



Módulo de Regressão e Séries Temporais

Parte 4

Mônica Barros, D.Sc.

Julho de 2007

monica@mbarros.com

1



Quem sou eu?

□ Mônica Barros

- Doutora em Séries Temporais – PUC-Rio
- Mestre em Estatística – University of Texas at Austin, EUA
- Bacharel em Matemática – University of Washington, Seattle, EUA
- Professora da PUC-Rio (Depto. De Eng. Elétrica)
- E-mails: monica@ele.puc-rio.br, monica@mbarros.com
- Home page: <http://www.mbarros.com>



monica@mbarros.com

2



REGRESSÃO DINÂMICA (Formulação Teórica)

monica@mbarros.com

3



REGRESSÃO DINÂMICA

□ Preliminares

- O termo “dinâmico” não quer dizer que os parâmetros do modelo evoluem no tempo.
- No contexto atual, o termo “dinâmico” está relacionado à estrutura de defasagem da série temporal e das variáveis causais do modelo.
- No caso mais geral também existe uma estrutura de dependência dos erros, ao contrário da especificação “iid” na regressão convencional.

monica@mbarros.com

4

ESTRUTURA DOS MODELOS DE REGRESSÃO DINÂMICA



$$\varphi(B).Y_t = \beta.Z_t + \varepsilon_t$$

□ onde:

- e_t são erros iid $N(0, \sigma^2)$
- Z_t é um vetor de variáveis exógenas no instante t ;
- β é o vetor de coeficientes das variáveis exógenas;
- $\varphi(B)$ é um polinômio AR(p). Se $\varphi(B) = 1$ o modelo se reduz a um modelo de regressão convencional.
- T é o tamanho da série;
- S é o período sazonal.

REGRESSÃO DE COCHRAN ORCUTT GENERALIZADA



- É uma generalização do modelo de Regressão Dinâmica apresentado anteriormente em que se admite uma estrutura de dependência dos erros.

$$\varphi(B).Y_t = \beta.Z_t + \pi_t$$

$$R(B).\pi_t = \varepsilon_t$$

- onde $R(B)$ é um polinômio AR de grau r e os
- e_t são erros iid $N(0, \sigma^2)$;
- Na formulação original de Cochran - Orcutt, o polinômio $R(B)$ tinha grau 1, e então os erros só apresentavam autocorrelação de lag 1, o que talvez seja o caso mais comum na prática.

REGRESSÃO DE COCHRAN ORCUTT GENERALIZADA



- O modelo também pode ser escrito como:

$$R(B)\{\varphi(B)Y_t - \beta.Z_t\} = \varepsilon_t$$

- Ou de maneira mais compacta como:

$$\varphi(B)Y_t^* - \beta.Z_t^* = \varepsilon_t$$

- onde
- $R(B)$ é conhecido como fator comum.

$$Y_t^* = R(B)Y_t ; Z_t^* = R(B).Z_t$$

REGRESSÃO DE COCHRAN ORCUTT



- No caso do modelo de Cochran- Orcutt original, temos $R(B) = 1 - r_1B$ e assim a estrutura do modelo é:

$$\varphi(B) [Y_t - r_1Y_{t-1}] = \beta [Z_t - r_1Z_{t-1}] + \varepsilon_t$$

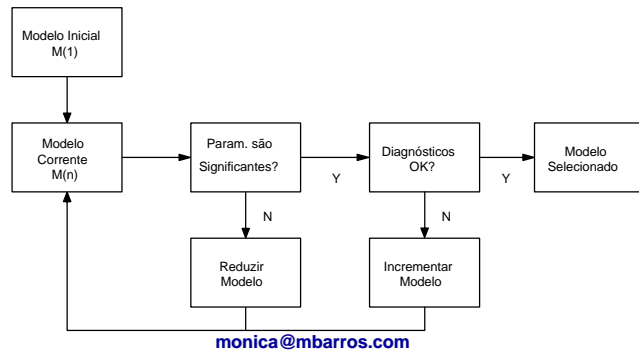
- A estimação de parâmetros é feita seqüencialmente.

REGRESSÃO DINÂMICA



□ Construção dos Modelos

- Empregamos uma estratégia "bottom-up" na construção de um modelo RD, partindo de um modelo mais simples até chegar num modelo mais elaborado.



9

REGRESSÃO DINÂMICA



□ Modelagem da sazonalidade

- Através de:
 - (i) "dummies" sazonais
 - (ii) defasagens sazonais na variável dependente Y_t ou erros estruturados sazonalmente (e_{t-5} , e_{t-25} etc...)
- A existência de autocorrelações significantes nos resíduos pode indicar:
 - (i) Necessidade de inclusão de lags da variável dependente ou,
 - (ii) Necessidade de inclusão de lags das variáveis causais ou novas causais

monica@mbarros.com

10

REGRESSÃO DINÂMICA



□ Modelos ARCH de Regressão

- Servem para modelar séries com elevadas volatilidades, i.e., variâncias dos erros variando no tempo segundo uma determinada lei.
- Modelo ARCH (Autoregressive Conditional Heteroscedastic), desenvolvido por Engle em 1982 consiste num modelo Cochran Orcutt usual, ou seja: $\phi(B)Y_t = \beta Z_t + \varepsilon_t$ onde ao invés dos erros serem iid $N(0, \sigma^2)$, eles são $N(0, h_t)$

monica@mbarros.com

11

REGRESSÃO DINÂMICA



□ Modelos ARCH de Regressão

- h_t é a variância condicional das observações, e é dada pela expressão:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2$$

- isto é, h_t é função dos quadrados dos erros passados.
- Os parâmetros $\alpha_0, \alpha_1, \dots, \alpha_p$ são estimados a partir dos dados.

monica@mbarros.com

12

REGRESSÃO DINÂMICA



- Note que h_t , a variância condicional, depende dos quadrados dos erros passados.
- No software FORECAST MASTER PLUS as seguintes estruturas para h_t são possíveis:
 - $\alpha_1 = \alpha_2 = \dots = \alpha_{p-1} = 0; \alpha_p \neq 0$
 - $\alpha_1 = \alpha_2 = \dots = \alpha_p = \alpha = \text{constante}$
- a_i 's decaem linearmente.

REGRESSÃO DINÂMICA



- Testes de Verificação
 - Como em todo modelo estatístico e de previsão de séries, podemos efetuar testes estatísticos para verificar a adequação do modelo aos dados ("goodness of fit"). Nos modelos RD estes testes podem ser agrupados em:
 - (i) Testes para verificar a especificação do modelo explicativo
 - (ii) Teste visando encontrar a dinâmica do modelo, i.e., a inclusão ou não de variáveis defasadas
 - (iii) Testes para verificar o ajuste do modelo

REGRESSÃO DINÂMICA



- Testes baseados na ACF dos resíduos
 - Idealmente, se a especificação do modelo está correta, os resíduos devem ser "brancos", isto é, não devem apresentar qualquer tipo de estrutura ou dependência.
 - Em particular, as autocorrelações entre resíduos com diferentes defasagens devem ser nulas, pois a existência de autocorrelações significantes nos resíduos revela que algum tipo de estrutura não foi devidamente considerado no modelo.

TESTES BASEADOS NA ACF DOS RESÍDUOS



- (i) Portmanteau ou Ljung-Box
 - Testa a hipótese de que as "k" primeiras autocorrelações são nulas, isto é:
 - $H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$
 - Estatística usada no teste: $Q = \frac{T(T+2) \sum_{j=1}^k r_j^2}{(T-i)}$
 - Regra de decisão
 - $(Q | H_0 \text{ Verdadeiro}) \sim \chi^2_{(k-m)}$, onde "m" é o número de parâmetros livres do modelo.
- **Rejeita-se a hipótese nula** (autocorrelações nulas) se Q é "grande" quando comparado a um percentil apropriado da densidade Qui-quadrado.

TESTES BASEADOS NA ACF DOS RESÍDUOS



□ (ii) Durbin Watson

- Testa apenas se a primeira autocorrelação é nula, isto é: $H_0 : \rho_1 = 0$

- Estatística de teste:
$$DW = \frac{\sum_{t=2}^r (e_t - e_{t-1})^2}{\sum_{t=1}^r e_t^2}$$

- (DW | H_0 Verdadeiro) é próximo de 2.

CRITÉRIOS DE SELEÇÃO E COMPARAÇÃO DE MODELOS



□ Objetivo: comparar diferentes modelos, e escolher o "melhor"

- Critério AIC (Akaike) $AIC = T \cdot \ln \hat{\sigma}_e^2 + 2m$

□ Critério BIC

$$BIC = T \cdot \ln \hat{\sigma}_e^2 - (T - m) \ln \left(1 - \frac{m}{T} \right) + m \cdot \ln T + m \cdot \ln \left[\frac{\left(\frac{\hat{\sigma}_y^2}{\hat{\sigma}_e^2} - 1 \right)}{m} \right]$$

□ Critério de Schwartz

$$SBC = T \cdot \ln \hat{\sigma}_e^2 + m \cdot \ln T$$

TESTES BASEADOS NA TEORIA ASSINTÓTICA



- São testes onde a hipótese nula H_0 representa uma restrição linear de rank "k" dos parâmetros do modelo, i.e. :

- $H_0 : q_1 = q_2 = \dots q_k = 0;$

- $H_1 : q_{1'} \dots q_{k'} \dots q_m \neq 0$

- Os 3 testes existentes são:

- (i) WOLD

- No cálculo da estatística, estima-se q sob a hipótese alternativa ("m" parâmetros estimados)

TESTES BASEADOS NA TEORIA ASSINTÓTICA



- (ii) RAZÃO DA VEROSSIMILHANÇA

- No cálculo da estatística, estima-se q sob ambas hipóteses H_0 e H_1 ("m-k" e "m" parâmetros estimados)

- (iii) MULTIPLICADORES DE LAGRANGE (LM)

- No cálculo da estatística, estima-se q somente sob a hipótese nula ("m-k" parâmetros estimados)

- Estatística do teste:

- onde

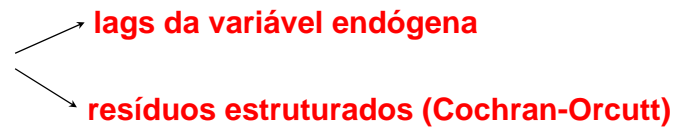
$$LM = T \cdot \lambda(\theta_0) \cdot I^{-1} \cdot \lambda(\theta_0) \sim \chi^2(k)$$

- e I é a matriz de informação de Fisher.

TESTES PARA VERIFICAR A DINÂMICA DO MODELO



□ Dinâmica



□ Nesta bateria de testes teremos:

- H_0 : a dinâmica do modelo está correta;
- H_1 : é necessário incluir outros lags da variável endógena e/ou erros estruturados.

TESTES PARA VERIFICAR A DINÂMICA DO MODELO



□ No FMP e Forecast PRO os seguintes testes de dinâmica estão implantados:

- A) Testes para verificar a dinâmica da variável endógena (Y_t)
 - Teste 1: lag simples da endógena
 - H_1 : O 1o lag da endógena Y_{t-p} ainda não presente no modelo é acrescentado ao modelo;

TESTES PARA VERIFICAR A DINÂMICA DO MODELO



- Teste 2: lag sazonal da endógena
- H_1 : O 1o lag sazonal da endógena Y_{t-ps} ainda não presente no modelo é acrescentado a este;

- Teste 3: Seqüência de Lags da Endógena
- H_1 : Uma seqüência de lags da endógena $Y_{t-1}, Y_{t-2}, \dots, Y_{t-s}$ ainda não presente no modelo é acrescentada a este.

□ B) Testes para verificar a dinâmica dos resíduos

- Teste 4: Lag Simples do Resíduo
- H_1 : O 1o lag do resíduo e_{t-p} ainda não presente no modelo é acrescentada a este

TESTES PARA VERIFICAR A DINÂMICA DO MODELO



- Teste 5: Lag Sazonal do Resíduo
- H_1 : O 1o lag sazonal do resíduo e_{t-ps} ainda não presente no modelo é acrescentado a este

- Teste 6: Seqüência de Lags do Resíduo
- H_1 : Uma seqüência de lags dos resíduos e_{t-1}, \dots, e_{t-s} ainda não presente no modelo é acrescentada a este.

TESTES PARA A ESPECIFICAÇÃO DAS VARIÁVEIS CAUSAIS



- Nesta bateria de testes temos:
 - H_0 : A parte "causal" do modelo está correta
 - H_1 : Outras variáveis causais e/ou outros lags das variáveis causais já incluídas devem ser acrescentados

- Teste 1: Causais Excluídas
 - H_1 : Cada variável causal excluída do modelo atual é acrescentada a este

TESTES PARA A ESPECIFICAÇÃO DAS VARIÁVEIS CAUSAIS



- Teste 2 : Componente de Tendência
 - H_1 : Verifica se a variável exógena tendência, $X_t = t$ (TREND) deve ser acrescentada ao modelo

- Teste 3 : Lags das Causais
 - H_1 : Verifica se as variáveis causais presentes no modelo com um lag extra devem ser acrescentadas ao modelo

TESTES PARA A ESPECIFICAÇÃO DAS VARIÁVEIS CAUSAIS



- Teste 4: Causais não-lineares
 - H_1 : Verifica se os quadrados das variáveis causais (X_{it2}) já incluídas no modelo devem ser acrescentadas a este

- Teste 5: Fator Comum
 - H_1 : Caso exista erro estruturado, este teste verificará se este pode ser substituído por lags de Y_t e/ou das causais X_{it} .

TESTES DE HOMOGENEIDADE



- São testes para verificar a existência de variações na variância dos resíduos (heterocedasticidade dos resíduos)

- H_0 : O modelo está "bom", i.e., os resíduos são totalmente aleatórios sem exibir estrutura de dependência;

- H_1 : Existe dependência nos resíduos.

TESTES DE HOMOGENEIDADE



- Teste 1: Variação com o tempo
- H_1 : Variância dos resíduos varia com o tempo. (Regressão e_{t2} versus t)

- Teste 2: Variação com a Endógena
- H_1 : Variância dos resíduos varia com a endógena Y_t (Regressão e_{t2} versus Y_t)

- Teste 3 : Variação com as Causais
- H_1 : Variância dos resíduos varia com as variáveis causais (Regressão e_{t2} versus X_{it})

TESTES DE HOMOGENEIDADE



- Teste 4 : Modelo ARCH sem restrições
- H_1 : Os resíduos seguem um modelo ARCH sem restrições nos coeficientes.

- Teste 5 : Modelo ARCH com restrição linear nos parâmetros
- H_1 : Os resíduos seguem um modelo ARCH cujos coeficientes decrescem linearmente com o lag de defasamento

TESTES DE HOMOGENEIDADE



- Teste 6 : Modelo ARCH com coeficientes iguais
- H_1 : Os resíduos seguem um modelo ARCH com coeficientes iguais.

- Teste 7 : Modelo ARCH Sazonal
- H_1 : Os resíduos seguem um modelo ARCH com coeficiente não nulo apenas no lag correspondente ao período Sazonal "S".