que $\rho_\eta = \pm 1$)

\hookrightarrow ν_t pode ser +, pois $\pi = \rho \sigma_{\eta_t} / \sigma_{\eta_1}$

Temos modelo dinâmico

$$y_{1t} = \mu_t^+ + \varepsilon_{1t}$$

de balanced growth

$$y_{2t} = \mu_t^+ + \bar{\mu} + \varepsilon_{2t}$$

$$= \mu_{t-1}^+ + \eta_{t-1}^+ + \bar{\mu} + \varepsilon_{2t} \quad \Rightarrow \quad \mu_{t-1}^+ - \mu_t^+ = \bar{\mu} : \text{diferença entre as tendências é constante } \nu_t$$

\Rightarrow "balanced growth"

• Se $\bar{\mu} = 0 : \mu_{t-1}^+ = \mu_t^+$

\Rightarrow tendências idênticas

- 11 -

Nível comum (\Rightarrow co-integrações)

(Varemos def. qual as co-integrações e vamos apoiar p/ este caso em particular).

equilíbrio

Definições de co-integração:

Um processo estocástico p-variado $y_t = (y_{1t}, y_{2t}, \dots, y_{pt})'$

é dito co-integrado de ordem d e b , $b \leq d$, i.e.

• Números inteiros.

y_t é CI(d, b) se

cada um dos elementos é integrado da mesma ordem.

(i) $y_{jt} \sim I(d)$, $j = 1, 2, \dots, p$

(ordem: n.º de vezes que precisas diferenças y_t para tornar estacionária)

(ii) $\exists \alpha \in \mathbb{R}^p - \{0\}$ tal que $\alpha'y_t \sim I(d-b)$

(Nenhum dos elementos de $\alpha = 0$)

$\delta^d y_t \sim \text{estac.}$

($d=0 \sim I(0)$)

Exemplo: Se $y_{jt} \sim I(1)$ e satisfaçõe $\alpha'y_t \sim I(0)$

$d=1$

• existe comb. linear das séries que seria estacionária

$b=1$

combinação linear perfeita
de ordem de
integração < do
que a de cada série individualmente.

Intuições:

Séries nas estacionárias nas tempos atuais

mas há uma classe que pode variar de forma nas estacionárias, mas podem estar em tendências, mas divergir muito (andam ± em fase) por conta de forças (económicas etc.) comuns que atuam sobre elas.

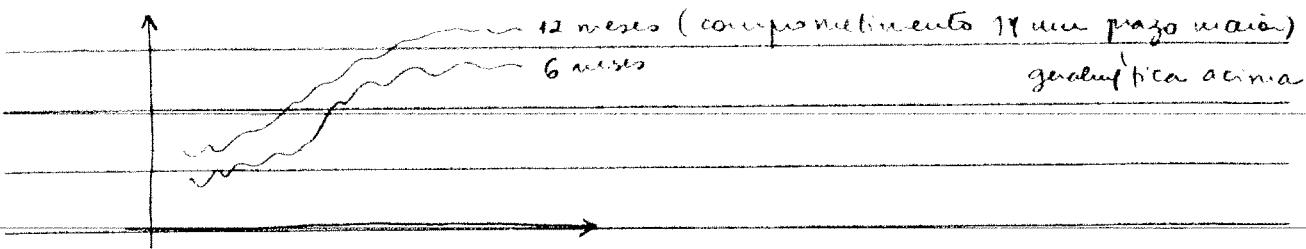
Serve y_t séries multivariadas, nas estacionárias

(No caso de estac. mas se aplica este conceito. Nesse caso las correlacionadas ou não).

Ex. de séries macroeconómicas co-integradas.

- Títulos de dívida pública e rendimentos +

(ambas variam de acordo com estruturas de juros).



Elas são co-integradas pois há E(1) de uma série de taxas comuns.
No processo formador desta estrutura a termos há um fator comum e relevante que afeta as séries.

Ex. Relações entre consumo e renda afogada.

Tb por razões macroeconómicas espera-se que sejam co-integradas.

Ex. Petróleo em várias maturidades

Petróleo e seus derivados.

O fato de terem co-integrações é a existência de um fator comum (ex: componente comum, como tendências por exemplo).

Voltando ao modelo:

$$\left\{ \begin{array}{l} y_{1t} = \mu_t^+ + \epsilon_{1t} \\ y_{2t} = \pi \mu_t^+ + \bar{\mu} + \epsilon_{2t} \end{array} \right.$$

$$\mu_t^+ = \mu_{t-1}^+ + \eta_{t-1}^+$$

$$\Rightarrow \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_t^+ \\ \pi \mu_t^+ + \bar{\mu} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \quad \rightarrow \text{cada } y_{jt} \text{ é um RW} \quad (I(1))$$

Se y_{1t} e y_{2t} sas CI $\Rightarrow \exists \alpha_1 y_{1t} \sim I(0)$ estacionário

$y_{1t} \sim I(1)$ (trivial)

$$\begin{pmatrix} \alpha_1 & \alpha_2 \\ y_{1t} \\ y_{2t} \end{pmatrix} = \alpha_1 y_{1t} + \alpha_2 y_{2t} = \alpha_1 \mu_t^+ + \alpha_2 \pi \mu_t^+ + \alpha_2 \bar{\mu} + (\alpha_1 \varepsilon_{1t} + \alpha_2 \varepsilon_{2t}) \sim I(1) ?$$

sas arbitrárias

Se conseguirmos escolher α_1 e α_2 tal que srie seja $I(0)$

\Rightarrow provamos que sas co-integradas.

Escolhendo $\alpha_1 = -\pi$ } conveniente de forma a anular
 $\alpha_2 = 1$ processo em estacionário μ_t^+

$$\begin{pmatrix} -\pi & 1 \\ y_{1t} \\ y_{2t} \end{pmatrix} = -\pi \mu_t^+ + \pi \mu_t^+ + \bar{\mu} + (-\pi \varepsilon_{1t} + \varepsilon_{2t}) \\ = \bar{\mu} + \varepsilon_t \sim I(0)$$

Sries sas co-integradas

pois CL tem ordem menor

do que cada uma separadamente

\Rightarrow concluas: $\rho = 1$ é condicão suficiente p/ existência
de componente comum
(sries ficam co-integradas)

($\pi = 1$ mas é necessário. Ele torna a diferença entre
tendências ck.)

Generalizações do que vimos acima para variáveis bivariadas. Agora y_t é p-variado

1) Modelos de Nível local Geral $\Rightarrow y_t \sim p \times 1$: quebramos em partes

$$(I) \quad y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon)$$

SUTSE $\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta)$
p-variado
de nível
local.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \xrightarrow{n \times 1} \begin{bmatrix} K \times 1 \\ K \times 1 \end{bmatrix}$$

$$(K+n=p)$$

considere a seguinte partição para $\Sigma_\eta = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \xrightarrow{n \times 1} \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \xrightarrow{n \times 1}$$

$$\begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} \xrightarrow{n \times 1} \begin{bmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} \xrightarrow{n \times 1}$$

onde $n = p - K$

$$\begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} \Rightarrow$$

matriz
de var covar Σ_η
também separam partes
como consequência.

$$\text{Seja } L = \begin{pmatrix} I_K & 0 \\ -\Pi & I_n \end{pmatrix} \Rightarrow \text{transformações matriciais}$$

(propriedades fundamentais
permanecem invariantes)

usamos novamente
esta sua particular

$$\therefore L \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} = L \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{pmatrix} + L \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

resolver e
avaliar a soma
de componentes
comuns

co-integrar

$$\begin{pmatrix} I & 0 \\ -\Pi & I \end{pmatrix} \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} = \begin{pmatrix} I & 0 \\ -\Pi & I \end{pmatrix} \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{pmatrix} + \begin{pmatrix} I & 0 \\ -\Pi & I \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

$$I \mu_{1t} = I \mu_{1,t-1} + I \eta_{1t} \Rightarrow \mu_{1t} = \mu_{1,t-1} + \eta_{1t}$$

$$-\Pi \mu_{1t} + I \mu_{2t} = -\Pi \mu_{1,t-1} + I \mu_{2,t-1} - \Pi \eta_{1t} + \eta_{2t}$$

$$\therefore \underbrace{\mu_{1t} - \Pi \mu_{1t}}_{\bar{\mu}_t} = \underbrace{\mu_{2,t-1} - \Pi \mu_{2,t-1}}_{\bar{\mu}_{t-1}} + (\eta_{2t} - \Pi \eta_{1t}).$$

íllibra

$$\bar{\mu}_t = \bar{\mu}_{t-1} + \bar{\eta}_t$$

E assim segue:

$$\begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} = \begin{pmatrix} \mu_{1t} \\ \bar{\mu}_t \end{pmatrix} \Rightarrow L\mu_t = \mu_t^*$$

mostra-se que $\text{cov} \begin{pmatrix} \eta_{1t} \\ \eta_t \end{pmatrix} = \begin{pmatrix} \Sigma_{\eta_1} & 0 \\ 0 & \bar{\Sigma}_{\eta} \end{pmatrix}$

provar!

$$\text{onde } \bar{\Sigma}_{\eta} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

Por outro lado:

$$y_t = \mu_t + \varepsilon_t$$

$$y_t = L^{-1}L\mu_t + \varepsilon_t$$

$$\therefore y_t = L^{-1}\mu_t^* + \varepsilon_t$$

$$\therefore \begin{pmatrix} y_t \\ \varepsilon_t \end{pmatrix} = \begin{pmatrix} I_K & \Pi \\ 0 & I_n \end{pmatrix} \begin{pmatrix} \mu_{1t} \\ \bar{\mu}_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \Rightarrow E o (II) \text{ separado}$$

$$\begin{pmatrix} \mu_{1t} \\ \bar{\mu}_t \end{pmatrix} = \begin{pmatrix} \mu_{1,t-1} \\ \bar{\mu}_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \bar{\eta}_t \end{pmatrix}, \quad y_t = \begin{pmatrix} \eta_{1t} \\ \bar{\eta}_t \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{\eta_1} & 0 \\ 0 & \bar{\Sigma}_{\eta} \end{pmatrix} \right)$$

$$\bar{\Sigma}_{\eta} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

(30.10.11)

$$\text{P.M.} \quad \text{P.M.} \quad \text{M.X.}$$

$$(I) \quad y_t = z_t x_t + \epsilon_t$$

$$\epsilon_t \sim N(0, H_t)$$

$$(II) \quad \hat{x}_{t+1} = T_t x_t + R_t \eta_t \quad \eta_t \sim N(0, Q_t)$$

M.X. $\approx x_1$

Passo da Previsão: $a_{t+1} = E[x_{t+1} | y_t] = T_t a_{t|t}$ Harvey
 (da aula) Notações
 (nas do

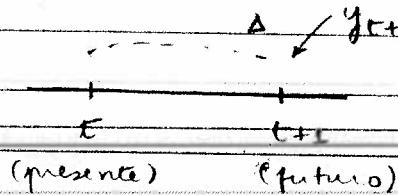
passada) $P_{t+1} = \text{Var}(x_{t+1} | y_t) =$ Koopman
 $= T_t P_{t|t} T_t' + R_t Q_t R_t'$

Queremos $a_{t|t}$ e $P_{t|t}$

$$\downarrow \quad \downarrow$$

$$E[x_t | y_t] \quad \text{Var}(x_t | y_t) \quad (\text{matriz de varcov})$$

A situação na prática é:



O passo $E[x_t | y_t]$ é a revisão em $t+1$ da previsão que havia sido feita

quando chegou y_{t+2}

$E[x_t | y_{t-1}]$: estamos em $t-1$ e queremos previsões $\# t$

$E[x_t | y_t]$: chega info y_t e queremos revisar a previsão passada.

Queremos epis que linkem as 2 funções dist. de prob. dos 2 períodos.

Obs: se fôr oltarmos média e variância pois estamos trabalhando em ambiente Gaussiano. Se não tivermos Gaussianidade, teríamos que recuperar a tábua a fdp a cada instante

base p/ nossas deduções

Resultados (da aula passada):

$$(1) \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim N \begin{pmatrix} \mu_X \\ \mu_Y \\ \mu_Z \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} & \Sigma_{XZ} \\ \Sigma_{YX} & \Sigma_{YY} & \Sigma_{YZ} \\ \Sigma_{ZX} & \Sigma_{ZY} & \Sigma_{ZZ} \end{pmatrix}$$

X, Y, Z : vetores

de dimensões
arbitrária

Supon: (i) $\mu_Z = 0$

(ii) $\Sigma_{ZY} = 0 \Rightarrow \Sigma_{YZ} = 0$

Dado isso, prova-se que:

$$E[X | Y=y, Z=z] = \mu_{X|Y} + \Sigma_{XZ} \Sigma_{ZZ}^{-1} z \quad (a)$$

$$\text{Var}[X | Y=y, Z=z] = \Sigma_{X|Y} - \Sigma_{XZ} \Sigma_{ZZ}^{-1} \Sigma_{ZX} \quad (b)$$

(2) Se X e Y são dois vetores aleatórios e

$p_{X|Y}$ $p_{X|Z}$

$g: \mathbb{R}^p \rightarrow \mathbb{R}^p$ (transf. bijetiva \Rightarrow inversa: g^{-1})

cond. necessária e
suficiente

então $F_{X|Y} = F_{X|g(Y)}$

$$\Rightarrow E[X | Y=y] = E[X | g(Y)=g(y)]$$

Calculando

decompõendo em Y_{t-1} e y_t

$$\Rightarrow E[x_t | Y_t] = E[x_t | Y_{t-1}, y_t] =$$

$$= E[x_t | Y_{t-1}, v_t]$$

onde $v_t \stackrel{\text{def}}{=} y_t - E[y_t | Y_{t-1}]$

tilibra

inovacões (nova de previsões + passo a frente)

$$\bullet \text{var}(v_t) = E[(v_t - E(v_t))(v_t - E(v_t))'] = E[v_t v_t']$$

$$\text{mas } E[v_t v_t'] = E[E[v_t v_t' | Y_{t-1}]]$$

sabemos que $v_t = y_t - z_t \alpha_t$
 $= z_t \alpha_t + \epsilon_t - z_t \alpha_t$
 $= z_t (\alpha_t - \alpha_t) + \epsilon_t$

$$\therefore v_t' = (\alpha_t - \alpha_t)' z_t' + \epsilon_t'$$

$$\Rightarrow E[v_t v_t' | Y_{t-1}] = E[(z_t (\alpha_t - \alpha_t) + \epsilon_t)((\alpha_t - \alpha_t)' z_t' + \epsilon_t') | Y_{t-1}]$$

$$= E[z_t (\alpha_t - \alpha_t)(\alpha_t - \alpha_t)' z_t' | Y_{t-1}] + E[\epsilon_t \epsilon_t' | Y_{t-1}]$$

(termos cruzados se anulam)

$$\text{Ex: } E[z_t (\alpha_t - \alpha_t) \epsilon_t' | Y_{t-1}] =$$

$$= E[z_t \alpha_t \epsilon_t' | Y_{t-1}] - E[z_t \alpha_t \epsilon_t' | Y_{t-1}]$$

$$= z_t E[\alpha_t \epsilon_t' | Y_{t-1}] - z_t \alpha_t E[\epsilon_t' | Y_{t-1}]$$

se iteramos dado passado
de reia função
de α_t , e $\eta_{t-j} \Rightarrow \emptyset$

obj. A cond. inicial $\alpha_0 \sim N(\alpha_0, P_0)$ denotulatadas das choques
faz com que termos cruzados se anulam.

$$\therefore E[v_t v_t' | Y_{t-1}] = z_t P_t z_t' + H_t$$

(valor condicional)

matriz de var/covar estimada no passo anterior

$$\text{logo: } E[v_t v_t'] = E[E[v_t v_t' | Y_{t-1}]] \Rightarrow E[v_t v_t'] = z_t P_t z_t' + H_t \hat{=} F_t$$

Matriz var. covar. da inovações

$$\bullet \text{cov}[\alpha_t, v_t] = M_t$$

Da definição:

$$\begin{aligned} M_t &= E[(\alpha_t - E[\alpha_t])(v_t - E[v_t])^T] \\ &= E[\alpha_t v_t^T] - E[\alpha_t] \cdot E[v_t]^T = \\ &= E[\alpha_t v_t^T] \quad \text{lei das exp. iteradas} \\ &\quad \downarrow \\ &= E[E[\alpha_t v_t^T | y_{t-1}]] \end{aligned}$$

$$\text{Calculemos } \alpha_t v_t^T = \alpha_t [(\alpha_t - a_t)^T z_t^T + \epsilon_t^T]$$

calculado no passo anterior

$$\text{Logo: } E[\alpha_t v_t^T | y_{t-1}] = E[\alpha_t (\alpha_t - a_t)^T z_t^T | y_{t-1}] + E[\alpha_t \epsilon_t^T | y_{t-1}]$$

biteado: $f(\alpha_t)$

$$= E[(\alpha_t - a_t)(\alpha_t - a_t)^T z_t^T | y_{t-1}]$$

$$\text{obs: } (E[\alpha_t \alpha_t^T | y_{t-1}] - E[\alpha_t | y_{t-1}] a_t^T) z_t^T \quad \text{é a m/maco}$$

$$(E[\alpha_t \alpha_t^T | y_{t-1}] - a_t a_t^T) z_t^T$$

$$\therefore E[\alpha_t v_t^T | y_{t-1}] = E[(\alpha_t - a_t)(\alpha_t - a_t^T) | y_{t-1}] z_t^T =$$

$$= P_t z_t^T$$

$$\text{Portanto: } M_t = E[E[\alpha_t v_t^T | y_{t-1}]] = P_t z_t^T$$

Usando os resultados obtidos $\text{var}(v_t)$ e $\text{cov}(\alpha_t, v_t)$, podemos escrever:

$$E[x_t | Y_t] = \alpha_t + M_t F_t^{-1} v_t$$

onde $M_t = P_t Z_t'$

$$F_t = Z_t P_t Z_t' + H_t$$

$$v_t = y_t - Z_t \alpha_t$$

Agora queremos:

$$\text{var}[x_t | Y_t] = ?$$

Tínhamos feito $x = x_t, y = y_{t+1}, z = v_t$

⇒ Usando a fórmula (b):

$$\text{var}[x_t | y_t] = \text{var}[x_t | y_{t-1}, v_t]$$

$$= \text{var}[x_t | Y_{t-1}] - \text{cov}(x_t, v_t) \text{var}(v_t) \text{cov}(x_t, v_t)$$

$$\therefore P_{t|t} = P_t - M_t F_t^{-1} M_t'$$

$$= P_t - M_t F_t^{-1} (P_t Z_t')' =$$

$$= P_t - P_t Z_t' F_t^{-1} Z_t P_t$$

$$\Rightarrow P_{t|t} = P_t - P_t Z_t' F_t^{-1} Z_t P_t$$

e simétrica

Resumindo:

Equações do FK

Previsões: $\hat{\alpha}_{t+1} = T_t \hat{\alpha}_{t|t}$

$$P_{t+1} = T_t P_{t|t} T_t' + R_t Q_t R_t'$$

onde $v_t = y_t - Z_t \alpha_t$

Atualizações: $\hat{\alpha}_{t+1|t} = \hat{\alpha}_t + M_t F_t^{-1} v_t \xrightarrow{(y_t - Z_t \alpha_t)} F_t = Z_t P_t Z_t' + H_t$

$$P_{t+1|t} = P_t - P_t Z_t' F_t^{-1} Z_t P_t \quad . . . \quad M_t = P_t Z_t'$$

\downarrow atualizada \downarrow prevista \downarrow conceitos

equilíbrio

na prática: temos uma previsão sequencial

$x_t \sim N(a_t, P_t)$: dist. a priori inicial \rightarrow é o passo de previsão inicial.
(conhecido)

Começamos na atualização usando y_t (1º obs).

Com $a_{t|t}$ e $P_{t|t}$, jogamos nas fórmulas da previsão.

Se rescrevermos as fórmulas em funções apenas de a_t e a_{t+1} (P_t e P_{t+1}), rest. ativo \rightarrow $P_{t|t}$ das eqs. de atualizações nas eqs. de previsão, temos:

$$F_t \text{ é em } 1 \Rightarrow a_{t+1} = T_t(a_t + M_t F_t^{-1} v_t)$$

(passo da
atualização
desaparece)

$$\therefore a_{t+1} = T_t a_t + T_t M_t F_t^{-1} v_t \quad \text{onde } K_t: \text{ ganho de Kalman}$$

$$\therefore a_{t+1} = T_t a_t + K_t(v_t)$$

observações entram aqui

fazendo pf: $P_{t+1} = T_t P_t L_t^{-1} + R_t Q_t R_t^{-1}$ | onde $L_t = T_t - K_t Z_t$

Não depende das observações

Dependendo do caso, converge H um valor. Se for assim, podemos trabalhar com \bar{P} (steady state value) e paramos de computar.

(11)

Exemplo:

$$\begin{cases} y_t = \mu_t + \epsilon_t & (m=p=1) \\ \mu_{t+1} = \mu_t + \eta_t \end{cases}$$

\Leftrightarrow : $E[y_t | Y_{t-1}] = E[\mu_t + \epsilon_t | Y_{t-1}] = \alpha_t$ (*)

$$\begin{cases} Z = T = R = 1 \text{ (escalares)} \\ H = \sigma_e^2 \\ Q = \sigma_p^2 \end{cases}$$

$$\bullet K_t = T_t M_t F_t^{-1} = P_t F_t^{-1} = \frac{P_t}{F_t} \quad \begin{matrix} \hookrightarrow P_t Z_t \\ F_t - \text{escalar} \end{matrix}$$

$$\therefore K_t = \frac{P_t}{P_t + \sigma_e^2} \rightarrow 0 < K_t < 1$$

$$\bullet v_t = y_t - Z_t \alpha_t = y_t - \alpha_t$$

$$\begin{aligned} a_{t+1} &= a_t + K_t v_t \\ &= a_t + K_t(y_t - a_t) \\ &= a_t + K_t y_t - K_t a_t \end{aligned}$$

$$\therefore a_{t+1} = (1 - K_t)(a_t) + K_t y_t \quad \rightarrow E[y_t | Y_{t-1}] \quad (*)$$

	$\hat{y}_{t+1 t} = K_t y_t + (1 - K_t) \hat{y}_{t t-1} \rightarrow \text{EWMA}$
--	---

FK está fazendo ajustamento exponencial

No caso clássico:

$$\hat{y}_{t+1|t} = \lambda y_t + (1 - \lambda) \hat{y}_{t|t-1} \quad 0 < \lambda < 1$$

$$\text{fazendo } w_j = \lambda(1-\lambda)^j \Rightarrow \sum_{j=0}^{t-1} w_j y_{t-j} \text{ média ponderada}$$

vibrar
não é um caso

Quando adotamos modelos de nível local, a função de previsões é um EWMA

Na hora que estimamos σ^2_ϵ temos um σ^2_ϵ ótimo.

Os pesos do EWMA estarão otimizados.

Estes métodos garantem a otimização das estimativas dos parâmetros.

Obs: Também $y_t = \mu_t + \gamma_t + \varepsilon_t$

$$\mu_{t+1} = \mu_t + \beta_t + \gamma_t$$

$$\beta_{t+1} = \beta_t + \xi_t$$

$$\gamma_t = \sum_{j=1}^{12} Y_{t-j} + w_t$$

Modelo

6

de previsões

sazonalidade

usando fórmula

de Holt-Winters

Como estamos trabalhando com modelos lineares,

todos os métodos acabarão trabalhando com médias das observações passadas.

Os pesos estarão sendo objetivamente calculados de forma ótima (parâmetros reais estimados por MV).

Sob Normalidade:

usamos média central como estimador de μ
 \rightarrow (= moda = mediana)

Usualmente fazemos isso, pois a média tem:
uma propriedade ótima \Rightarrow minimiza o erro quadrático médio

(1/1)

$\hat{x}_{t+1|t}^*$ \Rightarrow estimativa

$$\text{Queremos } \min_{\hat{x}_{t+1|t}^*} E[(\hat{x}_{t+1} - \hat{x}_{t+1|t}^*)(\hat{x}_{t+1} - \hat{x}_{t+1|t}^*)^T | Y_t]$$

$\rightarrow \text{MSE}(\hat{x}_{t+1|t}^*)$

$$\Rightarrow \hat{x}_{t+1|t}^* = E[\hat{x}_{t+1} | Y_t] \\ = a_{t+1}$$

O estimador que minimiza o MSE é a média condicional (sob norma lida)

Nas gaussianas

(Se usarmos as fórmulas obtidas nessa situação,
o que são os resultados que obtivemos?)

a_{t+1} : melhor estimador linear de x_{t+1}

$$\hat{x}_{t+1|t}^* = \sum_{j=0}^{t-1} w_j y_{t-j}$$

queremos que este estimador linear seja ótimo no sentido de min MSE

Se fizermos isso, obtemos o equivalente ao FK que calculamos.

Nas reais a média, mas terá uma característica de optimidade local.

(sob Gaussianidade, temos o melhor estimador entre os lineares e não lineares, como vimos)

Na prática: (para implementar o FK)

$$\sigma_e^2 = \langle \cdot \rangle \quad a_i = \langle a_i \rangle \quad \text{valores dados}$$

$$\sigma_\eta^2 = \langle \cdot \rangle \quad p_i = \langle p_i^* \rangle$$

$$\text{for } t = 1, n \quad \Rightarrow \quad a_{t+1} = a_t + k_t v_t$$

$$v_t = y_t - a_t$$

$$p_{t+1} = p_t (1 - k_t) + \sigma_\eta^2$$

$$k_t = \frac{p_t}{p_t + \sigma_e^2}$$

continue (loop até 'n')

$$F_t = p_t + \sigma_e^2$$

Temos uma ST qualquer e queremos modelá-la

Começa q a_1 e p_1^*

Calcula v_1, k_1

Calcula $a_2, p_2 \quad \text{loop } n \times (\text{teremos } n \text{ valores})$

Mas precisamos saber com que valores entramos p_1

parâmetros fixos

Temos uma faixa de variação

$$l = l - \frac{1}{2} (\ln f_t + v_t^2 / f_t)$$

(começa q $l=0$ e vai acumulando em l)

pp ser ótimo, devemos ter l máximo.

Temos que ter um algoritmo que varia

os parâmetros fixos pp otimizar $l \Rightarrow$ gera novos

valores.
No caso $\sigma_e^2 < \sigma_\eta^2$

Ex: $\Psi^{(i+1)} = \Psi^{(i)} - \lambda H^{-1}(\Psi^{(i)}) g(\Psi^{(i)})$. forma iterativa
bifida

H^{-1} Heriano gradiente λ itera
 2×2 \rightarrow itera para gerar novos valores

(11)

Iterações (i) $\ell^{(i)}$ $0 \quad \ell^{(0)}$ $1 \quad \ell^{(1)}$ $2 \quad \vdots$

I

\Rightarrow a cada iteração temos que
cheiar as condições de otimalidade
(gradiente = 0 e Hessiana)

difícil: implementar um algoritmo que converja n
valores ótimos dos parâmetros.

de otimização

Talvez se problema tiver restrições:

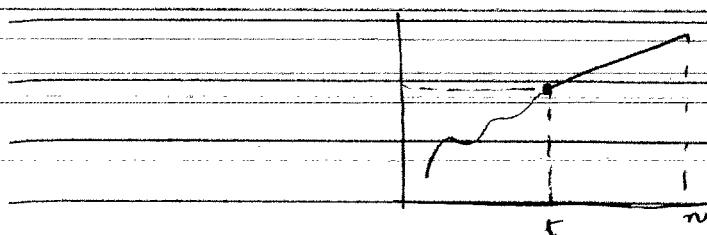
- se σ_e^2 sempre $> 0 \Rightarrow$ podemos fazer $\sigma_e^2 = e^\theta$ e otimizar por θ .

— II —

Estimativa suavizada ou alisada do estado
(smoothing)

$$f(x_t | y_n) \sim N(\hat{x}_t | V_t) \quad \rightarrow \text{notação de Koopman}$$

\downarrow
 $f(x_t | \text{dados o conjunto de observações completo})$



Notações: $y = y_n$ (do Koopman)
 $= (y_1, y_2, y_3, \dots)$

$$\hat{x}_t = E[\hat{x}_t | y] = E[x_t | y_{t-1}, y_t, y_{t+1}, \dots, y_n] \quad \left\{ \begin{array}{l} \text{como } \hat{x}_t \text{ é a} \\ \text{média suavizada do estado} \end{array} \right. \\ = E[x_t | y_{t-1}, v_t, v_{t+1}, \dots, v_n] \quad \left\{ \begin{array}{l} \text{com } v_t \text{ os} \\ \text{erros de} \\ \text{previstas} \\ \text{atualizações} \end{array} \right.$$

Temos o resultado:

$$E[x | y = y, z = z] = E[x | y] + \sum_{xz} \sum_{zz} z$$

Faremos: $x = \hat{x}_t$ erros de previstas um passo a frente
 $y = y_{t-1}$
 $z = (v_t, v_{t+1}, \dots, v_n) \rightarrow \text{dimensão: } (n-t+1) \times 1$

$$\therefore E[\hat{x}_t | y] = E[\hat{x}_t | y_{t-1}] + \text{cov}(\hat{x}_t, (v_t, v_{t+1}, \dots, v_n)') \cdot \text{var}(v_t, \dots, v_n)$$

$\cdot \begin{pmatrix} v_t \\ v_{t+1} \\ \vdots \\ v_n \end{pmatrix}$

(Ainda no passo de atualização)

(06/09/11)

Equações de Riccati (steady state KF)Observar a equação para var(α_{t+1}, γ_t) = P_{t+1}

$$P_{t+1} = T P_t L_t' + R Q R' \rightarrow F_K \text{ é un. } \pm \text{ (embutido nos passos de previsão)}$$

Considerar sistemas invariantes no tempo (matrizes do sistema independem de t)

$$P_{t+1} = T P_t L_t' + R Q R'$$

$$L_t = T - K_t Z = T - (T P_t Z' F_t^{-1}) Z \quad (K \text{ e } K_t \text{ não depende de } P_t)$$

$$F_t = Z P_t Z' + H$$

desta forma:

$$P_{t+1} = T P_t [I - (T P_t Z' F_t) Z]^{-1} + R Q R'$$

$$\therefore P_{t+1} = T P_t T' - T P_t Z' F_t^{-1} P_t T' + R Q R' \quad \begin{array}{l} \text{chamada} \\ \text{equações de} \\ \text{Riccati} \end{array}$$

Suponha que lim $P_{t+1} = P_t = \bar{P}$ - computacionalmente nas etapas de recálculo (mais eficiente)

$$\bar{P} = T \bar{P} T' - T \bar{P} Z' F_t^{-1} \bar{P} T' + R Q R' \quad \text{e} \quad F_t = Z \bar{P} Z' + H$$

resolver para \bar{P} equação matricial $H \bar{P}$

$$\text{Exemplo: } y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_e^2)$$

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

obs. Existem algumas cond. pf ep. de Riccati ter soluções. No nosso caso só nos preocupamos com a checagem

$$\bar{T} = \bar{z} = 1$$

$$\therefore \bar{P} = \bar{P} - \bar{P} \bar{F}' \bar{P} + \sigma_\eta^2$$

$$= \bar{P} - \frac{\bar{P}^2 + \sigma_\eta^2}{\bar{F}} \Rightarrow \bar{P} = \bar{P} \left(1 - \frac{\bar{P}}{\bar{P} + \sigma_\epsilon^2} \right) + \sigma_\eta^2$$

$$\Rightarrow x^2 - xh - h = 0 \quad \text{onde } x = \bar{P}/\sigma_\epsilon^2$$

$$h = \sigma_\eta^2/\sigma_\epsilon^2 = q \quad (\text{signal to noise ratio})$$

soluções: $x = \frac{h \pm \sqrt{h^2 + 4h}}{2}$, desde que $h > 0$ ou

$\rightarrow \hat{x}_t$ depende do tempo

obs: $y_t = y_{t-1} \varphi_t + \varepsilon_t$ Jamais haria

$$\varphi_t = \Phi \varphi_{t-1} + \eta_t \quad \text{soluções de Riccati}$$

(ver lista)

obs 2: computacionalmente

A cada iteração, verificaremos se diferença entre P_t e P_{t+1} é menor que um certo valor (definição de igualdade de matriz).

Se for assim η os n primeiros casos \Rightarrow abandona cálculo de P_{t+1}
(Ao invés de calcular se há solução da eq. de Riccati)

— " —

Passo de suavizações (ver folha)

(11)

O que vimos até agora:

$$\hat{\alpha}_j = E[\alpha_t | y_j]$$

$$\left\{ \begin{array}{l} \alpha_t \quad \text{se } j = t-1 \\ \hat{\alpha}_{t|t} \quad \text{se } j = t \end{array} \right.$$

Agora, vamos expandir o conjunto de observações Π o qual tomamos a esperança.

Faremos $j = n$

(trá' passado, $t-1$, presente, t , e futuro, ate' n)

Existem vários tipos de smoothing.

Veemos, um desenvolvido por Koopman.

(obs: Na literatura de Eng, os livros que deram origem à Teoria de Espaço de Estado: Anderson & Moore; Jagwinkar.

def: Π significa que y_t é independente

temos que other Π dist. conjunta.

Como é serie de tempo

$$p(y_1, \dots, y_n) = \prod_{t=1}^n p(y_t | y_{t-1}), \quad p(y_1 | y_0) = p(y_1)$$

$$\text{Mas } y_t = z_t \alpha_t + \epsilon_t$$

$$p(y_t | y_{t-1}) = \int_{-\infty}^{\infty} p(y_t, \alpha_t | y_{t-1}) d\alpha_t =$$

$$= \int_{-\infty}^{\alpha} p(y_t | \alpha_t) \cdot p(\alpha_t | y_{t-1}) d\alpha_t$$

Normal

\hat{y}_{t-1}

$$E[y_t | Y_{t-1}] = \hat{y}_{t-1}$$

$$F_t = \text{var}(y_t | Y_{t-1}) = \hat{\sigma}_t^2 P_t \hat{\sigma}_t + H_t$$

Se pudemos provar que prob. conjunta das inovações = produto das marginais

\Rightarrow inovações são independentes

— II —

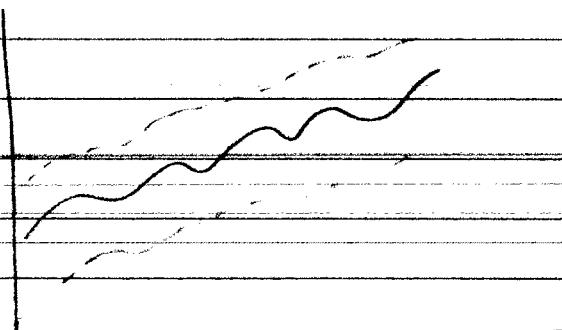
Intervalos de confiança à tendência

$$\text{Ex: } y_t = \mu_t + \gamma_t + \epsilon_t$$

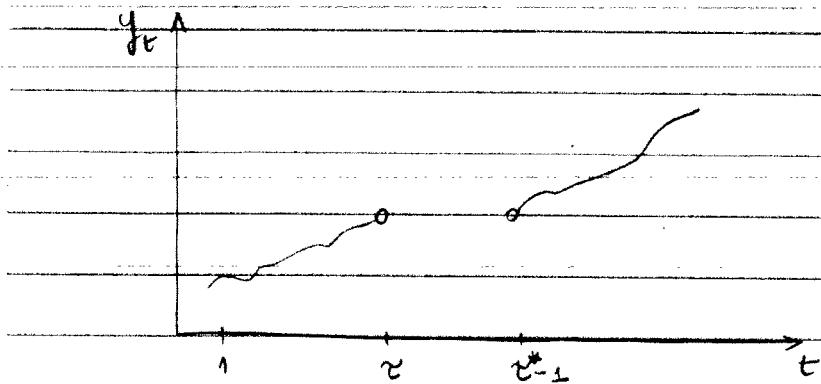
$$x_t = \begin{pmatrix} \mu_t \\ \gamma_t \end{pmatrix} \begin{array}{l} \text{tendência} \\ \text{sazonalidade} \end{array} = \begin{matrix} \alpha_{1,t} \\ \alpha_{2,t} \end{matrix}$$

$$\hat{\alpha}_t \pm t_{\alpha/2} \sqrt{V_{tt}}$$

da variação suavizada



Observações faltantes em MEE



Observações faltantes em $t = \tau, \tau+1, \dots, \tau^*-1$



$(\tau^* - \tau)$ observações faltantes

FK

$$\hat{a}_{t+1} = T_t \hat{a}_t + K_t v_t$$

$$P_{t+1} = T_t P_t T_t' + R_t Q_t R_t'$$

$$L_t = T_t - K_t Z_t$$

Nos inst. em que tiverem obs faltantes faremos $v_t = 0$ (mas variar os t. pf suas inovações correspondentes).

E como n̄ variar gasto (pois n̄ há inovações) $\Rightarrow K_t = 0$.

$$\text{Em } t = \tau, \tau+1, \dots, \tau^*-1 \Rightarrow v_t = 0 \quad \left. \begin{array}{l} \\ \end{array} \right\} L_t = T_t \\ K_t = 0$$

$$\therefore \hat{a}_{t+1} = T_t \hat{a}_t$$

$$P_{t+1} = T_t P_t T_t' + R_t Q_t R_t'$$

(/ /)

último t com observações é $t = T - 1$

Supondo sistema invariante no tempo (pf contas + simples)

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

$$P_{t+1} = TP_tT' + RQR'$$

$$t+1 = T \Rightarrow \alpha_T = T\alpha_{T-1}$$

$$\alpha_{T+1} = T\alpha_T = T^2\alpha_{T-1}$$

:

$$\alpha_{T+j} = T^j\alpha_{T-1}$$

Da equações do estado:

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

Iterando até $t+j$:

$$\alpha_{t+j} = T^j\alpha_t + (\dots)$$

$$\text{Previsões: } E[\alpha_{t+j} | y_t] = T^j E[\alpha_t | y_t] = T^j \alpha_{T-1}$$

Nos períodos de previsão, o FK projeta j passos a frente.

Pega-se o último vetor de estados e projeta a média

ao longo do período de dados faltantes

Para a variância valerá a mesma coisa.

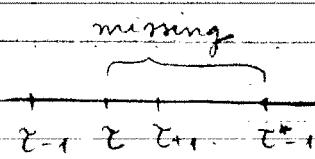
$$P_{t+j} = TP_{t-1}T' + RQR'$$

$$= T(TP_tT' + RQR')T' + RQR =$$

$$= T^2P_t(T')^2 + TRQRT' + RQR$$

$$\text{Ex: } y_t = \alpha_t + \epsilon_t$$

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



$$T = 2 = 1$$

$$\hat{x}_t = x_{\tau} + \sum_{j=\tau}^{t-1} \eta_j, \quad t = \tau, \tau+1, \dots, \tau^*-1 \quad (\star)$$

$$E[\alpha_t | y_{t-1}] = E[\alpha_t | y_{\tau-1}]$$

= α_{τ} ↑ termos obs ate' $\tau-1$

→ ultima previsão ante de faltar dados

para os períodos de obs faltantes:

se α_{τ} ainda $\hat{\alpha}_{t+1} = \alpha_{\tau}, \tau, \dots, \tau^*-1$

podemos $\hat{\alpha}_{t+1} = \alpha_{\tau}$

fazer previsões $\hat{\alpha}_{t+1} = \alpha_{\tau}$

pois $\alpha_{\tau} = E[\alpha_t | y_{t-1}]$ $\hat{\alpha}_{t+2} = \hat{\alpha}_{t+1} = \alpha_{\tau}$

⋮

$$(\star) \text{ var}(\hat{\alpha}_t | y_{t-1}) = \text{var}(\alpha_t | y_{t-1}) + \text{var}\left(\sum_{j=\tau}^{t-1} \eta_j\right)$$

$$= P_{\tau} + (t-\tau) \sigma_{\eta}^2$$

↓

variância tb é projetada

Nos algoritmos de smoothing:

$$\text{Equações originais: } \hat{x}_t = \alpha_t + P_t \gamma_{t-1} \quad e \quad V_t = P_t - P_t N_{t-1} P_t$$

$$\gamma_{t-1} = Z_t^T F_t^{-1} Y_t + L_t \gamma_t \quad N_{t-1} = Z_t^T F_t^{-1} Z_t + L_t^T N_t L_t$$

$$\gamma_n = 0$$

$$N_n = 0$$

Entonces $t = 2, 3, \dots, T^* - 1 \Rightarrow k_t = 0 \Rightarrow l_t = T_t$

$$v_t = 0$$

$$\therefore \begin{cases} n_{t+1} = l_t' n_t = T_t' n_t \\ N_{t+1} = T_t' N_t L_t \end{cases}$$

Example: $y_t = \alpha_t + \epsilon_t$

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

• $\hat{\alpha}_t = \alpha_t + p_t n_{t-1}$ (no perdido de obs. faltantes)

• $\hat{\alpha}_{T+j} = \alpha_T + p_{T+j} + n_{T+j-1}, j = 1, 2, \dots, T^*-1$

$$= \alpha_T + (p_T + j\sigma_\eta^2) n_T$$



do resultado anterior

usando o que de dados

faltantes

$$= (\alpha_T + (p_T n_T)) + j(\sigma_\eta^2 n_T)$$

a.

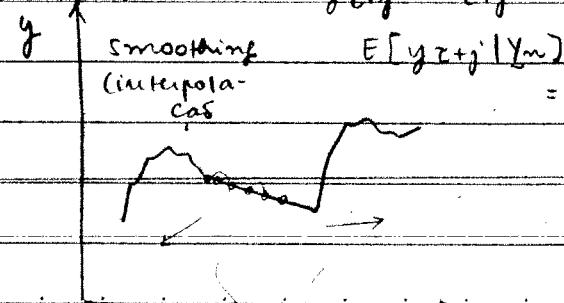
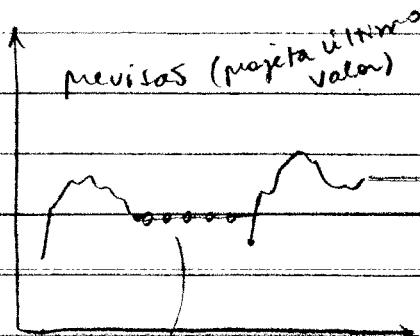
b

$$\therefore \hat{\alpha}_{T+j} = a + bj$$

$$y_t = \alpha_t + \epsilon_t$$

$$y_{T+j} = \alpha_{T+j} + \epsilon_{T+j}$$

$$\text{smoothing (interpolacão)} \quad E[y_{T+j} | Y_m] = E[\hat{\alpha}_{T+j}] = \alpha_{T+j}$$



é característica
deste modelo.
Afinal da amostra temos de ter
a final da amostra temos de ter
a final da amostra temos de ter

a e b com dados partados e tudo

Inicializações do EK

pois tecnicamente não significa

mas no EK precisamos olhar atentamente.

¶ processos estacionários, cond. inicial perde importância com $t \rightarrow \infty$.

¶ processos n'estacionários \Rightarrow isso não é verdade.

Veamos

✓ mas maneras de fazer inicializações

$$\begin{cases} y_t = Z_t x_t + \epsilon_t \\ x_{t+1} = T_t x_t + R_t \eta_t \end{cases}$$

$$t=1 : x_2 = T_1(\underline{x}) + R_1 \eta_1$$



$N(\underline{x}, P_1) \rightarrow$ aqui apoia, tratamos como conhecidos



vetor px, matriz var, covar pxp

O que determina como tratar cond. inicial é a estacionariedade ou não de cada componente do vetor de estado.

¶ comp. estacionária \Rightarrow usa como cond. a dist. de LP do processo

y comp. nas estacionárias \rightarrow dist. a priori difusa

$$y_t = \mu_t + \varepsilon_t$$

$$\mu_{t+1} = \mu_t + \eta_t$$

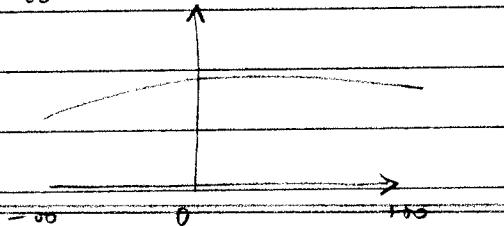
$$\alpha_1 = \mu_1 \sim N(\bar{\alpha}_1, P_1)$$

(

já que não sabemos nada sobre α_1 , não devemos priorizar nenhum valor sobre outro.

Seria como tomar $\alpha_1 \sim p$ dist. uniforme

Equivale a fazer $P_1 = K \rightarrow \infty$



obs: Mas do ponto de vista estatístico é desapropriado

pois este é uma dist. imprópria, ou seja:

$$\text{diz-se } \int_{-\infty}^{\infty} f(\alpha_1) d\alpha_1 = \infty.$$

Prior difusa: é imprópria pois tem var infinita

e dize $\int f(\alpha_1) d\alpha_1$ não converge.

Na prática, não há problema pois só começaremos a trabalhar os dados quando o processo tiver prior própria (deixar de perder em a priori difusa).

obs: A apresentação da forma que vemos é particular
do livro do Koopman.

(Prova!)

• Fazer cálculos / manipulações pp models + geral

$$y_t = \mu_t + \varepsilon_t$$

$$\alpha_t \sim N(0, K I_d)$$

$$\mu_{t+1} = \mu_t + \beta_t + \eta_t$$

$$\beta_{t+1} = \beta_t + \varsigma_t$$

tabalha com
matriz var/covar
diagonais

$K \rightarrow \infty$

Neste processo, pode-se mostrar que a partir de $t=3$
têmos priors bem definidas (môpuas) e assim o
cômputo das quocês do FK não apresentaS problema

Cf d componentes n estacionário \Rightarrow a partir de $t+1$ podemos
computar FK de forma
adequada

pg(8) : Algumas contas:

Tendência linear local

$$y_t = (+ 0) \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \Rightarrow \begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_{t+1} = \mu_t + \eta_t \\ \beta_{t+1} = \beta_t + \varsigma_t \end{cases}$$

$$\begin{pmatrix} \mu_{t+1} \\ \beta_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_t \\ \varsigma_t \end{pmatrix}$$

$$I = (1 \ 0) \quad Q_t = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\varsigma^2 \end{pmatrix} = \sigma_q^2 \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix}$$

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

$$R_t = I_{2x2}$$

$$H_t = \sigma_q^2$$

$$\text{onde } \frac{\sigma^2}{q_1} = q_1, \quad \frac{\sigma^2}{q_2} = q_2$$

Razões nível ruído. Utiliza

Mais fácil se estimar \hat{q}_y , \hat{q}_y por mV
(reparametrizando o sistema e com eles calcula $\hat{\sigma}_\epsilon^2$)

sistema pode deixar

um parâmetro a

menos se estima (se

$$\hat{\sigma}_\epsilon^2 = \frac{1}{T-2} \sum_{t=3}^T v_t^2 / f_t$$

$$\text{onde } v_t = f(\hat{q}_y, \hat{q}_z)$$

$$f_t = g(\hat{q}_y, \hat{q}_z)$$

A cond. inicial é dada por:

$$\alpha_1 = a + A\delta + R_0 \eta_0$$

$$P_1 = K P_\infty + P_\star$$

$$\text{onde } P_\infty = AA'$$

$R_0 = 0$ pois não há componentes estacionárias.

$$\Rightarrow \alpha_1 = \begin{pmatrix} \mu_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$$

ligado a μ

onde $\begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, K I_2 \right]$

$$\therefore P_0 = AA' = I_2 I_2' = I_2 = P_\infty, \quad \text{ligado a } \beta$$

$$P_\star = 0 \quad = P_{\star,1}$$

↓

tb reas iteradas

$$\Rightarrow P_t = K P_\infty = K I_2$$

primeira estimativa:

$$P_t \text{ ou } F_k \Rightarrow \text{usamos } \alpha_{t+1} \in P_{t+1}$$

Enquanto estimamos na fase de inicialização exata, usaremos notações diferentes: $\hat{\alpha}_2^{(0)}$

Quando atingirmos $P_{\infty,t} = 0 \Rightarrow t = d$: aí começamos a considerar F_k padroas.

$$\hat{x}_2^{(0)} = \underbrace{T\hat{\alpha}_1^{(0)}}_{\begin{pmatrix} 0 \\ 0 \end{pmatrix}} + K_1^{(0)} v_1^{(0)} - K_1^{(0)} v_2^{(0)}$$

fazer contas como na folha

— 11 —

Estimações de hiperparâmetros

Além das componentes que são estimadas pelo FK

Há alguns hiperparâmetros do modelo que precisam ser estimados.

São constantes fixas desconhecidas nas matrizes $\{Z_t, T_t, H_t, Q_t\}$

Normalmente R nas tem

(pois é matriz de seleção)

Em modelos estruturais, normalmente em que $\Phi \in H$ (podem ser definidos em diagonais)

\Downarrow
mais comum

$\alpha_t \rightarrow FK$

$$\Phi = \{ \text{hiperparâmetros} \} \Rightarrow Z_t, T_t, H_t, Q_t$$

\Downarrow
usaremos MV

Quando dist. iid \Rightarrow prova-x que estimadores MV

1. consistentes

2. assintoticamente ii viadas

3. eficientes

4. dist. na mal assintótica

Gilbert

iid \Rightarrow 1. consistência: $\text{plim } \hat{\theta} = \theta$

2. asymptoticamente nas vizinhadas: $\lim_{n \rightarrow \infty} E[\hat{\theta}_n] \Rightarrow \theta$

3. eficientes: $\text{var}(\hat{\theta})$ é mínima

4. dist. normal asymptótica $\Rightarrow \hat{\theta} \sim N(\theta, \text{var}(\hat{\theta}))$

dados dependentes

quando não são iid, prova é mais complicada, mas tem as mesmas propriedades (desde que, não estejam na fronteira do espaço paramétrico)

\Rightarrow Para isso usamos Verossimilitude

obs: se p/ pequenas amostras, propriedades asymptóticas não são muito boas \Rightarrow usa bootstrap

Verossimilitude: \rightarrow parte de uma densidade conjunta, mas

\rightarrow o argumento da função

$$L(\psi) = p(y_1, y_2, \dots, y_n; \psi)$$

\rightarrow diferente (faz para metas). Nesse caso,

$$= \prod_{i=1}^n p(y_i | y_{i-1}; \psi)$$

faz para metas mas precisa ser = 1

densidade

preditiva

(se dados dependentes)

duas formas de obter $p(y_t | y_{t-1})$:

(11)

No, modelos curto

$$1) p(y_t | \underline{y_{t-1}}) = \int p(y_t, x_t | \underline{y_{t-1}}) dx_t =$$

$$= \int p(y_t | x_t) p(x_t | \underline{y_{t-1}}) dx_t \quad \begin{matrix} \text{resultado ge ale} \\ \text{dependente} \end{matrix}$$

Mas, $y_t = z_t x_t + \varepsilon_t$, $\varepsilon_t \sim N(0, H_t)$

$x_{t+1} = T_t x_t + R_t \eta_t$, $\eta_t \sim N(0, Q_t)$

de estarmos em
ambiente Gaussiano

$$\Rightarrow p(y_t | x_t) \sim N(z_t x_t, H_t)$$

$$p(x_t | \underline{y_{t-1}}) \sim N(a_t, P_t)$$

Logo: $N(z_t x_t, H_t) \cdot N(a_t, P_t) \Rightarrow$ convoluções da normal.

f

Trabalhando esta expressão, achamos

$$E[\cdot] = V[\cdot]$$

Mas podemos olhar diretamente as egs:

$$E[y_t | \underline{y_{t-1}}] = z_t a_t$$

$$\text{var}[y_t | \underline{y_{t-1}}] = \cancel{z_t P_t z_t^T} + R_t \cancel{Q_t R_t^T} = F_t$$

$$H_t$$

$$\Rightarrow p(y_t | \underline{y_{t-1}}) \sim N(z_t a_t, F_t)$$

$$\therefore p(y_t | \underline{y_{t-1}}) = \frac{1}{(2\pi)^{p/2}} \frac{1}{|F_t|^{1/2}} \exp \left\{ -\frac{1}{2} \underbrace{(y_t - z_t a_t)^T}_{V_t} \underbrace{F_t^{-1} (y_t - z_t a_t)}_{V_t} \right\}$$

p-variaada

$$= (2\pi)^{-p/2} |F_t|^{-1/2} e^{-\frac{1}{2} V_t^T F_t^{-1} V_t}$$

ilibra

$$\therefore L(\Psi) = p(y_1, y_2, \dots, y_n, \Psi) = \prod_{t=1}^n p(y_t | y_{t-1}, \Psi) =$$

$$= \prod_{t=1}^n (2\pi)^{-p_{12}} |F_t|^{-1/2} e^{-\frac{1}{2} v_t' F_t^{-1} v_t}$$

$$\log L(\Psi) = l(\Psi) = -\frac{pn}{2} \log(2\pi) - \frac{1}{2} \sum_t (\log |F_t| + v_t' F_t^{-1} v_t)$$

, entra p (do FK)

6

estimar Ψ depende do FK

e FK depende de Ψ

cond. 1ª ordem (necessária)

$$\frac{\partial l}{\partial \Psi} = 0 \quad \therefore \left(\frac{\partial l}{\partial \Psi_1}, \frac{\partial l}{\partial \Psi_2}, \frac{\partial l}{\partial \Psi_3}, \dots, \frac{\partial l}{\partial \Psi_p} \right)' = 0$$

vetor nulo

ajundo Ψ com dimensão l

cond. 2ª ordem (necessária)

$$H = \begin{pmatrix} \frac{\partial^2 l}{\partial \Psi \partial \Psi'} \end{pmatrix}_{l \times l} \quad \text{é} \quad \begin{array}{l} \text{tiva definida} \\ \text{(negativa)} \end{array} \Rightarrow \exists z' H z < 0$$

$$H z \neq 0.$$

matriz Hessiana

. Nas vezes conseguimos enver Ψ como solução analítica de y_t 's.
(teremos que usar métodos não-lineares de otimização)

. Se l for cônica \Rightarrow resolvido problema encontraremos

maximum global

Mas l não é cônica

Quando calcularmos

Na funções $\ell(\Psi)$: $F_t = F_t(\Psi)$

$$V_t = V_t(\Psi)$$

Inicializações. Duas formas \rightarrow difusa por big kappa
 Exata

(i) . Difusa por big kappa

Computar verossimilhança:

for $t = 1, \dots, n$ \leftarrow pega cond. inicial a_1, P_1

$$a_{t+1}$$

$$P_{t+1} =$$

$$V_t =$$

$$F_t =$$

$$d = \ell - \frac{1}{2} \log |F_t| - \frac{1}{2} V_t^T F_t^{-1} V_t$$

No modelo linear

$$\text{local, } p \propto: a_1(0)$$

$$P_t = K I_{2n, K}$$

continua

$$y_t = z_t a_t$$

$$z_t P_t z_t^T + H_t$$

esperanto

estas infos não forem

bem definidas, não podemos

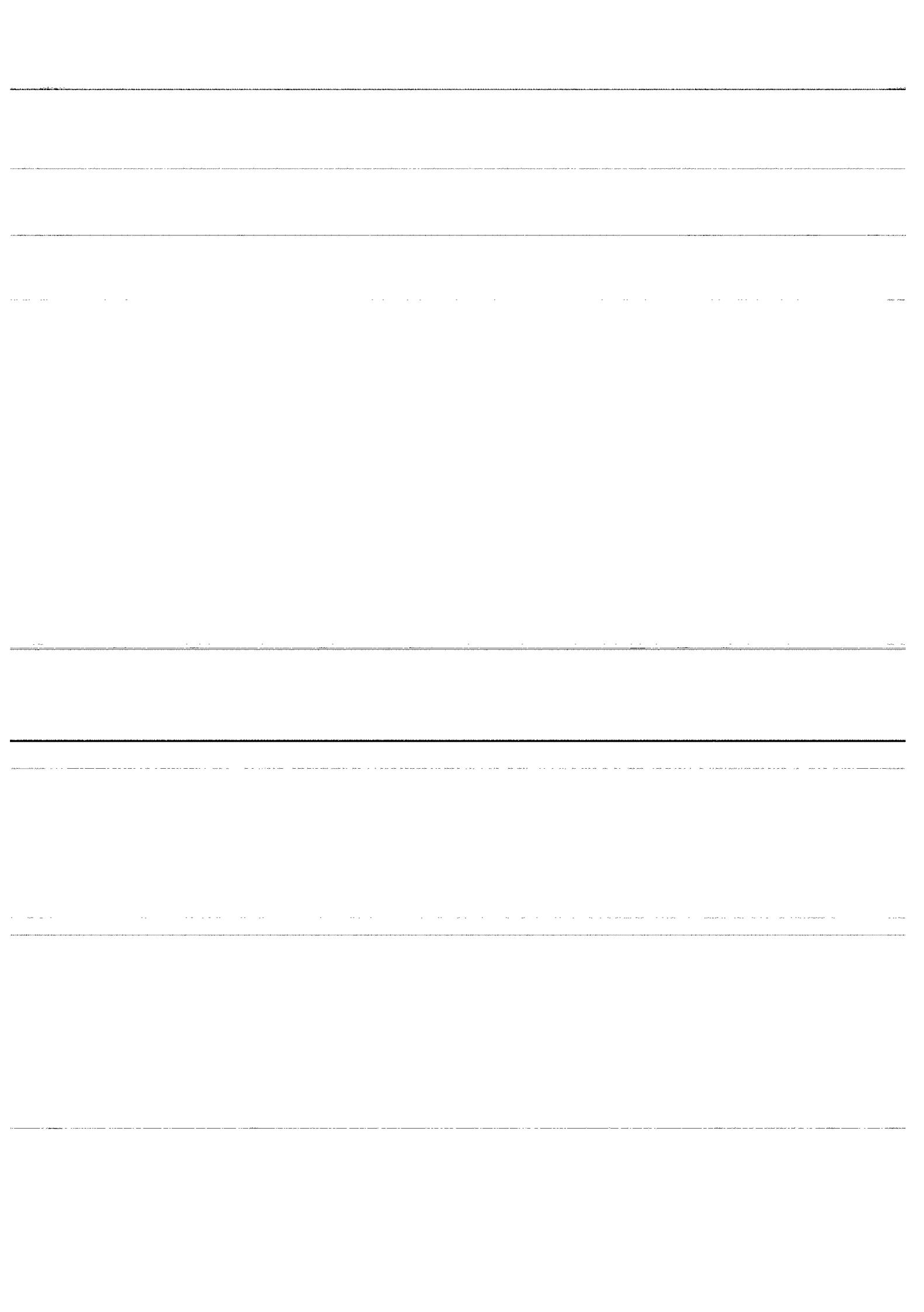
computar verossimilhança.

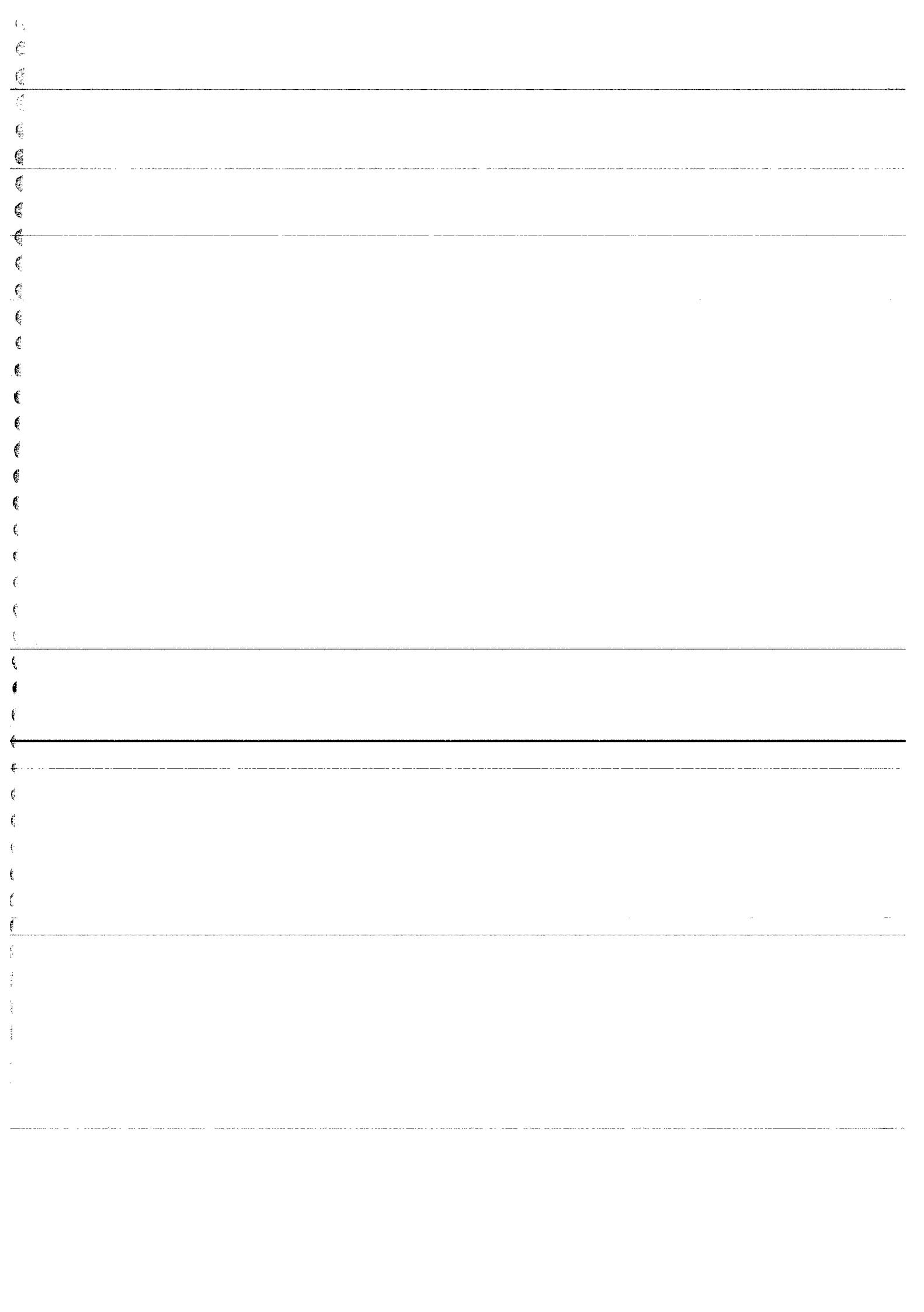
foi vamos calculá-la a partir

de $t = d$.

letras de AK são calculadas desde o início, mas

$\ell(\Psi)$ apenas quando dist. difusa bem definida $(\sum_{t=d}^n)$





(ii). Difusa Exata

contribuições
dada pelo
termo $\log|F_t|$

funcão Verosimilitude: $L(4) \propto -\frac{q}{2} \log k$

Verosimilitud

$$L_d = L(4) + \frac{q}{2} \log k$$

q : n.º componentes nas estacionárias
do modelo

padras

nas c.
adequadas

p/ trabalhar

termos que
necessitam ser
corrigidos

$$- np \log(2\pi) - \frac{1}{2} \sum_{t=1}^d w_t - \frac{1}{2} \sum_{t=d+1}^n [v_t F_t^{-1} v_t + \log|F_t|]$$

de $t=d+1$ a n : forma padras

$d=t=d$: 1^{as} observações tratadas de forma diferente

$$w_t = \log|F_{0,t}| \propto F_{0,t} \text{ é p.d.}$$

$$\log|F_{x,t}| + v_t^{(0)} F_{t,x}^{-1} v_t^{(0)} \propto F_{0,t=0}$$

Otimizações: (método numérico, tipo quase Newton)

6

ver software

Stamp

sóli casas
decimais.

usando rufos de

cond. de 1^a e 2^a

orden (gradiente,

Hessiano)

FSS. \Rightarrow forma qual de Algoritmos:

BHHH

$$\psi^{(i+1)} = \psi^{(i)} - \lambda H^{-1}(\psi^{(i)}) g(\psi^{(i)}) \quad i=1, 2, \dots, I^{\max}$$

$(\mathbf{x}_1) \quad (\mathbf{x}_1) \quad \uparrow \quad \text{escalar} \quad \mathbf{x}_1 \quad \mathbf{x}_2$

Conecta com soluções iniciais: $\mathbf{x} \Rightarrow \sigma_y^2 = 0,001$

$$\sigma_E^2 = 0,1$$

λ : passo p/ controlar a busca pela solução

MUD

se $H(\psi)$ for sempre neg. definida \Rightarrow garante que resultado
mais econômico (n^*
fim de passos)

\Rightarrow A cada instante de $t \rightarrow$ temos um $\psi^{(t+1)}$

Times que avaliam x chegamos no ótimo

Há critérios de parada (ver notas scanneadas sobre otimizações)

$$1. \text{ Verossimilhança: } \frac{l(\psi^{(t)}) - l(\psi^{(t-1)})}{|l(\psi^{(t)})|} < \epsilon$$

$$2. \text{ Gradiente: } \frac{dl}{d\psi} = 0$$

mais importante

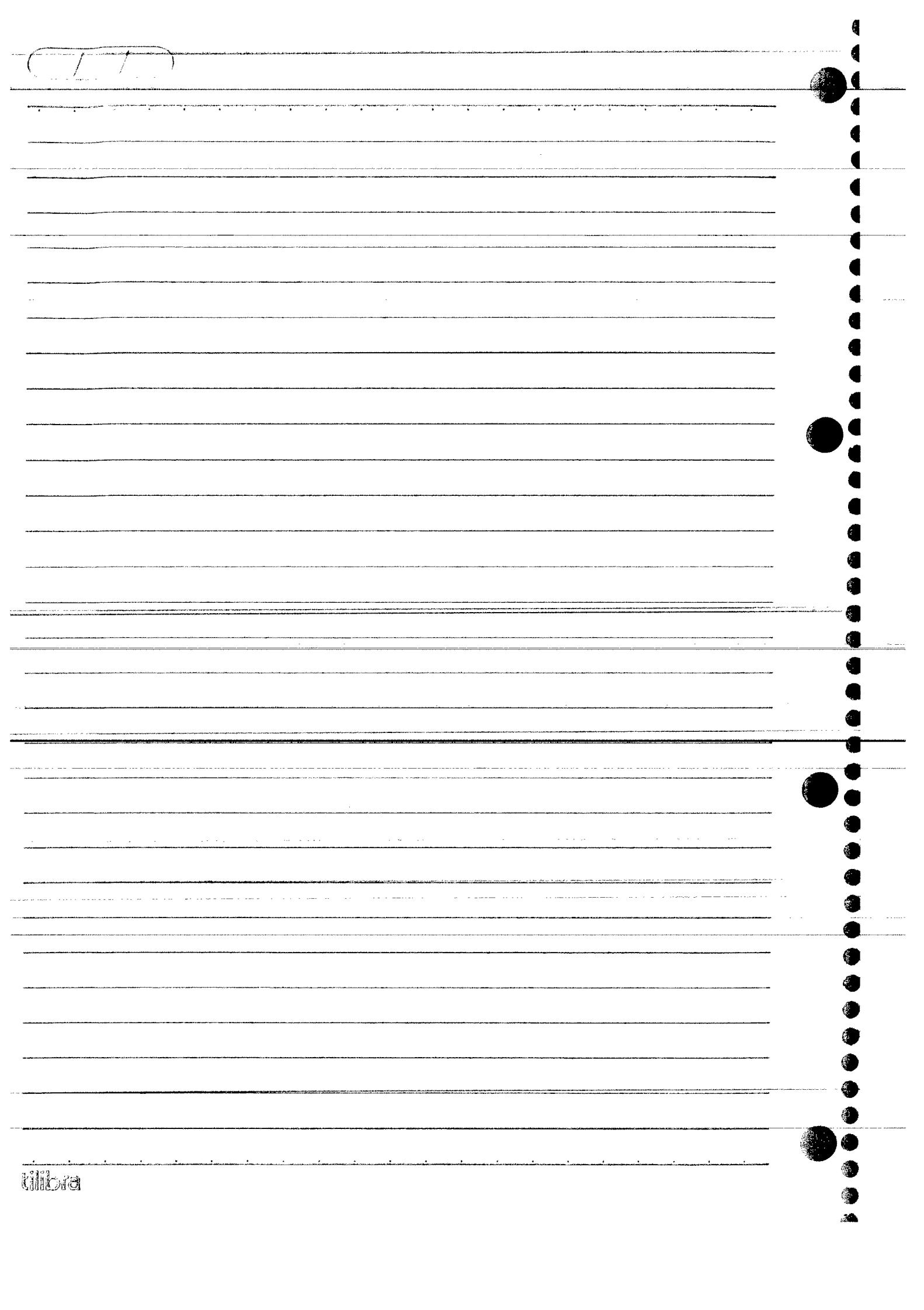
Imaginando que temos:

$$\begin{pmatrix} \frac{dl}{d\psi_1} \\ \frac{dl}{d\psi_2} \\ \vdots \\ \frac{dl}{d\psi_n} \end{pmatrix} = \begin{pmatrix} \\ \\ \\ \end{pmatrix}$$

$$\text{se } \sum_{i=1}^n \left| \frac{dl}{d\psi_i} \right| \approx 0$$

Se na média for muito pequeno \Rightarrow
podemos considerar que grad = 0.

3. Parâmetros:



(20/04/11)

Estimadores de Hiperparâmetros

$$\hookrightarrow \text{para MLE} \quad \hat{\psi} \sim \mathcal{N}(x_t, \alpha_t)$$

Estimados pelo princípio da máxima Verossimilhança
 (apresentam propriedades desejáveis)

- 1) viciados $\lim_{n \rightarrow \infty} E(\hat{\psi}_n) \rightarrow \psi$
- 2) consistência $\lim_{n \rightarrow \infty} \hat{\psi}_n = \psi$
- 3) Eficiência $\text{var}(\hat{\psi}_n) \rightarrow \text{LCR}$ (Cramer Rao)
- 4) Dist. Assintótica normal $\hat{\psi} \sim N(\psi, I'(\hat{\psi}))$

(importante \forall IC's e TH's
 \forall os estimadores)

Problema \forall considerar amostras

muito grandes (\rightarrow que $n = \infty$)

Método Bootstrap: \forall avaliar as distribuições
 \forall diferentes tamanhos de amostra.

Bootstrap: técnicas \forall obter dist. de estimadores \forall tamanhos
 de amostra aleatórios.

$$L(\psi) = \prod_{t=1}^n p(y_t | \psi_{t-1}; \psi)$$

$$\text{Vimos que } p(y_t | \psi_{t-1}) = (2\pi)^{-p/2} |\mathbf{F}_t|^{-1/2} e^{-\frac{1}{2} \mathbf{v}_t^\top \mathbf{F}_t^{-1} \mathbf{v}_t}$$

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{z}_t \boldsymbol{\alpha}_t$$

precisa do FK \forall implementar verossimilhança

Mais adiante trabalhar com $\ell(\psi) = \log L(\psi)$

$$\frac{\partial \ell}{\partial \psi} = 1 \cdot \frac{\partial L}{\partial \psi} = 0 \Rightarrow \max_{\psi} \text{de } L(\psi) = \max_{\psi} \text{de } \ell(\psi)$$

$$\ell(\psi) = \log L(\psi) = \sum_{t=1}^n \log p(y_t | x_{t-1})$$

ℓ_t contribuições da t-ésima componente em $\ell(\psi)$

$$\therefore \ell(\psi) = -\frac{p}{2} n \log(2\pi) - \frac{1}{2} \sum_{t=1}^n [\log |F_t| + v_t' F_t^{-1} v_t]$$

mas é constante

pois é cte.

(n depende dos hiperparâmetros)

Kernel das funções de Verossimilhança

Verossimilhança por
chamada de decomposições do erro de previsão

inovações

(verossimilhança conta tendo v_t como parâmetro)

$$\text{onde: } v_t = y_t - z_t a_t$$

$$F_t = z_t' P_t z_t + h_t$$

Problema: como funciona a questão da inicialização
do FK na verossimilhança?

Como será tratada?

Inicializações do FK na verossimilhança:

duas formas \rightarrow prior difusa via "big kappa"
 \rightarrow inicializações exatas

$$\pi_i \alpha_i \sim N(\bar{\alpha}_i, \bar{P}_i)$$

i) Prior difusa

$$\text{Ex: } \eta_t = \mu_E + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_E^2)$$

$$\mu_{t+1} = \mu_E + \eta_E, \quad \eta_E \sim N(0, \sigma_\eta^2)$$

$$\text{Hyperparâmetros} \Rightarrow \Psi = (\sigma_E^2, \sigma_\eta^2) \text{ ou } (\eta, \sigma_E^2)$$

Vantagem

$$\hat{\sigma}_E^2 = \frac{\sum v_t' F_t^{-1} v_t}{n} \rightarrow \text{so' depen. de } q$$

Procura da verossimilhança pode

ser feita só em relações a q e encontrar

σ_E^2 em funções de q .

Reparametrizações: Reseve P_t e F_t em funções de σ_E^2 (dividindo por)

acha P_t^* e F_t^*

funções de q

$$l(\sigma_E^2, q) \rightarrow dl = 0 \rightarrow \text{terí' termo}$$

$$\frac{d\sigma_E^2}{d\sigma_E^2} \quad v_t' F_t^{-1} v_t$$

apenas como função de q .

No exemplo: $p = m = 1$

$$Z = T = R = 1$$

corresponde a

$$l(\Psi) = -n \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \left[\log F_t + \frac{v_t^2}{F_t} \right]$$

$\sum v_t' F_t^{-1} v_t$ unidimensional

Inicialização difusa: $\alpha_0 \sim N(0, K)$, $K \sim 10^3$

Em $t=1$: distribuição \bar{n} é bem definida

=) considera tudo começando em $t=2$

$$P_t | t=1 \Rightarrow \text{calcula } \tilde{F}_t = P_t + \sigma_E^2 \rightarrow P_{t+1}$$

até

$$l(\Psi) = - \frac{1}{2} (n-1) \log(2\pi) - \frac{1}{2} \sum_{t=2}^n (\log f_t + v_t^2/f_t)$$

Outra questão é o número de componentes nas estacionárias
no vetor de estado.

Caso Mais comum é ter comp.
nas estacionárias

x_t : tem q componentes nas estacionárias

$x_t \sim m \times 1$ onde $q \leq m$

$$\Rightarrow l(\Psi) = - \frac{1}{2} p(n-q) \log(2\pi) - \frac{1}{2} \sum_{t=q+1}^n [\log |f_t| + v_t^2 f_t^{-1} v_t]$$

Na implementação, só passa a considerar t p/ $t = q+1$

$$\alpha = 0 \quad P = K$$

$$t = 0 \quad \rightarrow v_0 = y_0 - \alpha_0$$

$$\text{for } t = 1, n \quad \begin{matrix} P_t = \dots \\ f_t = \dots \end{matrix}$$

$$l = l - \frac{1}{2} (\log f_t + v_t^2/f_t)$$

If $t = 1$ a q

(computa informações, mas n p/ a
verossimilhança)

2) Inicializações Exatas

normalmente considera $q \times n^2$ de
variáveis n estacionárias

Faz de: $t=1$ a d a $d+1$ a n

FK "equitativo". Rada. FK. padras.

Podeu acontecer 2 coisas: $F_{0,t} \neq p.d.$ → mais comum

$$F_{0,t} = 0$$

Na verossimilhança:

$$\ell(\psi) = -\frac{pn}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^d \left(\quad \right) - \frac{1}{2} \sum_{t=d+1}^n \left(\quad \right)$$

usa eq.

cf alterações

p. o FK (que

vimos na

aula passada)

usa equações

do FK padrão

$$\ell_d(\psi) \hat{=} \lim_{k \rightarrow \infty} (\ell(\psi) + g \log k)$$

Nº de componentes não estacionárias

a contribuição em $\ell(\psi)$ de α_i é

dada por $-g \log k$

(Koopman não mostra,

apenas informa)

Término é anulado e fica apenas $\ell(\psi)$

$$\therefore \ell_d(\psi) = -\frac{(np)}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^d w_t + \sum_{t=d+1}^n (\log |F_{0,t}| + v_t^\top F^{-1} v_t)$$

$$w_t = \begin{cases} \log |F_{0,t}|, & F_{0,t} \text{ é p.d.} \rightarrow \text{caso mais comum} \\ 0, & F_{0,t} = 0. \end{cases}$$

$$\log |F_{0,t}| + v_t^\top F_{0,t}^{-1} v_t, \quad F_{0,t} = 0.$$

Obs: Interessante fazer no prox exercício inc. b) kappa,

initialização exata, m.veros. e ver se faz muita diferença

Hamilton: tem uma regras boa

Otimizações Numérica

de otimizações numérica

(mas nem todos os modelos atendem)

$$\hat{\psi} = \arg \max_{\psi \in \mathbb{R}} l(\psi)$$

$\cup \times \downarrow$

$$1^{\text{a}} \text{ ordem: } \frac{dl}{d\psi} = 0 \Rightarrow \begin{pmatrix} \frac{dl}{d\psi_1} \\ \vdots \\ \frac{dl}{d\psi_v} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

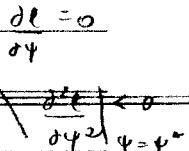
condições
necessárias

$g(\psi)$ pode ser
 $\cup \times \downarrow$ mínimo de

é o vetor gradiente
ou vetor de slope

ponto de xela

No caso univariado:



Mas funções l não são necessariamente côncavas.

Pode ter maximum locais

$$2^{\text{a}} \text{ ordem: } H(\psi) = - \frac{\partial^2 l}{\partial \psi \partial \psi} \Big|_{\psi = \hat{\psi}} > 0$$

Hessiano

$$\text{onde } H(\hat{\psi}) = - \frac{\partial^2 l}{\partial \psi \partial \psi} = \begin{pmatrix} \frac{\partial^2 l}{\partial \psi_1^2} & \frac{\partial^2 l}{\partial \psi_1 \partial \psi_2} & \dots & \frac{\partial^2 l}{\partial \psi_1 \partial \psi_v} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 l}{\partial \psi_v \partial \psi_1} & \frac{\partial^2 l}{\partial \psi_v^2} & & \end{pmatrix}_{v \times v}$$

$$\Rightarrow H(\hat{\psi}) = H|_{\psi = \hat{\psi}}$$

Como $H = - \frac{\partial^2 l}{\partial \psi \partial \psi} \hat{\psi}$, precisaremos que H positiva

definida (e não negativa dependendo)

Entrada: $\mathbf{z}^T \mathbf{H} \mathbf{z} > 0$, i.e., \mathbf{H} é positiva definida (p.d.)
 \mathbf{z} é vetor $V \times 1 \neq 0$.

Se $v=2 \Rightarrow \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \Rightarrow h_{11} = \frac{\partial^2 \ell}{\partial \psi_1^2} > 0$
 $h_{22} = \frac{\partial^2 \ell}{\partial \psi_2^2} > 0$

↓
 valores da diagonal de \mathbf{H}

Entretanto, dificilmente temos soluções analítica.

Precisaremos de um método numérico para realizar a otimização.

a) Métodos do gradiente: algoritmos de otimizações numérica definidos pela seguinte estrutura

- i. procedimento iterativo inicializado com condições inicial
- ii. iterações produzem soluções aproximadas $\hat{\psi}^{(1)}, \hat{\psi}^{(2)}, \dots, \hat{\psi}^{(k)}$
- iii. a cada iteração k , a solução é testada para a parada a partir de uma regra de parada

Dependendo da cond. inicial \rightarrow pode parar em um máximo local

Vale a pena mexer na cond. inicial. Se parar em pontos diferentes \Rightarrow indicações de vários máximos locais

Por exemplo, uma das formas de implementar é:

$$\psi^{(k+1)} = \psi^{(k)} + \lambda \mathbf{H}^{-1}(\psi^{(k)}) g(\psi^{(k)}) \Rightarrow \text{após cada iteração, compare a regra de parada.}$$

$\lambda = 1, 2, \dots, k_{\max}$

λ é implementado de diferentes maneiras. Pode ser um $\lambda^{(k)}$.

ilibra

Esses tipos de algoritmos são chamados quasi-Newton.

Fornecemos $\ell(\psi)$ e fazemos expansões de Taylor de 2º ordem em torno de uma iteração arbitrária de ψ dada por $\psi^{(k)}$.

$$\ell(\psi) = \ell(\psi^{(k)}) + \left(\frac{\partial \ell}{\partial \psi} \right)' (\psi - \psi^{(k)}) + \frac{1}{2} (\psi - \psi^{(k)})' \frac{\partial^2 \ell}{\partial \psi^2} \frac{(\psi - \psi^{(k)})'}{v \times v}$$

\downarrow

$1 \times v \quad v \times 1 \quad 1 \times v \quad v \times v$

Por construção, queremos que ψ seja ponto de máximo de modo que satisfaça $\frac{\partial \ell}{\partial \psi} = 0$

Derivando a expressão:

$$g(\psi^{(k)}) - H(\psi^{(k)})(\psi - \psi^{(k)}) = 0$$

$$H(\psi^{(k)})(\psi - \psi^{(k)}) = g(\psi)$$

$$\boxed{\psi^{(k+1)} = \psi^{(k)} + \lambda H^{-1}(\psi^{(k)}) g(\psi^{(k)})}$$

Se fosse função quadrática, chegaria no máximo na 1ª iteração.

Mas no caso trabalhamos com iterações.

chama ψ de $\psi^{(k+1)}$

e λ reai o tamanho do passo.

Critérios de Parada:

(Normalmente é usado 1 ou mais critérios simultaneamente.
Mais importante é o do gradiente)

(i) critério do gradiente: $g(\psi^{(k)}) = 0$

equivalente a:

$$\begin{pmatrix} \frac{\partial L}{\partial \psi_1^{(k)}} \\ \frac{\partial L}{\partial \psi_2^{(k)}} \\ \vdots \\ \frac{\partial L}{\partial \psi_n^{(k)}} \end{pmatrix} = 10^{-m}$$

dos módulos

Na prática, toma-se a média \bar{v} e verifica se é menor de cada um muito pequeno.

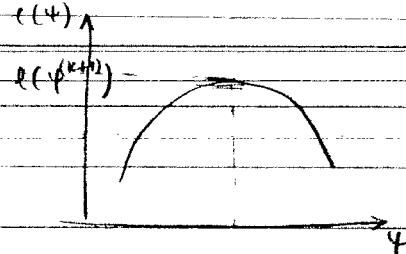
$$c_1 = \sqrt{\sum_{i=1}^n |g_i(\psi^{(k)})|^2} < 10\varepsilon$$

\rightarrow Nas é a única forma.

Poderia pegar o máximo do vetor $g(\psi^{(k)})$ e ver se é $< 10\varepsilon$

(ii) critério de verossimilhança

No max: variações da função e do ponto tem que ser muito pequena



$$c_2 = \left| \frac{e(\psi^{(k)}) - e(\psi^{(k+1)})}{e(\psi^{(k)})} \right| < \varepsilon$$

↳ variação relativa

(iii) critério paraâmetro

$\hat{\theta}$ é a média do valor do parâmetro

se medindo muitas

(equivalente ao que fizemos com
a verossimilhança)

$$G_3 = \frac{1}{n} \sum_{j=1}^n \left| \frac{\psi_j^{(k+1)} - \psi_j^{(k)}}{\psi_j^{(k)}} \right| < \epsilon$$

Normalmente, alguns parâmetros têm restrições de valores.

Ex: $\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix}$ $\sigma_e^2 > 0$

Para usar otimizações sem restrições: faz uma reparametrização η os estimadores que necessitam.

Ex: $\hat{\psi}_1 = e^{\hat{\theta}}$ $\theta \in \mathbb{R}$

\hookrightarrow variância xida dada por $\sigma_e^2 = e^{2\hat{\theta}}$

$$\therefore \sigma_e = e^{\hat{\theta}}$$

qualquer $g(\cdot) : (0, \infty) \rightarrow \mathbb{R}$ é permitida.

(ver nas anotações as transformações propostas)

Ex: $y_t = \mu_t + \varepsilon_t \sim \sigma_e^2$ $\left. \begin{array}{l} l(\Psi) \propto v_t^2 / f_t \\ f_t = p_t + \underbrace{e^{\hat{\theta}}}_{\sigma_e^2} \end{array} \right\}$ parâmetro 1.

$$\mu_{t+1} = \mu_t + \gamma_F \sim \sigma_\gamma^2$$

Quando fizemos otimizações numéricas, ela estará sendo realizada não sobre os parâmetros originais, mas sim em θ .
Teremos:

$$\theta^{(k+1)} = \theta^{(k)} + \lambda H^{-1}(\theta^{(k)}) g(\theta^{(k)})$$

\downarrow

$$\hat{\theta} \sim \mathcal{N}(\theta), \quad \theta = T(\gamma)$$

Incerteza nos estimados: IC

$$\hat{\theta}_j \pm z \sqrt{\text{var}(\hat{\theta}_j)}$$

\downarrow

precisamos da estimativa de $\text{var}(\hat{\theta}_j)$

$$\hat{\theta} \sim N\left(\theta, \frac{I'(\theta)}{n}\right) \text{ onde } I'(\theta) \text{ é a informação de Fisher}$$

$$I(\theta) = E\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta'}\right) = E\left(\frac{\partial \ell}{\partial \theta} \frac{\partial \ell}{\partial \theta'}\right)$$

\downarrow

$$-H$$

\downarrow

$$\left(\frac{\partial \ell}{\partial \theta}\right)'$$

$$\text{Na prática } l = l(\theta) = \sum_{t=1}^n \ell_t(\theta) =$$

$$= -\frac{m}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^n \left[\ln |F_t| + v_t' F_t' v_t \right]$$

dn: q os modelos em EE, teremos

$$E\left[\frac{\partial \ell}{\partial \theta}\right] = \underbrace{\square}_{2 \text{ partes}} + \underbrace{\square}_{\text{uma que nem}}$$

⇒ cálculos dasas valores operados sas complicados

Tomaremos un estimador $\hat{I}(\theta) \rightarrow \hat{I}(\hat{\theta})$

Dous formas de construir $\hat{I}(\theta) \rightarrow$ ver Hamilton (ver notas)

$$\begin{aligned}\hat{I}(\theta) &= \left\{ \begin{array}{l} -1 \frac{\partial^2 l}{\partial \theta \partial \theta'} = -1 \sum \frac{\partial^2 l_t}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}} \end{array} \right. \xrightarrow{\text{I}_{20}} \text{(II)} \\ (\text{estimativa da matriz de informac}) & \quad \text{ou} \\ &= \left\{ \begin{array}{l} \frac{1}{n} \frac{\partial l}{\partial \theta} \frac{\partial l}{\partial \theta'} = \frac{1}{n} \sum \frac{\partial l_t}{\partial \theta} \left(\frac{\partial l_t}{\partial \theta} \right)' \Big|_{\theta=\hat{\theta}} \end{array} \right. \xrightarrow{\text{I}_{op}} \text{(II)}\end{aligned}$$

Na prática, esas derivadas sas calculadas numericamente (nas analiticamente)

$$\frac{de}{dy} = \lim_{h \rightarrow 0} \frac{e(y+h) - e(y)}{h}$$

Supondo que conseguimos fazer cálculos analiticamente:

$$\text{Var}(\hat{\theta}) = E \left[(\hat{\theta} - \theta)(\hat{\theta} - \theta)' \right] = \hat{I}^{-1}(\theta) \quad \text{(III)}$$

por definição
da dist. amostral
de $\hat{\theta}$

De (I) e (II) em (III), temos:

$$\Rightarrow \text{var}(\hat{\theta}) = \frac{1}{n} I'(\theta) = \left\{ - \left[\sum \frac{\partial^2 \ell_{\theta}}{\partial \theta \partial \theta'} \right]_{\theta=\hat{\theta}}^{-1} \right. \\ \left. \left[\sum \left(\frac{\partial \ell_{\theta}}{\partial \theta} \left(\frac{\partial \ell_{\theta}}{\partial \theta} \right)' \right) \right]^{-1} \right|_{\theta=\hat{\theta}}$$

Estimamos a matriz de info $\Sigma \theta$, que são os parâmetros transformados

$$\hat{\theta}_j \pm \sqrt{\text{var}(\hat{\theta}_j)}$$

veem da diagonal de $I'(\theta)$

Mas queremos nos parâmetros originais:

$$\hat{\psi}_j \pm \sqrt{\text{var}(\hat{\psi}_j)}$$

\Rightarrow passar de $\hat{\theta}$ p/ $\hat{\psi}$

Para o vetor $\hat{\psi}$, temos: $\text{var}(\hat{\psi}) = T \text{var}(\hat{\theta}) T'$

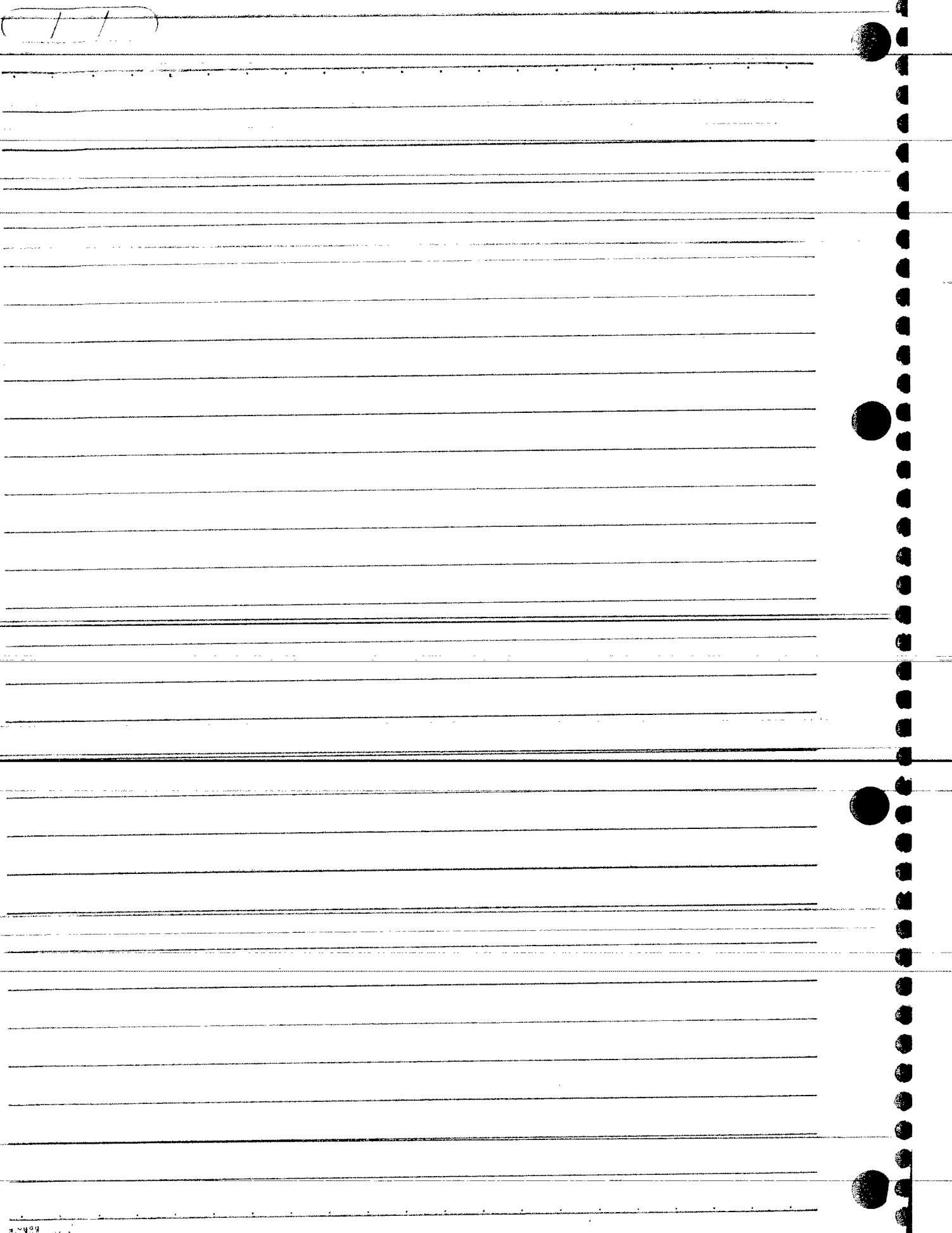
? dividida
(rever)

transformações de $\text{var}(\hat{\theta})$ em
 $\text{var}(\hat{\psi})$

obs: Para cada ψ_j , temos

$\psi_j = T_j(\theta) \Rightarrow$ transformações de ψ original p/
 θ estimado

clássica



Million

cont. otimizações

$$(*) \quad \alpha_t = T_{t-1} \alpha_{t-1} + R_{t-1} \eta_{t-1}$$

$$\therefore E[\alpha_t] = T_{t-1} \alpha_{t-1}$$

$$\text{var}(\alpha_t | \alpha_{t-1}) = E[(\alpha_t - E[\alpha_t | \alpha_{t-1}])(\alpha_t - E[\alpha_t | \alpha_{t-1}])^T]$$

$$\text{mas } \alpha_{t-1} - E[\alpha_t | \alpha_{t-1}] = R_{t-1} \eta_{t-1}$$

$$\therefore \text{var}(\alpha_t | \alpha_{t-1}) = E[R_{t-1} \eta_{t-1} \eta_{t-1}^T R_{t-1}^T]$$

$$= R_{t-1} Q_{t-1} R_{t-1}^T = \text{var}(\alpha_t | \alpha_{t-1})$$

↓

Esta é a variação de $\alpha_t | \alpha_{t-1}$.

mas não podemos escrever dessa forma. Tem a ver com a natureza da matriz R .

Ver exemplo (**)

Temos problema p/ calcular $\det(R_{t-1} Q_{t-1} R_{t-1}^T)$ que aparece na variação.

Problema:

Temos uma transformação do tipo

$$z = a + bx$$

$$\text{p/ calcular } f_Y(y) \Rightarrow f_Y(y) = f_X(x = \frac{y-a}{b}) |J|$$

No caso, temos exatamente isso:

$$\therefore \alpha_t = T_{t-1} \alpha_{t-1} + R_{t-1} \eta_{t-1} \sim N(0, Q_{t-1})$$

última

Como R : $n \times n$ (pois $x = u \alpha + \eta = u x_0$)

ou seja, como dimensões de α e dimensões de η , η calcula
p(α) precisamos usar transformações usando jacobiano.

Nas foram apresentadas todas as contas.

Mas consideraremos sempre R_t da forma

$$R_t = [I_n \ 0']' \Rightarrow R_{t-1}^T R_{t-1} = I_n$$

Trataremos: $\log |R_{t-1}^T R_t + R_{t-1}'| \rightarrow \log q_{t-1}$: precisa fazer o jacobiano das transformações (não explicitado)

Na verdade, temos: $(\alpha_t - T_{t-1} \alpha_{t-1})' (R_t Q R_t)^{-1} (\alpha_t - T_{t-1} \alpha_{t-1})$

quando usarmos o jacobiano

$$(\underbrace{\alpha_t - T_{t-1} \alpha_{t-1}}_{\eta_{t-1}^T} \underbrace{R_{t-1}}_{R_{t-1}^T} \underbrace{Q^T R_t^T}_{R_t^T} \underbrace{\alpha_t - T_{t-1} \alpha_{t-1}}_{\eta_{t-1}})$$

$$\eta_{t-1}^T R_{t-1}^T R_t^T \eta_{t-1}$$

$$\text{de } \alpha_t = T_{t-1} \alpha_{t-1} + R_{t-1} \eta_{t-1}$$

$$\therefore \eta_{t-1} = R_{t-1}^T (\alpha_t - T_{t-1} \alpha_{t-1})$$

— II —

(***) \Rightarrow Resultados:

é uma fórmula analítica de calcular o vetor de sítore (= gradiente)

Tra' utilidade em 2 situações:

a) Em otimizações, métodos de Newton

"

$$\psi^{(i+1)} = \psi^{(i)} + \lambda H^{-1}(\psi^{(i)}) g(\psi^{(i)}) \Rightarrow \text{ótimo.}$$

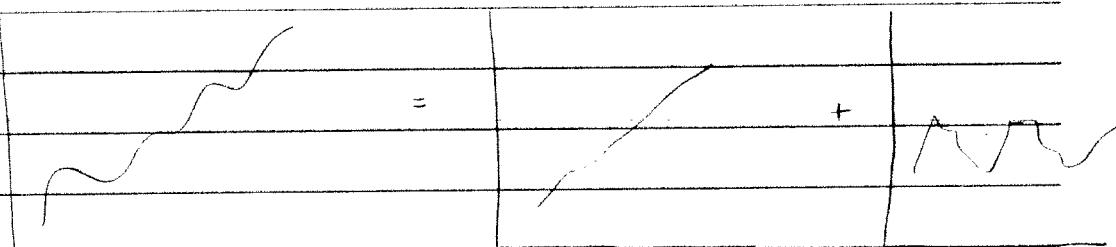
l. d.e. (pode ser analítico,
d ψ mas tbx n'tiverma info.
máscara de otimização de qualquer função objetivo.)

• Macroeconomia 1:

tendência + irregular / ciclo

↳ em economia há irregularidades

na frequência, no tempo etc.



← Choques permanentes
Choques transientes

• Macroeconomia 2:

tendência + sazonalidade + irregular

obj:

Para esta metodologia, é preciso que se

realizem um estacionarização da série.

(como fazemos nos modelos ARIMA - Box & Jenkins)

$$z_t = \Delta^D \Delta^d y_t$$

Aqui, ajusta-se a série original, desde que ela
tenha as componentes propostas.

Se uma série é estacionária de 2º ordem, vale:

Outro algoritmo = EM

pt maximizaç



Tb é um todo iterativo qd adira máximos da função, mas toma + a cara da função que está se do maximizada

Toma a direção do máximo, mas fica lento próximo do máximo.

Normalmente: usa-se o EM qd ter uma cond. inicial melhor.

Quando começar a ficar lento \Rightarrow joga condiç inicial no algoritmo quasi-Newton.

Vantagem: ter uma cond. inicial melhor.

usar apenas 1^o derivada (no EM)

No quasi-Newton entra 2^o derivada tb.

Modelos

Vamos olhar pt srie como componentes nos observáveis.

ST = Tendência + Sazonalidade + Ciclo + Irregular



demandas etapas do

aumento de ano, conven-

população, cões sociais

aumento de

riqueza

11

representações em MA infinito:

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

$$\begin{aligned} y_t &= y_t - \mu = \sum \psi_j \\ &= \Psi(L) \epsilon_t \\ &= \theta_q(L) \epsilon_t \quad \Rightarrow \text{ARMA} \end{aligned}$$

$$\Phi_p(L) y_t^* = \Theta_q(L) \epsilon_t$$

Mas x séries não estacionárias, mas há formalizações de que componentes devem ser usadas, justificativa.

Mas são ferramentas bem usadas (tendência + sazonalidade + ...)

Fatores estilizados



de séries de retornos financeiros, por exemplo:
- excesso de cointe etc.

de séries macroeconômicas

- tendência e ciclo

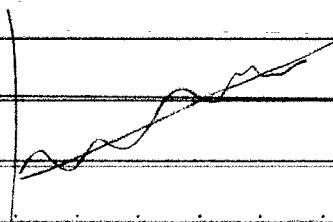
- tendência e sazonalidade

• Estas componentes mais estocásticas e locais

determinística

global

Obs: se considerarmos tendência linear



$$y_t = a + b t + \epsilon_t$$

$$\hat{y}_t = \hat{a} + \hat{b} t$$

$y_t - \hat{y}_t = \hat{\epsilon}_t \Rightarrow$ fazendo assim, amplia o erro, o ciclo.

equilíbrio

se tendência fornece estocástica \Rightarrow "erro" diminui

Nelson e Plosser (1972) \Rightarrow propõem como melhor
tendência estocástica

com tendência e sazonalidade

$$y_t = \mu_t + \gamma_t + \epsilon_t$$

\downarrow \downarrow

tendência sazonalidade

Série ajustada dessazonalizando:

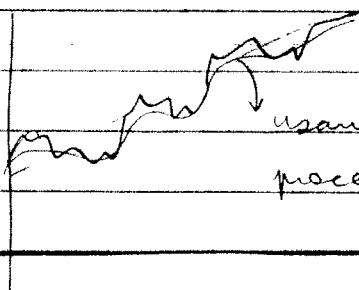
$$y_t^{(a)} = y_t - \hat{\gamma}_t \quad \left\{ \begin{array}{l} \text{x12-ARIMA} \\ \text{SEATS TRAMO (órgãos da CE).} \end{array} \right. \quad \begin{array}{l} \text{metodologia de dessazonalização,} \\ (\text{IBGE; órgãos dos EVA}) \end{array}$$

dependendo de como é especificado,
pode ter séries ajustadas diferentes

Modelos de Tendência

Componente de baixa freqüência ("mudanças no longo do tempo")

Ex:



usando médias móveis
procedimento descritivo

$$\hat{y}_t^* = \frac{1}{3} (y_{t-1} + y_t + y_{t+1})$$

eliminação de
altas freqüências
Assim mais conseguimos
tirar componente sazonal

77

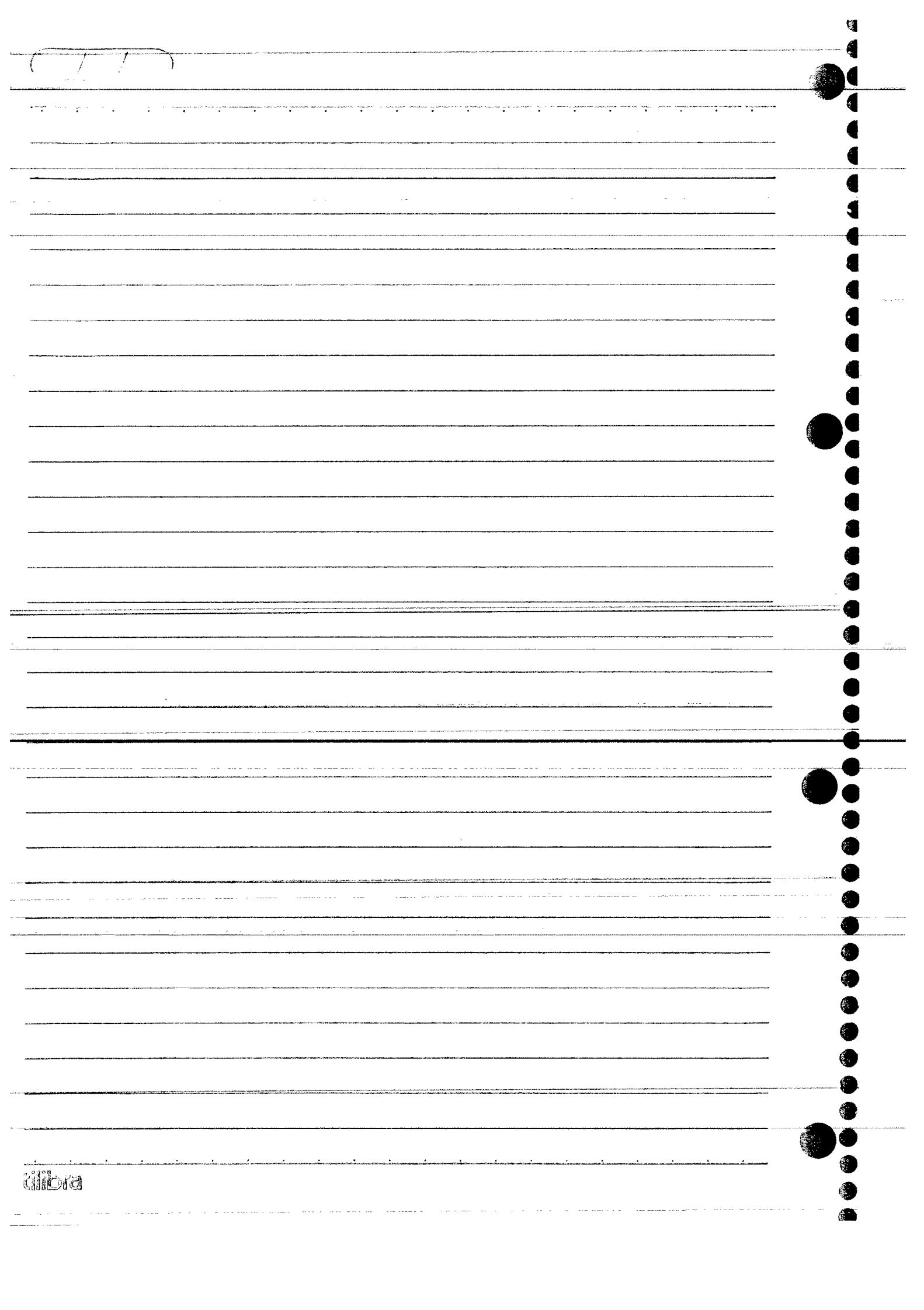
Obs importante:

Precisa especificar a correlação entre os componentes.

Só pode ser 0 ou $\pm \frac{1}{2}$.

Se for $\neq 0$, sistema é não identificável

Dar eq. de FR \Rightarrow achar eqs. de trânsito



Da equações (I) do modelo:

$$\begin{aligned} E[y_t | y_{t-1}] &= E[2_t \alpha_t + \varepsilon_t | y_{t-1}] \\ &= 2_t E[\alpha_t | y_{t-1}] \\ &= 2_t \alpha_t \end{aligned}$$

Logo: $v_t = y_t - 2_t \alpha_t$

intuições: o que queremos dizer é que o conteúdo informativo
mas é modificado se substituirmos y_t por v_t .
(pois α_t só depende de passado)

Podemos fazer isso pois g é bijetiva.
é como se $g(y)$ fosse v_t (a inovação).

Então:

$$E[x_t | y_t] = \alpha_t x = E[x_t | y_{t-1}, v_t]$$

Usando (a): fazendo $x = \alpha_t$

$$y = y_{t-1}$$

$$z = v_t$$

precisamos provar

que a situação é de

retros tri-variações onde $\mu_2 = 0$ e $\Sigma_{zy} = 0$.

Temos que provar que estas condições são satisfeitas.

$$(i) E[z] = \mu_2 = 0 \Rightarrow E[v_t] = 0$$

$$E[v_t] = E[E[v_t | y_{t-1}]] \quad (\text{lei da expectativa iterada})$$

$$\therefore E[v_t | y_{t-1}] = E[y_t - z_t a_t | y_{t-1}] =$$

$$= E[y_t | y_{t-1}] - z_t a_t$$

do modelo
 $\Rightarrow z_t a_t - z_t a_t = 0.$

$$\therefore E[v_t] = E[0] = 0 \quad \text{ok! (condição (i) é satisfeita)}$$

$$(ii) \sum y_{tj} = 0 \Rightarrow \sum_{y_{t-1}, v_t} \text{cov}(y_{t-1}, v_t) = 0$$

$$\text{cov}(y_{t-1}, v_t) = E[(y_{t-1} - E[y_{t-1}]) (v_t - E[v_t])']$$

0 (já provamos)

$$y_{t-1} = \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_0 \end{pmatrix} = E[y_{t-1} v_t']$$

(cada elemento dessa matriz é dada por:
 $\{E[y_{t-j} v_t']\}_{j=1,2,\dots,t-1}$)

$$\text{mas: } E[y_{t-j} v_t'] = E[E[(y_{t-j} v_t') | y_{t-1}]]$$

lai das
exp. iteradas
para calcular
valor esperado
incondicional

como $t-j$ é passado
pode sair de $E[\cdot]$

$$= E[y_{t-j} E[(v_t' | y_{t-1})]] = 0$$

(ii) ok!

Portanto, podemos usar o resultado (a):

$$E[x_t | y_t] = E[x_t | y_{t-1}] + \text{cov}(x_t, v_t) (\text{var}[v_t])^{-1} v_t$$

↓
já sabemos

que é Precisamos calcular

(passo da prova, em $t-1$) as outras componentes da equação.

04/05/11

Tendência e ciclo: caráter subjetivo

depende muito de onde e de que horizonte estamos olhando
a série

Normalmente, ciclo pode ser tratado tb de forma não periódica

podemos ajustar uma tendência estocástica (e variabilidade em um ciclo).

em séries econômicas,
por exemplo.

Tendência:

→ Determinística (TS)

(no tempo)

→ Estocástica (DS)

a) Tend. Determinística

processo ARMA

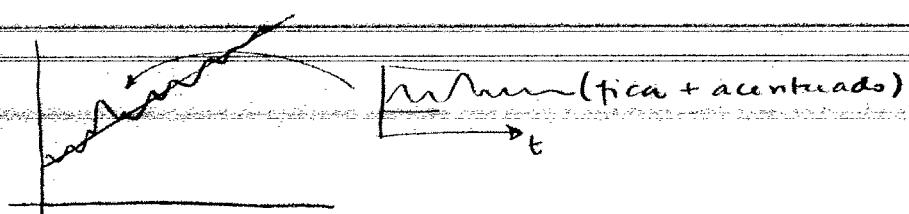
$$y_t = \sum_{j=0}^d \beta_j t^j + \psi(L) a_t, \quad a_t \sim NID(0, \sigma^2) \quad (\text{podríamos ter log } y_t, \text{ ou inversas})$$

onde $\psi(L) = 1 - \psi_1 L - \psi_2 L^2 - \dots$ (se colocarmos apenas ruído branco, ainda teríamos correlações)

função polinomial do tempo (normalmente $t=2$)

9º mais suave para a tendência \Rightarrow maior é o ciclo (flutuações
em torno da tendência)

(mas é adequado para nas séries econômicas, a tendência
é essencialmente estocástica)



\Rightarrow Modelo típico de ST: parte estacionária é ARMA

"n" "é tend. determ."

Cov e autocovar desse modelo \Rightarrow Vem do ARMA

Ex: $d=1$ e $a_t = AR(1)$

boa correção PRB nas autocorrelações.

$$y_t = \beta_0 + \beta_1 t + \epsilon_t$$

$$a_t = \phi a_{t-1} + \epsilon_t$$

$$(1-\phi L) a_t = \epsilon_t$$

$$\therefore a_t = \frac{\epsilon_t}{1-\phi L}$$

$$y_t = \beta_0 + \beta_1 t + \frac{\epsilon_t}{1-\phi L}$$

$$(1-\phi L) y_t = (1-\phi L) \beta_0 + (1-\phi L) \beta_1 t + \epsilon_t$$

$$(1-\phi L) y_t = (1-\phi L) \beta_0 + \beta_1 (t - \phi(t-1)) + \epsilon_t$$

$$(1-\phi L) y_t = (1-\phi) \beta_0 + \beta_1 (t - \phi t + \phi) + \epsilon_t$$

$$(1-\phi L) y_t = (1-\phi) \beta_0 + \beta_1 \phi + (\phi - 1) \beta_1 t + \epsilon_t$$

Queremos:

$$w_t = y_t - y_{t-1}$$

$$\begin{aligned} y_t - y_{t-1} &= (1-L) y_t = (1-L) \left(\beta_0 + \beta_1 t + \frac{\epsilon_t}{1-\phi L} \right) \\ &= \beta_1 (t - (t-1)) + \frac{\epsilon_t}{1-\phi L} (1-L) \end{aligned}$$

$$= \beta_1 + (1-L) \frac{\epsilon_t}{1-\phi L}$$

$$(1-\phi L)(1-L) y_t = (1-\phi) \beta_1 + (\epsilon_t - \epsilon_{t-1})$$

MA de raiz unitária
(A=1)

Nas é bom pois nas é inversível (problema na identificação, na previsão etc.)

\Rightarrow Processo q. tendência determinística: não é recomendado tirar tendências por 1ª diferença

(**) Modelos de tendência linear estocástica

$$y_t = \mu_t + (\varepsilon_t)$$

$$\mu_{t+1} = \mu_t + \beta_t + \eta_t$$

$$\beta_{t+1} = \beta_t + \zeta_t$$

→ quando $\sigma_\eta^2 = \sigma_\zeta^2 = 0 \Rightarrow$ passa a ser determinística.

$$\begin{cases} H_0: \sigma_\eta^2 = \sigma_\zeta^2 = 0 \Rightarrow \text{tão complicado de fazer pois estamos} \\ \text{na fronteira do espaço paramétrico (pois var} \\ \text{entre } 0 \text{ e } \infty) \\ H_A: \text{c.c.} \end{cases}$$

$\left(\begin{array}{l} \sigma_\eta^2 = 0 \\ \sigma_\zeta^2 \neq 0 \end{array} \right) \text{ ou } H_0 = \text{equivalente ao processo TS}$

$\left(\begin{array}{l} \sigma_\eta^2 \neq 0 \\ \sigma_\zeta^2 = 0 \end{array} \right) \text{ ou } H_A = \text{ " " " " DS.}$

→ Diferente do caso da regressão em que testamos $H_0: \beta_1 = \beta_2 = 0$

$$(y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t + \beta_3 x_t + \varepsilon_t) \quad \text{pois } \beta_1 \text{ e } \beta_2 \text{ estão em } \mathbb{R}^2$$

Ajusta o modelo e estima σ_η^2 e σ_ζ^2

se estimativas muito pequenas: evidência
de tendência determinística

(obs: na verdade, devemos olhar q_η e q_ζ)

— //

→ Vantagem do arcabouço dos ME: há uma transição suave entre
TS e DS

é que o suficiente M caracterizar os 2 tipos.

Na outra formatação, é necessário pré-definir

no caso ME: vai se ajustar naturalmente. Já faria na estimativa
dos parâmetros (das variâncias)

— //

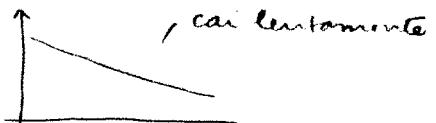
Artigo: Nelson e Plosser.

- Séries tabela 2 \Rightarrow cálculo das autocorrelações de cada série
(caem lentamente, exceto tx. desemprego)
- Tabela 3 \Rightarrow pega a 1^a diferença

1) $z_t = \log y_t \rightarrow$ FAC (Tabela 2)

2) $\Delta z_t = \Delta \log y_t \rightarrow$ FAC (Tabela 3)

Se $z_t = \alpha + z_{t-1} + a_t \Rightarrow$ FAC



Se $\Delta z_t = \alpha + a_t \Rightarrow$ FAC



\Rightarrow típica de processo estacionário

Pode ser consistente com os ou com DS.

Não é suficiente pegar diferenças ainda

3) $z_t = \beta_0 + \beta_1 t + a_t$ (Tabela 4)

$\hat{a}_t = z_t - (\hat{\beta}_0 + \hat{\beta}_1 t) \rightarrow$ FAC

\rightarrow Faz regressão, estimando via MPO

Estima processo DS via TS

\rightarrow FAC decai lentamente se o erro

Muita estrutura de dependência

$\quad \quad \quad$ típico de séries DS
estimada via TS

Cores 1, 2, 3 \Rightarrow evidência de que para séries macroeconómicas,
grande maioria é DS (e nas TS)

Aleia das evidências:

Faz teste da raiz unitária H_0 o modelo considerando tendência estocástica e determinística.

Tabela 5: estimativa de δ (na parcela δt) \Rightarrow muito pequeno
estimativa de ρ (na parcela $\rho \cdot z_{t-1}$)

$$\text{Estatística: } t = \frac{\hat{\rho} - 1}{\text{SE}(\hat{\rho})}$$

\Rightarrow Pr maioria das séries não pode rejeitar $H_0: \rho = 1 \Rightarrow$ tem raiz unitária

$H_0: \delta = 0 \Rightarrow$ pode ser
testado pela
estatística t-padrão

\hookrightarrow mas pode usar estatística
t-padrão (ataria na
fronteira do espaço
paramétrico)

t_b nas rejeita $H_0: \delta = 0$ \forall algumas delas estatisticamente.
mas em termos absolutos, δ é muito pequeno.

É significante, mas não é relevante.

\rightarrow Evidência de tendência estocástica.

(típico de séries macroeconômicas)

