

Modelos Autorregressivos-Defasagens Distribuídas (ADL)

Pedro Valls¹
CEQEF e EESP-FGV

¹CEQEF e EESP-FGV, Rua Dr Plinio Barreto 365 s.1319, 01313-020, , São Paulo, SP, Brasil. E-mail:pedro.valls@fgv.br. © 2024 - Pedro Valls

Sumário

1	Especificação do Modelo	6
1.1	Relação de Equilíbrio com Elasticidade Constante	7
1.2	Reparametrização do Modelo ADL - Modelo de Correção de Erros (M.C.E.)	8
1.3	Equivalência entre estimadores, variância dos estimadores e resíduos de (1) e (10)	8
1.4	Equilíbrio Dinâmico	11
1.5	Modelo de Correção de Erros Homogeneo - Elasticidade Unitária	12
1.6	Modelo ADL Geral	13
2	Estimação e Testes	14
2.1	Exclusão de Informação Relevante	15
2.2	Inclusão de Informação Irrelevante	16
18	subsection.2.3	
2.3.1	Densidade Conjunta, condicional e marginal	18
2.3.2	Parâmetros de Interesse, Variação Livre e Exogeneidade Fraca	21
2.3.3	Causalidade e Exogeneidade Forte	23
2.3.4	Exogeneidade, Causalidade e Invariância	25
2.4	Testes de Exogeneidade	30
2.4.1	Testes de Exogeneidade de Durbin-Wu-Hausman	32
2.4.2	Testando exogeneidade fraca de INC para CONS em	
	DATAGIVE	35
2.5	Exercícios	36
3	Tipologia de Modelos Dinâmicos	38
3.1	Regressão Estática	39
3.1.1	Exemplo usando DATAGIVE.XLSX	40
3.2	Modelo Autorregressivo Univariado	42
3.2.1	Continuação do Exemplo	42
3.3	Indicadores Antecedentes	44
3.3.1	Continuação do Exemplo	44
3.4	Modelos de Taxa de Crescimento	47
3.4.1	Continuação do Exemplo	47
3.5	Defasagens Distribuídas	50

3.5.1	Continuação do Exemplo	50
3.6	Ajustamento Parcial	51
3.6.1	Continuação do Exemplo	52
3.7	Regressão Estática com Erros AR(1) ou Modelos de Fatores Comuns	54
3.7.1	Estimação com Erros Correlacionados	55
3.7.2	Testes para Autocorrelação - o teste de Durbin & Watson	58
3.7.3	Programa no R para gerar distribuição da estatística d de Durbin & Watson	59
3.7.4	Teste de Autocorrelação de quarta ordem	60
3.7.5	Teste h de Durbin	61
3.7.6	Teste de autocorrelação de Breusch-Godfrey	62
3.7.7	Continuação do Exemplo	63
3.8	Forma Reduzida	66
3.8.1	Continuação do Exemplo	66
3.9	Modelo de Correção de Erros Homogeneo	68
3.9.1	Exemplo - Cointegração	68
3.9.2	Estimação e Teste da equação de cointegração	69
3.10	Exemplo - DATAGIVE	70
3.11	Exercícios	82
4	Métodos de Seleção de Variáveis - Eliminação Reversa, Métodos de Penalização e Autometrics	85
4.1	Eliminação Reversa	85
4.2	Seleção Para Frente - <i>Forward Selection</i> (FWD)	86
4.3	Regressão de Menor Ângulo - <i>Least angle regression</i> - (LARS)	87
4.4	Métodos de Penalização	87
4.5	Avaliando todos os sub-modelos	89
4.6	Busca Multicaminho - <i>Multipath search</i> e Autometrics	90
5	Apêndice 1	98

Lista de Tabelas

1	Teste de Exogeneidade primeiro estágio usando (109)	35
2	Teste de Exogeneidade usando (111)	36

3	Regressão Estática	40
4	$AR(1)$	43
5	Indicador Antecedente	45
6	Taxa de Crescimento com $\Delta(\log(Cons_t))$ com variável dependente	48
7	Taxa de Crescimento com $\log(Cons_t)$ como variável dependente	48
8	Defasagens Distribuídas	50
9	Ajustamento Parcial	53
10	Regressão Estática com Erros $AR(1)$	64
11	Forma Reduzida	66
12	Valor Críticos para teste de raiz unitária dos resíduos do primeiro passo do procedimento de Engle and Granger (1987)	70
13	Matriz de Correlação entre Cons, Inc e Inflat e suas defasagens	71
14	$ADL(1,1)$	72
15	Elasticidade de Longo Prazo bem definida	74
16	$ADL(1,1,1)$	75
17	Omissão das variáveis $Inflat_t$ e $Inflat_{t-1}$	75
18	Elasticidade de Longo Prazo bem definida no $ADL(1,1,1)$ dado por (166)	77
19	Elasticidade de Longo Prazo bem definida no $ADL(1,1,1)$ dado por (166)	77
20	ADF Test for CI	78
21	Modelo EqCM dado por (168)	78
22	Teste de Autocorrelação dos Resíduos até ordem 5	80
23	Teste de Heteroscedasticidade Condicional dos Resíduos até ordem 4	80
24	Teste de Heteroscedasticidade Não Condicional dos Resíduos até ordem 4	80
25	Modelo Final para (173)	95
26	Tipologia de Modelos Dinâmicos	99

Lista de Figuras

1	Observado, Previsto e Intervalo de Confiança para (114)	41
2	Resíduos e Intervalo de Confiança para (114)	41

3	Observado, Previsto e Intervalo de Confian	
	ca para (116)	43
4	Resíduos e Intervalo de Confian	
	ca para (116)	44
5	Observado, Previsto e Intervalo de Confiança para (118) . . .	46
6	Resíduos e Intervalo de Confiança para (118)	46
7	Observado, Previsto e Intervalo de Confiança para (121) . . .	49
8	Resíduos e Intervalo de Confiança para (121)	49
9	Observado, Previsto e Intervalo de Confiança para (123) . . .	51
10	Resíduos e Intervalo de Confiança para (123)	51
11	Observado, Previsto e Intervalo de Confiança para (129) . . .	53
12	Resíduos e Intervalo de Confiança para (129)	54
13	Histograma da Estatística DW onde o P.G.D. é (145)	60
14	Observado, Previsto e Intervalo de Confiança para (159) e	
	(160)	65
15	Resíduos e Intervalo de Confiança para (159) e (160)	65
16	Observado, Previsto e Intervalo de Confiança para (162) . . .	67
17	Resíduos e Intervalo de Confiança para(162)	67
18	Cons, Inc e Inflat	71
19	Observado, Previsto e Intervalos de Confianças para (165) . .	73
20	Resíduos e Intervalos de Confianças para (165)	73
21	Observado, Previsto e Intervalos de Confianças para (166) . .	76
22	Resíduos e Intervalos de Confianças para (166)	76
23	Observado, Previsto e Intervalos de Confianças para (168) . .	79
24	Resíduos e Intervalos de Confianças para (168)	79
25	CUSUM para (168)	81
26	CUSUM para (168)	81
27	Coeficientes Recursivos USUM para (168)	82
28	Autometrics Arvore	92
29	Todas as observações	92
30	Dummies nas últimas observações	92
31	Dummies selecionadas nas últimas observações	93
32	Dummies nas primeiras observações	93
33	Dummies selecionadas nas primeiras observações	93
34	Junta as Dummies das duas partes	93
35	Dummies selecionadas	93
36	Observado e Previsto para o Modelo Final	96
37	Resíduos Recursivos para o Modelo Final	96

38	Estimativa Recursiva dos Coeficientes e teste de Chow para o Modelo Final	97
----	--	----

1 Especificação do Modelo

1

A especificação mais simples do modelo ADL é dada por:

$$y_t = m + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \quad (1)$$

e é denotado por $ADL(\mathbf{1}, \mathbf{1})$ onde o primeiro $\mathbf{1}$ representa o ordem do polinômio autorregressivo da variável dependente e o segundo $\mathbf{1}$ representa a ordem do polinômio no operador defasagem para a (única) variável independente.

Observe que no caso geral, de várias variáveis independentes o modelo seria denotado por $ADL(n_0, n_1, \dots, n_k)$ onde n_0 representaria a ordem da estrutura autorregressiva da variável dependente e n_i representa a ordem do polinômio no operador defasagem para a i -ésima variável independentes.

Quando a raiz do polinômio autorregressivo estiver dentro do círculo unitário, i.e. $|\alpha_1| < 1$, então podemos inverter o polinômio AR e podemos reescrever (1) da seguinte forma:

$$y_t = (1 + \alpha_1 + \alpha_1^2 + \dots)m + (1 + \alpha_1 L + \alpha_1^2 L^2 + \dots)(\beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t) \quad (2)$$

que expressa os valores correntes de y em relação aos valores correntes e passados de x .

Podemos também definir o impacto que mudanças nos valores correntes de x (x_t) terão em valores correntes de y (y_t) e também em valores futuros de y (y_{t+j} para $j = 1, 2, \dots$), i.e., os multiplicadores de impacto, que serão dados por:

$$\frac{\partial y_t}{\partial x_t} = \beta_0 \quad (3)$$

$$\frac{\partial y_{t+1}}{\partial x_t} = \beta_1 + \alpha_1 \beta_0$$

$$\frac{\partial y_{t+2}}{\partial x_t} = \alpha_1 \beta_1 + \alpha_1^2 \beta_0$$

Deste modo podemos definir as respostas (multiplicadores) de impacto ($\frac{\partial y_t}{\partial x_t}$), interim ($\frac{\partial y_{t+j}}{\partial x_t}$ para $j = 1, 2, \dots$) e de longo prazo ($\frac{\partial y_t}{\partial x_t} + \frac{\partial y_{t+1}}{\partial x_t} + \dots$).

¹Esta sessão é baseada em Hendry (1995) e Johnston and DiNardo (1996)

Por exemplo o multiplicador de impacto é definido por (3) e o de longo prazo pela soma de todas as derivadas parciais, isto é:

$$\begin{aligned}
& \beta_0 + (\beta_1 + \alpha_1\beta_0) + (\alpha_1\beta_1 + \alpha_1^2\beta_0) + (\alpha_1^2\beta_1 + \alpha_1^3\beta_0) + \dots \\
= & \beta_0(1 + \alpha_1 + \alpha_1^2 + \dots) + \beta_1(1 + \alpha_1 + \alpha_1^2 + \dots) \\
= & (\beta_0 + \beta_1)(1 + \alpha_1 + \alpha_1^2 + \dots) \\
= & \frac{\beta_1 + \beta_0}{1 - \alpha_1}
\end{aligned} \tag{4}$$

Observe que (4) corresponde ao coeficiente da variável independente na solução de longo prazo de (1) uma vez que aplicando o operador esperança matemática em (1) temos:

$$E(y_t) = m + \alpha_1 E(y_{t-1}) + \beta_0 E(x_t) + \beta_1 E(x_{t-1}) \tag{5}$$

e como no longo-prazo $E(y_t) = E(y_{t-1}) = \bar{y}$ e $E(x_t) = E(x_{t-1}) = \bar{x}$ então podemos escrever (5) da seguinte forma:

$$\bar{y} = \frac{m}{1 - \alpha_1} + \frac{\beta_1 + \beta_0}{1 - \alpha_1} \bar{x} \tag{6}$$

A expressão (6) é chamada de **solução de longo-prazo** ou **equilíbrio estático**.

1.1 Relação de Equilíbrio com Elasticidade Constante

Se em (1) as variáveis y e x representam o logaritmo de Y e X , a expressão (6) representa uma relação de equilíbrio com elasticidade constante. Aplicando a transformação exponencial em (6) temos

$$Y = AX^\gamma \tag{7}$$

onde

$$a = \ln(A) = \frac{m}{1 - \alpha_1} \tag{8}$$

$$\gamma = \frac{\beta_1 + \beta_0}{1 - \alpha_1} \tag{9}$$

1.2 Reparametrização do Modelo ADL - Modelo de Correção de Erros (M.C.E.)

Considere o modelo dado em (1), isto é:

$$y_t = m + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

e como

$$x_t = \Delta x_t + x_{t-1}$$

$$y_t = \Delta y_t + y_{t-1}$$

temos que (1) pode ser re-escrito da seguinte forma:

$$\Delta y_t + y_{t-1} = m + \alpha_1 y_{t-1} + \beta_0 (\Delta x_t + x_{t-1}) + \beta_1 x_{t-1} + \varepsilon_t$$

$$\Delta y_t = m + \beta_0 \Delta x_t + (\alpha_1 - 1)y_{t-1} + (\beta_0 + \beta_1)x_{t-1} + \varepsilon_t \quad (10)$$

$$\Delta y_t = \beta_0 \Delta x_t - (1 - \alpha_1)[y_{t-1} - a - \gamma x_{t-1}] + \varepsilon_t \quad (11)$$

Esta especificação é conhecida por **Modelo de Correção de Erros**, onde as variações corrente na variável dependente são explicadas pela soma de dois componentes: o primeiro é a variação corrente na variável independente e o segundo um termo de correção parcial dos desvios de y_{t-1} da sua relação de equilíbrio dada por $a + \gamma x_{t-1}$. Este desequilíbrio ou desvio (erro) do equilíbrio se positivo a correção no período corrente será para baixo e se negativo será para cima uma vez que $(1 - \alpha_1) < 0$.

O modelo (10) pode ser estimado por MQO através da regressão de Δy_t em constante, Δx_t , y_{t-1} e x_{t-1} . Através dos coeficientes estimados e da matriz de variância e covariância das estimativas é possível obter os quatro parâmetros e seus respectivos desvios padrões de (1).

1.3 Equivalência entre estimadores, variância dos estimadores e resíduos de (1) e (10)

O modelo (1) escrito em forma matricial é dado por:

$$y = X\beta + \varepsilon \quad (12)$$

com

$$X = [i, y_{-1}, x, x_{-1}] \quad (13)$$

onde i é um vetor de uns e os coeficientes são agrupados no vetor β , i.e.

$$\beta' = [m, \alpha_1, \beta_0, \beta_1] \quad (14)$$

E o modelo (10) escrito em forma matricial é dado por:

$$\Delta y = Z\gamma + \varepsilon \quad (15)$$

onde

$$Z = [i, y_{-1}, x - x_{-1}, x_{-1}] \quad (16)$$

e

$$\gamma' = [m, \alpha_1 - 1, \beta_0, \beta_0 + \beta_1] \quad (17)$$

Observe que existe a seguinte relação entre β e γ

$$\gamma = A^{-1} \left(\beta - \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right) \quad (18)$$

onde

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

e entre Z e X

$$Z = XA \quad (19)$$

O estimador de γ em (15) é dado por:

$$\hat{\gamma} = (Z'Z)^{-1}Z'\Delta y = (Z'Z)^{-1}Z'y - (Z'Z)^{-1}Z'y_{-1} \quad (20)$$

$$\hat{\gamma} = (AXXA)^{-1}AX'y - (ZZ)^{-1}Zy_{-1} \quad (21)$$

Agora a primeira expressão do lado direito de (21) é dada por:

$$\begin{aligned} (AXXA)^{-1}AX'y &= A^{-1}(XX)^{-1}(A)^{-1}AX'y \\ &= A^{-1}(XX)^{-1}Xy \\ &= A^{-1}\hat{\beta} \end{aligned} \quad (22)$$

para a segunda expressão observe que

$$I = (ZZ)^{-1}(ZZ) = (ZZ)^{-1}Z' [i, y_{-1}, x - x_{-1}, x_{-1}]$$

o que implica que

$$(ZZ)^{-1}Zy_{-1} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad (23)$$

Agora multiplicando (23) por A^{-1} , temos:

$$A^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad (24)$$

juntando (22) e (25) temos o resultado.

Os resíduos do modelo (12) são dados por:

$$\hat{\varepsilon}_X = M_X \varepsilon \quad (25)$$

e do modelo (15) são dados por:

$$\hat{\varepsilon}_Z = M_Z \varepsilon \quad (26)$$

Mas

$$\begin{aligned}
M_Z &= I - Z(Z'Z)^{-1}Z' \\
&= I - XA(A'X'XA)^{-1}A'X' \\
&= I - XAA^{-1}(X'X)^{-1}(A')^{-1}A'X' \\
&= I - X(X'X)^{-1}X' = M_X
\end{aligned}$$

logo (25) e (26) são idênticos implicando também que o estimador da variância residual é o mesmo nos dois modelos.

Agora por (18) a variância do estimador do modelo (12) é dada por:

$$\begin{aligned}
Var(\hat{\beta}) &= Var\left(\hat{\gamma} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}\right) \\
&= AVar(\hat{\gamma})A' \\
&\quad s^2 A(Z'Z)^{-1}A' \\
&= s^2 A(A'X'XA)^{-1}A' \\
&= s^2 A(A^{-1})(X'X)^{-1}(A')^{-1}A' \\
&= s^2 (X'X)^{-1}
\end{aligned}$$

que implica que a variância dos dois estimadores é igual.

1.4 Equilíbrio Dinâmico

A expressão (6) foi obtida assumindo equilíbrio estático. Vamos agora assumir que X cresce a uma taxa constante g_X tal que $\Delta x_t = g_X$ para todo t . Assumindo elasticidade constante e igual a γ a taxa de crescimento de Y será dada por $g_Y = \gamma g_X$.

Substituindo-se estas expressões em (11) obtemos:

$$\gamma g_X = \beta_0 g_X - (1 - \alpha_1)\left[y - \frac{m}{1 - \alpha_1} - \gamma x\right]$$

\Downarrow

$$y = \frac{m - (\gamma - \beta_0)g_X}{1 - \alpha_1} + \gamma x \quad (27)$$

e como $y = \ln(Y)$ e $x = \ln(X)$ temos que (27) pode ser escrito da seguinte forma:

$$Y = AX^\gamma \quad (28)$$

com

$$A = \exp\left(\frac{m - (\gamma - \beta_0)g_X}{1 - \alpha_1}\right) \quad (29)$$

Observe que a diferença entre a relação de equilíbrio estática e a dinâmica é a constante A dada acima e em (8).

1.5 Modelo de Correção de Erros Homogeneo - Elasticidade Unitária

No modelo (28) a razão de equilíbrio Y/X varia com o nível de X uma vez que

$$\frac{Y}{X} \sim X^{\gamma-1} \quad (30)$$

Se a elasticidade γ for positiva mas menor do que um a razão de equilíbrio tenderá a zero quando X crescer, e por outro lado se for maior do que um a razão de equilíbrio crescerá sem limites.

Em várias relações economicas tanto a tendência para zero quando para o infinito é pouco provável, sendo mais razoável a hipótese de elasticidade unitária. Observe que elasticidade unitária é equivalente a

$$\gamma = 1 \Leftrightarrow \frac{\beta_1 + \beta_0}{1 - \alpha_1} = 1 \Leftrightarrow \beta_1 + \beta_0 + \alpha_1 = 1 \quad (31)$$

e se esta hipótese for verdadeira o modelo de correção de erros é do tipo homogêneo, isto é

$$\Delta y_t = m + \beta_0 \Delta x_t + (\alpha_1 - 1)(y_{t-1} - x_{t-1}) + \varepsilon_t \quad (32)$$

Observe que esta hipótese pode ser testada das seguintes formas:

- (i) usando o modelo (1) e testando a homogeneidade dos coeficientes, isto é, se a soma dos coeficientes de y_{t-1} , x_t e x_{t-1} somam um;

- (ii) usando o modelo (10) e testando se a soma dos coeficientes de y_{t-1} e x_{t-1} somam um; e
- (iii) usando o modelo (32) mas com a inclusão do termo x_{t-1} e testando se este coeficiente é igual a zero.

A racionalidade para (iii) acima vem do fato de que (10) pode ser reparametrizado da seguinte forma:

$$\Delta y_t = m + \beta_0 \Delta x_t + (\alpha_1 - 1)(y_{t-1} - x_{t-1}) + (\beta_0 + \beta_1 + \alpha_1 - 1)x_{t-1} + \varepsilon_t \quad (33)$$

logo testar a significância do coeficiente de x_{t-1} é equivalente a testar elasticidade unitária.

1.6 Modelo ADL Geral

O modelo ADL(p,q), assumindo uma única variável explanatória, é dado por:

$$A(L)y_t = m + B(L)x_t + \varepsilon_t \quad (34)$$

onde

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$$

e

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q$$

e assumindo que as variáveis estão em logaritmo temos que a elasticidade de longo-prazo é dada por

$$\gamma = \frac{B(1)}{A(1)} = \frac{\beta_0 + \beta_1 + \beta_2 + \dots + \beta_q}{1 - \alpha_1 - \alpha_2 - \dots - \alpha_p}$$

Para obter a representação MCE é necessário que $p = q$ e neste caso (34) pode ser reescrita da seguinte forma:

$$\begin{aligned}
\Delta y_t + y_{t-1} = & m + \alpha_1 y_{t-1} + \alpha_2 (y_{t-1} - \Delta y_{t-1}) + \\
& + \alpha_3 (y_{t-1} - \Delta y_{t-1} - \Delta y_{t-2}) + \dots + \\
& + \alpha_p (y_{t-1} - \Delta y_{t-1} - \dots - \Delta y_{t-p+1}) + \\
& + \beta_0 (\Delta x_t + x_{t-1}) + \beta_1 x_{t-1} + \beta_2 (x_{t-1} - \Delta x_{t-1}) + \\
& + \beta_3 (x_{t-1} - \Delta x_{t-1} - \Delta x_{t-2}) + \dots + \\
& + \beta_p (x_{t-1} - \Delta x_{t-1} - \dots - \Delta x_{t-p+1}) + \varepsilon_t
\end{aligned} \tag{35}$$

e pode ser escrita da seguinte forma:

$$\begin{aligned}
\Delta y_t = & m + \sum_{j=1}^{p-1} \pi_j \Delta y_{t-j} + \sum_{j=1}^{p-1} \delta_j \Delta x_{t-j} \\
& - (1 - \alpha_1 - \alpha_2 - \dots - \alpha_p) (y_{t-1} - \gamma x_{t-1}) + \varepsilon_t
\end{aligned} \tag{36}$$

onde

$$\begin{aligned}
\pi_j = & - \sum_{k=j+1}^p \alpha_k \quad \text{para } j = 1, \dots, p-1 \\
\delta_j = & - \sum_{k=j+1}^p \beta_k \quad \text{para } j = 1, \dots, p-1
\end{aligned}$$

Observe que (34) só inclui uma variável, que é um caso particular do modelo geral com k variáveis que pode ser representado por:

$$A(L)y_t = m + \sum_{i=1}^k B_i(L)x_{it} + \varepsilon_t$$

2 Estimação e Testes

Dada a especificação do modelo ADL a primeira pergunta que se apresenta é como estimar e testar hipóteses nestes modelos.

Dado que o modelo ADL é de uma única equação esta simplificação pode implicar em perda de generalidade. Mas como veremos um pouco mais a

frente esta simplificação, sob certas condições (exogeneidade fraca), não implica em perda de informação relevante.

A questão prática que se coloca é relacionada a quais e quantas variáveis incluir no modelo e também quantas defasagens usar.

Um procedimento que era usado no passado era começar de um modelo simples, na maioria das vezes estático, e caso a parte não sistemática, isto é, os erros não fossem um ruído branco modificava-se o modelo, ou através de correção por autocorrelação, ou por inclusão de novas variáveis, ou por inclusão de dinâmica, até que as hipóteses sobre a parte não sistemática fossem alcançadas. Este procedimento é chamado de **S**imples **P**ara **G**eral, denotado por **SPGE**, e em inglês, **S**imple **T**o **G**eneral (**STGE**).

Este tipo de procedimento foi e, é criticado uma vez que excluir informação relevante, ou por omissão de variáveis ou de dinâmica, implica em perda de eficiência e em algumas situações em perda de consistência dos estimadores dos parâmetros.

O exemplo a seguir ilustra este fato

2.1 Exclusão de Informação Relevante

Suponha que é estimado o seguinte modelo:

$$y_t = x_t\beta + u_t \quad (37)$$

mas o modelo verdadeiro é:

$$y_t = x_t\beta + \gamma z_t + \varepsilon_t \quad (38)$$

O estimador de β em (37) é dado por:

$$\begin{aligned} \hat{\beta} &= \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t (x_t \beta + \gamma z_t + \varepsilon_t)}{\sum x_t^2} \\ &= \beta + \gamma \frac{\sum x_t z_t}{\sum x_t^2} + \frac{\sum x_t \varepsilon_t}{\sum x_t^2} \\ &= \beta + \gamma b_{xz} \\ \implies E(\hat{\beta}) &= \beta + \gamma b_{xz} \end{aligned} \quad (39)$$

onde em (39) foi assumido que x_t é não correlacionado com ε_t e que b_{xz} é o coeficiente de regressão de z_t em x_t . Logo $\hat{\beta}$ é um estimador viesado e inconsistente para β exceto quando $b_{xz} = 0$, isto é quando z_t não correlacionado com x_t .

Por outro lado a variância do estimador do modelo (39) é dada por

$$Var(\hat{\beta}) = \frac{\sigma_\varepsilon^2}{\sum x_t^2} \quad (40)$$

mas a pseudo parte não sistemática de (37) é dada por

$$u_t = \gamma z_t + \varepsilon_t \quad (41)$$

então o estimador da variância residual, s^2 , usando (37) tenderá a superestimar σ_ε^2 .

Isto mostra que a omissão de informação relevante faz com que o estimador do parâmetro seja inconsistente e como o estimador da variância residual superestima a variância a inferência a respeito dos parâmetros será inválida.

O exemplo a seguir ilustra o caso em que informação desnecessária é incluída.

2.2 Inclusão de Informação Irrelevante

Suponha que o modelo verdadeiro é (37) mas o modelo estimado é (38), isto é incluímos informação desnecessária.

Os estimadores de MQO dos parâmetros são dados por:

$$\hat{\beta} = (X'M_Z X)^{-1} X'M_Z y \quad (42)$$

onde $M_Z = I - Z(Z'Z)^{-1}Z'$ e

$$\hat{\gamma} = (Z'M_X Z)^{-1} Z'M_X y \quad (43)$$

onde $M_X = I - X(X'X)^{-1}X'$

Como o modelo verdadeiro é (37), temos que (42) pode ser re-escrito da seguinte forma:

$$\begin{aligned} \hat{\beta} &= (X'M_Z X)^{-1} X'M_Z (X\beta + u) \\ &= \beta + (X'M_Z X)^{-1} X'M_Z u \end{aligned} \quad (44)$$

Agora aplicando o operador esperança em (44) temos que

$$E(\hat{\beta}) = \beta$$

logo $\hat{\beta}$ é um estimador não viesado para β .

Como o modelo verdadeiro é (37), temos que (43) pode ser re-escrito da seguinte forma:

$$\begin{aligned}\hat{\gamma} &= (Z'M_X Z)^{-1} Z'M_X (X\beta + u) \\ &= (Z'M_X Z)^{-1} Z'M_X X\beta + (Z'M_X Z)^{-1} Z'M_X u \\ \implies E(\hat{\gamma}) &= 0\end{aligned}$$

Observe que o resíduo do modelo (38) é dado por:

$$\begin{aligned}\hat{\varepsilon} &= y - W\hat{\delta} = y - W(W'W)^{-1}W'y \\ &= [I - W(W'W)^{-1}W'] y \\ &= [I - P_W] y = M_W y\end{aligned}\tag{45}$$

onde $W = [X : Z]$.

Assumindo que o modelo verdadeiro é (37) e como $P_W X = X^2$ temos que (45) reduz-se a:

$$\hat{\varepsilon} = M_W (X\beta + u) = M_W u\tag{46}$$

e a soma dos quadrados dos resíduos é dada por:

$$\hat{\varepsilon}'\hat{\varepsilon} = u' M_W u\tag{47}$$

e como $u \sim N(0, \sigma_u^2 I_T)$ e M_W é uma matriz idempotente com posto $T - k_1 - k_2$ onde k_1 é a dimensão do vetor β e k_2 é a dimensão do vetor γ temos que $E(u' M_W u) = \sigma_u^2 (T - k_1 - k_2)$ logo $\frac{\hat{\varepsilon}'\hat{\varepsilon}}{T - k_1 - k_2}$ é um estimador não viesado para σ_u^2 .

²Intuitivamente P_W gera o espaço formado por W logo $P_W W = W \Leftrightarrow P_W [X : Z] = [X : Z] \implies P_W X = X$.

2.3 Conceito de Exogeneidade³

O ponto de partida para a metodologia denominada na literatura por "LSE Methodology" (veja Gilbert (1989), Juselius (1999) ou Ericsson (2004-08)) é a distribuição conjunta das observações e através de condicionamentos e marginalizações é possível simplificar o processo gerador dos dados. Onde estas idéias foram apresentadas pela primeira vez por Haavelmo (1943).

Observe que, em geral, marginalizações e condicionamentos implicam em perda de informação em relação a distribuição conjunta.

Nesta mesma modelagem o conceito de exogeneidade de uma variável em relação a um conjunto de parâmetros é fundamental para ser possível fazer inferência válida no modelo condicional, isto é a perda de informação em se descartar o modelo marginal é irrelevante.

Observe que nesta metodologia a modelagem deve ser feita do **GE**ral **PA**ra o **S**imples ou particular (**GEPS** e em inglês **GE**neral **T**o **S**pecific - **GETS**), isto é, o ponto de partida deve ser um modelo que contemple dinâmica e com uma ordem de defasagem grande o suficiente para incluir a defasagem verdadeira, que é desconhecida, e isto é feito para que não se incorra em omissão de informação relevante, que é o caso do Simples para o Geral (em inglês **STGE**). Através de testes de especificação e falta de especificação simplificasse o modelo.

O exemplo abaixo ilustra o caso em que podemos descartar a distribuição marginal sem perda de informação relevante.

2.3.1 Densidade Conjunta, condicional e marginal

Seja $z_t = (y_t, x_t)$ um processo estocástico serialmente não e gaussiano que será denotado por:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} \sim NI(\mu, \Omega) \text{ para } t = 1, \dots, T \quad (48)$$

Por exemplo y_t e x_t podem representar moeda e taxa de juros, salários e preços etc. Defina ε_t a parte não sistemática do modelo definida por:

$$\varepsilon_t = \mathbf{z}_t - E(\mathbf{z}_t) \quad (49)$$

logo por construção temos que

³Esta sessão é inspirada em Ericsson (1992), Hendry (1995) e Hendry and Nielsen (2007).

$$\varepsilon_t \sim NI(0, \Omega) \quad (50)$$

Agora (48) pode ser re-escrito da seguinte forma:

$$\mathbf{z}_t = \mu + \varepsilon_t \quad (51)$$

que representa o mecanismo estatístico gerador ou na nomenclatura usual (51) está na “ forma de um modelo ” .

A seguir será necessário particionar tanto o vetor de médias μ quanto a matriz de covariância Ω conforme a partição do vetor \mathbf{z}_t , isto é,

$$\mu = \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix} \text{ e } \Omega = \begin{bmatrix} \omega_{yy} & \omega_{yx} \\ \omega_{xy} & \omega_{xx} \end{bmatrix} \quad (52)$$

Sem perda de generalidade e informação, a distribuição conjunta dada por (48) pode ser fatorada na distribuição de y_t condicional em x_t e na marginal de x_t da seguinte forma:

$$y_t \mid x_t \sim NI(\alpha + \beta x_t, \sigma^2) \quad (53)$$

$$x_t \sim NI(\mu_x, \omega_{xx}) \quad (54)$$

onde $\beta = \frac{\omega_{yx}}{\omega_{xx}}$, $\alpha = \mu_y - \beta\mu_x$ e $\sigma^2 = \omega_{yy} - \frac{\omega_{yx}^2}{\omega_{xx}}$. Então a ”forma de modelo ” de (53-54) é dada por:

$$y_t = \alpha + \beta x_t + \nu_{1t} \text{ com } \nu_{1t} \sim NI(0, \sigma^2) \quad (55)$$

$$x_t = \mu_x + \varepsilon_{2t} \text{ com } \varepsilon_{2t} \sim NI(0, \omega_{xx}) \quad (56)$$

onde $\nu_{1t} = \varepsilon_{1t} - \left(\frac{\omega_{yx}}{\omega_{xx}}\right) \varepsilon_{2t} = y_t - E(y_t \mid x_t)$ e este componente contém parte da informação de y_t que é não correlacionada com x_t porque:

- (i) $E(\nu_{1t} \mid x_t) = E(y_t - E(y_t \mid x_t) \mid x_t) = E(y_t \mid x_t) - E(E(y_t \mid x_t)) = 0$;
- (ii) $E(\nu_{1t} x_t) = E(E(\nu_{1t} x_t \mid x_t)) = E(x_t E(\nu_{1t} \mid x_t)) = E(x_t \cdot 0) = 0$

e também é não correlacionado com ε_{2t} porque:

$$\begin{aligned}
E(\nu_{1t}\varepsilon_{2t}) &= E\left(\left(\varepsilon_{1t} - \left(\frac{\omega_{yx}}{\omega_{xx}}\right)\varepsilon_{2t}\right)\varepsilon_{2t}\right) \\
&= E(\varepsilon_{1t}\varepsilon_{2t}) - \left(\frac{\omega_{yx}}{\omega_{xx}}\right)E(\varepsilon_{2t}^2) \\
&= \omega_{yx} - \left(\frac{\omega_{yx}}{\omega_{xx}}\right)\omega_{xx} \\
&= 0
\end{aligned}$$

Observe que a relação entre (51) e (53-54) é dada por:

$$D_Z(z_t; \theta) = D_{Y|X}(y_t \mid x_t; \lambda_1) D_X(x_t; \lambda_2) \quad (57)$$

onde $\theta = (\mu, \Omega) \in \Theta \subseteq \mathbb{R}^n$ representa o conjunto de (todos) parâmetros da distribuição conjunta, $\lambda_1 = (\alpha, \beta, \sigma^2) \in \Lambda_1 \subseteq \mathbb{R}^{n_1}$ representa os parâmetros da distribuição condicional e $\lambda_2 = (\mu_x, \omega_{xx}) \in \Lambda_2 \subseteq \mathbb{R}^{n_2}$ os parâmetros da distribuição marginal e temos que $n_1 + n_2 = n$. Observe também que $\lambda = (\lambda_1, \lambda_2) = f(\theta)$ é uma transformação um-a-um que sempre pode ser definida embora os parâmetros λ_1 e λ_2 possam estar relacionados.

Desta forma em (57) não há perda de informação, mas ao se analisar o modelo condicional ignorando o modelo marginal tem-se perda de informação.

Deste modo modelar somente o modelo condicional, ignorando a informação que possa existir entre os parâmetros do modelo condicional e marginal, por exemplo, restrições entre equações, há perda de informação. Observe que esta dependência está presente na própria definição dos parâmetros α, β, σ^2 .

A hipótese de não dependência entre λ_1 e λ_2 é, em geral, uma condição muito forte para que inferência seja válida. Vamos apresentar um outro conceito, corte sequencial, da função densidade que é mais fraco do que a "não dependência".

Definição 1: Dizemos que (57) opera um corte sequencial se e somente se λ_1 e λ_2 são variação livre, isto é $(\lambda_1, \lambda_2) \in \Lambda_1 \times \Lambda_2$.

Observe que esta definição implica que o espaço paramétrico Λ_1 não é função do parâmetro λ_2 e nem vice-versa. De uma forma intuitiva, podemos dizer que o conhecimento do valor de um dos parâmetros não nos dá informação sobre o intervalo dos possíveis valores que o outro parâmetro pode assumir.

Observe que se o parâmetro de interesse for μ , o vetor de médias da distribuição conjunta, devemos estimar tanto o modelo marginal quanto o modelo condicional para poder recuperar tal parâmetro. É portanto fundamental deixar claro quem são os parâmetros de interesse, que denotaremos por $\psi \in \mathbb{R}^q$ com $q < n$. Tem-se então a seguinte definição de exogeneidade fraca.

Definição 2: A variável x_t é exogena fraca para os parâmetros de interesse ψ se e somente se:

- (i) $\psi = \psi(\lambda_1)$, isto é ψ é função somente de λ_1 ;
- (ii) (57) opera um corte sequencial, isto é, λ_1 e λ_2 são variação livre.

A condição (i) acima diz que os parâmetros de interesse não dependem de λ_2 direta ou indiretamente, logo nenhuma informação sobre os parâmetros de interesse pode ser obtida através do modelo marginal.

Deste modo os parâmetros de interesse são recuperados unicamente e completamente através do modelo condicional e portanto exogeneidade fraca é uma condição suficiente para inferência válida (estimação e testes eficientes) de ψ através do modelo condicional.

Dado que os parâmetros θ são constantes, isto é $\theta_t = \theta \forall t$ as definições acima podem ser usadas para toda a amostra, e usando-se a fatorização sequencial temos que a densidade conjunta é dada por:

$$D_Z(Z_1^T; \theta) = \prod_{t=1}^T D_Z(y_t, x_t; \theta) = \prod_{t=1}^T D_{Y|X}(y_t | x_t; \lambda_1) \prod_{t=1}^T D_X(x_t; \lambda_2) \quad (58)$$

e em (58) podemos descartar a distribuição marginal de x_t .

O exemplo a seguir aborda os conceitos de parâmetros de interesse, variação livre e exogeneidade fraca.

2.3.2 Parâmetros de Interesse, Variação Livre e Exogenidade Fraca

Consider o modelo (55-56) mas com a média de x_t agora sendo uma função linear de y_{t-1} , isto é:

$$y_t = \alpha + \beta x_t + \nu_{1t} \text{ com } \nu_{1t} \sim NI(0, \sigma^2) \quad (59)$$

$$x_t = \kappa y_{t-1} + \varepsilon_{2t} \text{ com } \varepsilon_{2t} \sim NI(0, \omega_{xx}) \quad (60)$$

Este modelo é conhecido na literatura por modelo de Cobwell se $y_t = \ln(P_t) = p_t$ representa ao preço de um produto agrícola e $x_t = \ln(Q_t) = q_t$ representa a quantidade produzida deste produto. Tem-se então, assumindo que $\alpha = 0$,

$$p_t = \beta q_t + \nu_{1t} \text{ com } \nu_{1t} \sim NI(0, \sigma^2) \quad (61)$$

$$q_t = \kappa p_{t-1} + \varepsilon_{2t} \text{ com } \varepsilon_{2t} \sim NI(0, \omega_{qq}) \quad (62)$$

e como antes temos $E(q_t \nu_{1t}) = E(\varepsilon_{2t} \nu_{1t}) = 0$.

Observe que (61) representa uma equação de demanda, isto é, o preço de equilíbrio do mercado para uma dada quantidade q_t . O valor $\frac{1}{\beta}$ é, portanto, a elasticidade preço de demanda.

Observe que (62) é uma equação de oferta, isto é, a produção hoje depende do preço recebido no período anterior. O valor de κ é, portanto, a elasticidade preço de oferta.

Em algumas situações a “estabilidade” do sistema é de interesse e temos que “resolver” o sistema acima de tal sorte que preço e quantidade dependam somente de preço e quantidade em instante anteriores. Isto é obtido reescrevendo o sistema (61-62) da seguinte forma:⁴

$$\begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_t \\ q_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \kappa & 0 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ q_{t-1} \end{bmatrix} + \begin{bmatrix} \nu_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (63)$$

agora invertendo a matriz do lado direito de (63) temos:⁵

$$\begin{bmatrix} p_t \\ q_t \end{bmatrix} = \begin{bmatrix} \beta\kappa & 0 \\ \kappa & 0 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ q_{t-1} \end{bmatrix} + \begin{bmatrix} \nu_{1t} + \beta\varepsilon_{2t} \\ \varepsilon_{2t} \end{bmatrix} \quad (64)$$

que é equivalente a:

$$p_t = \rho p_{t-1} + \varepsilon_{1t} \text{ onde } \rho = \beta\kappa \text{ e } \varepsilon_{1t} \sim NI(0, \omega_p) \quad (65)$$

$$q_t = \kappa p_{t-1} + \varepsilon_{2t} \text{ onde } \varepsilon_{2t} \sim NI(0, \omega_{qq}) \quad (66)$$

e para que o sistema seja estável devemos ter $|\rho| < 1$.

⁴A especificação (63) é conhecida como Forma Estrutural do Modelo de Equações Smultâneas Dinâmico.

⁵A especificação (64) é conhecida como Forma Reduzida do Modelo de Equações Smultâneas Dinâmico.

A seguir ficará claro como os parâmetros de interesse e o espaço paramétrico afetam o “*status*” da exogeneidade fraca de q_t em (64).

Observe que $\lambda_1 = (\beta, \sigma^2)$ e $\lambda_2 = (\kappa, \omega_{qq})$. A primeira condição para exogeneidade fraca exige que ψ seja função somente de λ_1 .⁶

Observe que se $\frac{1}{\beta}$, a elasticidade de demanda, for o parâmetro de interesse, então a primeira condição de exogeneidade fraca é satisfeita. Por outro lado se ρ for o parâmetro de interesse, isto é, estamos interessados na estabilidade do sistema, então para estimar ρ necessitamos de $\beta \in \lambda_1$ e $\kappa \in \lambda_2$, isto é, necessitamos de todo o sistema e portanto q_t não é exógeno fraco para ρ , mas pode ser exógeno fraco para $\frac{1}{\beta}$ se a segunda condição for satisfeita.

Suponha que β e κ são irrestritos, isto é, $(\beta, \kappa) \in \lambda_1 \times \lambda_2 \subseteq \Re \times \Re = \Re^2$. Então λ_1 e λ_2 são variação livre e, portanto, q_t é exógeno fraco para $\frac{1}{\beta}$.

Agora suponha que o sistema seja estável, isto é, $|\rho| < 1 \Leftrightarrow |\beta\kappa| < 1$ logo β e κ são restritos. Neste caso o valor de κ afeta o intervalo de valores possíveis para β . Por exemplo, se $\kappa = 0,5$ então $\beta \in (-2, 2)$ e se $\kappa = 0,2$ então $\beta \in (-5, 5)$. Logo λ_1 e λ_2 não são variação livre e portanto q_t não é exógeno fraco para $\frac{1}{\beta}$.

Observe também que o espaço paramétrico Λ não é $\Lambda_1 \times \Lambda_2$ porque quando $\kappa = 0,2$ temos que $\Lambda_1 = (-5, 5)$ e quando $\beta = 1 \in \Lambda_1$ temos que $\Lambda_2 = (-1, 1) \ni \kappa = 0,2$. Deste modo $\Lambda_1 \times \Lambda_2 = (-5, 5) \times (-1, 1)$ não é um subespaço de Λ porque este é definido por $\Lambda = \{(\beta, \kappa) \mid |\beta\kappa| < 1\}$. Logo q_t não é exógeno fraco para $\frac{1}{\beta}$.

Suponha que $\kappa \in [0, 1)$ e que $\frac{1}{\beta}$ é negativo e maior do que 1, isto é, $\beta \in [-1, 0)$. Agora $\Lambda_1 \times \Lambda_2$ é um subespaço de Λ e Λ_1, Λ_2 são variação livre. Portanto sob estas restrições nas elasticidades q_t é exógeno fraco para $\frac{1}{\beta}$.

Vamos considerar um terceiro exemplo em que as densidades marginal, condicional e conjunta contemplam defasagens de modo que os conceitos de causalidade e exogeneidade possam ser definidos e comparados.

2.3.3 Causalidade e Exogeneidade Forte

Seja dado o seguinte processo gerador dos dados:

$$\mathbf{z}_t = \Pi \mathbf{z}_{t-1} + \varepsilon_t \quad \varepsilon_t \sim NI(0, \Omega) \quad (67)$$

então os modelos condicional e marginal são dados por:

⁶Primeiramente não consideraremos σ^2 e ω_{qq} .

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 y_{t-1} + v_{1t} \quad v_{1t} \sim NI(0, \sigma^2) \quad (68)$$

$$x_t = \pi_{22} x_{t-1} + \pi_{21} y_{t-1} + \varepsilon_{2t} \quad \varepsilon_{2t} \sim NI(0, \omega_{xx}) \quad (69)$$

Observe que por (67) temos que:

$$E(\mathbf{z}_t \mid \mathbf{Z}_{t-1}) = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} \quad (70)$$

$$Var(\mathbf{z}_t \mid \mathbf{Z}_{t-1}) = \Omega = \begin{bmatrix} \omega_{yy} & \omega_{yx} \\ \omega_{xy} & \omega_{xx} \end{bmatrix} \quad (71)$$

e através das propriedades de normal multivariada temos:

$$\begin{aligned} E(y_t \mid x_t, \mathbf{Z}_{t-1}) &= (\mu_{y|Z} - \omega_{yx}\omega_{xx}^{-1}\mu_{x|Z}) + \omega_{yx}\omega_{xx}^{-1}x_t \\ &= (\pi_{11}y_{t-1} + \pi_{12}x_{t-1} - \omega_{yx}\omega_{xx}^{-1}(\pi_{21}y_{t-1} + \pi_{22}x_{t-1})) + \omega_{yx}\omega_{xx}^{-1}x_t \\ &= \omega_{yx}\omega_{xx}^{-1}x_t + (\pi_{12} - \omega_{yx}\omega_{xx}^{-1}\pi_{22})x_{t-1} \\ &\quad + (\pi_{11} - \omega_{yx}\omega_{xx}^{-1}\pi_{21})y_{t-1} \end{aligned} \quad (72)$$

\Updownarrow

$$E(y_t \mid x_t, \mathbf{Z}_{t-1}) = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 y_{t-1} \quad (73)$$

onde $\beta_0 = \omega_{yx}\omega_{xx}^{-1}$, $\beta_1 = (\pi_{12} - \omega_{yx}\omega_{xx}^{-1}\pi_{22})$ e $\beta_2 = (\pi_{11} - \omega_{yx}\omega_{xx}^{-1}\pi_{21})$ e

$$Var(y_t \mid x_t, \mathbf{Z}_{t-1}) = \omega_{yy} - \omega_{yx}\omega_{xx}^{-1}\omega_{xy} = \sigma^2 \quad (74)$$

Suponha que $(\beta_0, \beta_1, \beta_2, \sigma^2) \in \Lambda_1$ são os parâmetros de interesse. Observe que os parâmetros do modelo marginal são $(\pi_{21}, \pi_{22}, \omega_{xx}) \in \Lambda_2$. Se b_i para $i = 0, 1, 2$ forem irrestritos então $\Lambda_1 \times \Lambda_2$ é variação livre. Portanto x_t é exógeno fraco para os parâmetros de interesse.

Mas se desejarmos fazer previsões l -passos à frente para y_t usando o modelo condicional (68), a exogeneidade fraca não será suficiente porque y_{t-1} influencia x_t no modelo marginal se $\pi_{21} \neq 0$ e, portanto, para fazer previsões para y_t condicional nas previsões de x_t não podemos assumir que x_t seja fixo no modelo condicional uma vez que existe *feedback* entre y_t e x_t que deve ser levado em consideração nestas previsões.

Deste modo a condição de exogeneidade fraca continua valendo mas, para se fazer previsões é necessário considerar tanto o modelo condicional quanto o marginal.

Mas se $\pi_{21} = 0$ não existe *feedback* entre y_t e x_t e, portanto, é possível fazer previsões para y_t usando somente o modelo condicional (68).

Esta condição é conhecida na literatura como não causalidade no sentido de Granger, que de forma intuitiva diz que uma variável y_t não causa uma outra, x_t , se o passado de y_t não ajuda a prever x_t .

Portanto, para se fazer previsões condicionais necessitamos de não causalidade além de exogeneidade fraca que é necessária para se fazer inferência válida. Tem-se o segundo conceito de exogeneidade que está relacionado a previsão.

Definição 3: Dizemos que x_t é exógeno forte para ψ se e somente se:

- (i) x_t é exógeno fraco para ψ ;
- (ii) y_t não causa x_t no sentido de Granger.

A seguir apresentamos o quarto exemplo que além de clarificar as diferenças entre causalidade e exogeneidade apresenta também o conceito de invariância dos parâmetros do modelo condicional à mudanças que possam ocorrer no modelo marginal. Este é a definição de super-exogeneidade e está relacionado a análise de políticas.

2.3.4 Exogeneidade, Causalidade e Invariância

Considere o seguinte processo gerador de dados:

$$y_t = \beta x_t + \varepsilon_{1t} \quad (75)$$

$$x_t = \delta_1 x_{t-1} + \delta_2 y_{t-1} + \varepsilon_{2t} \quad (76)$$

onde

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim NI \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right] \quad (77)$$

e estamos assumindo que $(\beta, \delta_1, \delta_2, \sigma_{11}, \sigma_{12}, \sigma_{22})$ são variação livre.

Seja β o parâmetro de interesse e observe que o sistema acima foi obtido através da seguinte distribuição conjunta:

$$\begin{matrix} y_t \\ x_t \end{matrix} \bigg| Z_{t-1} \sim NI \left[\begin{bmatrix} \beta\delta_1 x_{t-1} + \beta\delta_2 y_{t-1} \\ \delta_1 x_{t-1} + \delta_2 y_{t-1} \end{bmatrix}, \begin{bmatrix} \sigma_{11} + 2\beta\sigma_{12} + \beta^2\sigma_{22} & \sigma_{12} + \beta\sigma_{22} \\ \sigma_{12} + \beta\sigma_{22} & \sigma_{22} \end{bmatrix} \right]$$

então temos que

$$\begin{aligned} E(y_t \mid x_t, \mathbf{Z}_{t-1}) &= (\mu_{y|Z} - (\sigma_{12} + \beta\sigma_{22})\sigma_{22}^{-1}\mu_{x|Z}) + (\sigma_{12} + \beta\sigma_{22})\sigma_{22}^{-1}x_t \\ &= (\beta\delta_1 x_{t-1} + \beta\delta_2 y_{t-1} - (\sigma_{12} + \beta\sigma_{22})\sigma_{22}^{-1}(\delta_1 x_{t-1} + \delta_2 y_{t-1}) \\ &\quad + (\sigma_{12} + \beta\sigma_{22})\sigma_{22}^{-1}x_t) \\ &= (\sigma_{12} + \beta\sigma_{22})\sigma_{22}^{-1}x_t + c_1 x_{t-1} + c_2 y_{t-1} \end{aligned}$$

\Updownarrow

$$E(y_t \mid x_t, \mathbf{Z}_{t-1}) = bx_t + c_1 x_{t-1} + c_2 y_{t-1} \quad (78)$$

onde $b = (\sigma_{12} + \beta\sigma_{22})\sigma_{22}^{-1}$ e $c_i = \delta_i\sigma_{12}\sigma_{22}^{-1}$ para $i = 1, 2$ e

$$\begin{aligned} Var(y_t \mid x_t, \mathbf{Z}_{t-1}) &= \sigma_{11} + 2\beta\sigma_{12} + \beta^2\sigma_{22} - \frac{(\sigma_{12} + \beta\sigma_{22})^2}{\sigma_{22}} \\ &= \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21} = \sigma^2 \end{aligned} \quad (79)$$

Então o modelo condicional é dado por:

$$y_t = bx_t + c_1 x_{t-1} + c_2 y_{t-1} + u_t \quad \text{com } u_t \sim NI(0, \sigma^2) \quad (80)$$

onde $u_t = \varepsilon_{1t} - \sigma_{12}\sigma_{22}^{-1}\varepsilon_{2t}$.

Observe que quando $\sigma_{12} = 0$ as densidades do modelo condicional e marginal são normais com os seguintes parâmetros

$$E(y_t \mid x_t, \mathbf{Z}_{t-1}) = \beta x_t \quad \text{e} \quad Var(y_t \mid x_t, \mathbf{Z}_{t-1}) = \sigma_{11} \quad (81)$$

$$E(x_t \mid \mathbf{Z}_{t-1}) = \delta_1 x_{t-1} + \delta_2 y_{t-1} \quad \text{e} \quad Var(x_t \mid \mathbf{Z}_{t-1}) = \sigma_{22} \quad (82)$$

então (y_t, x_t) define um corte sequencial dado que $(\beta, \sigma_{11}) \in \Lambda_1$ e $(\delta_1, \delta_2, \sigma_{22}) \in \Lambda_2$ são variação livre e o parâmetro de interesse, β , é função somente de Λ_1 .

Desde modo x_t é exógeno fraco para β . Isto mostra que β pode ser recuperado sem o conhecimento do processo gerador de x_t quando a covariância entre y_t e x_t for zero.

Assim quando $\sigma_{12} = 0$ temos que $b = \beta$, mas se $\sigma_{12} \neq 0$, β não poderá ser recuperado de (80), isto é, através dos parâmetros do modelo condicional (b, c_1, c_2, σ^2) .

Agora se $\delta_2 \neq 0$, valores passados de y_t afetam x_t , e, portanto, y causa x no sentido de Granger logo x_t não é exógeno forte para β mesmo que $\sigma_{12} = 0$. Mas se $\delta_2 = \sigma_{12} = 0$ então x_t será exógeno forte para β .

Na literatura usual existem dois outros conceitos que, as vezes são confundidos com o conceito de exogeneidade. Esses conceitos são: variáveis estritamente exógenas e variáveis predeterminadas. Apresentaremos a seguir a definição destes dois conceitos.

Definição 4: Dizemos que x_t é estritamente exógeno para u_t se $E(x_t u_{t+i}) = 0$ para todo i .

Definição 5: Dizemos que x_t é pre-determinada para u_t se $E(x_t u_{t+i}) = 0$ para todo $i \geq 0$.

Observe que o conceito de exogeneidade está relacionado a parâmetros enquanto que estes dois últimos conceitos dizem respeito a ortogonalidade dos regressores com os erros.

No exemplo anterior se $\delta_1 \neq 0$ então x_t não será estritamente exógena, mas será pre-determinada pois:

$$\begin{aligned} E(x_t u_{t+i}) &= E[x_t(\varepsilon_{1t+i} - \sigma_{12}\sigma_{22}^{-1}\varepsilon_{2t+i})] \\ &= E(x_t\varepsilon_{1t+i}) - \sigma_{12}\sigma_{22}^{-1}E(x_t\varepsilon_{2t+i}) \end{aligned} \quad (83)$$

Agora usando (75-76) o processo gerador de x_t pode ser escrito da seguinte forma:

$$x_t = \frac{\delta_1\varepsilon_{1t-1}}{1 - (\delta_2 + \delta_1\beta)L} + \frac{\varepsilon_{2t}}{1 - (\delta_2 + \delta_1\beta)L} \quad (84)$$

Substituindo (84) em (83) temos dois termos onde o primeiro é dado por:

$$E \left[\left(\frac{\delta_1\varepsilon_{1t-1}}{1 - (\delta_2 + \delta_1\beta)L} + \frac{\varepsilon_{2t}}{1 - (\delta_2 + \delta_1\beta)L} \right) \varepsilon_{1t+i} \right] \quad (85)$$

e o segundo por

$$-\sigma_{12}\sigma_{22}^{-1}E\left[\left(\frac{\delta_1\varepsilon_{1t-1}}{1-(\delta_2+\delta_1\beta)L}+\frac{\varepsilon_{2t}}{1-(\delta_2+\delta_1\beta)L}\right)\varepsilon_{2t+i}\right] \quad (86)$$

Agora (85) implica em:

$$\delta_1 E\left[\left(\sum_{j=0}^{\infty}\theta^j\varepsilon_{1t-1-j}\right)\varepsilon_{1t+i}\right]+E\left[\left(\sum_{j=0}^{\infty}\theta^j\varepsilon_{2t-j}\right)\varepsilon_{1t+i}\right] \quad (87)$$

onde $\theta = \delta_2 + \delta_1\beta$.

Observe que a primeira expressão de (87)

(i) é zero para $i \geq 0$

(ii) para $i < 0$ vale $\theta^{|i|-1}\delta_1\sigma_{11}$

e a segunda expressão de (87)

(iii) é zero para $i > 0$,

(iv) para $i = 0$ vale σ_{12}

(v) para $i < 0$ vale $\theta^{|i|}\sigma_{12}$.

Agora (86) implica em:

$$-\sigma_{12}\sigma_{22}^{-1}\delta_1 E\left[\left(\sum_{j=0}^{\infty}\theta^j\varepsilon_{1t-1-j}\right)\varepsilon_{2t+i}\right]-\sigma_{12}\sigma_{22}^{-1}E\left[\left(\sum_{j=0}^{\infty}\theta^j\varepsilon_{2t-j}\right)\varepsilon_{2t+i}\right] \quad (88)$$

Observe que a primeira expressão de (88)

(vi) é zero para $i \geq 0$

(vii) para $i < 0$ vale $-\beta\delta_1\theta^{|i|-1}\sigma_{12}$

e a segunda expressão de (88)

(viii) é zero para $i > 0$

(ix) para $i = 0$ vale $-\beta\sigma_{22}$

(xii) para $i < 0$ vale $-\beta\theta^{|i|}\sigma_{22}$

Agora por (i), (iii), (iv), (vi), (viii) e (ix) temos que para $i \geq 0$ as expressões (85-86) resultam em:

$$\sigma_{12} - \beta\sigma_{22} = 0$$

o que implica que x_t é predeterminada

Agora por (ii), (v), (vii) e (xii) temos que para $i < 0$ as expressões (85-86) resultam em:

$$\begin{aligned}
& \theta^{|i|-1} \delta_1 \sigma_{11} + \theta^{|i|} \sigma_{12} - \beta \delta_1 \theta^{|i|-1} \sigma_{12} - \beta \theta^{|i|} \sigma_{22} \\
\Rightarrow & \theta^{|i|-1} (\delta_1 \sigma_{11} + \theta \sigma_{12} - \beta \delta_1 \sigma_{12} - \beta \theta \sigma_{22}) \\
\Rightarrow & \theta^{|i|-1} (\delta_1 \sigma_{11} + \theta \sigma_{12} - \beta \delta_1 \sigma_{12} - \theta \sigma_{12} \sigma_{22}^{-1} \sigma_{22}) \\
\Rightarrow & \theta^{|i|-1} \delta_1 (\sigma_{11} - \sigma_{12} \sigma_{22}^{-1} \sigma_{21}) = 0 \text{ se } \delta_1 = 0
\end{aligned}$$

Logo necessitamos da condição de que $\delta_1 = 0$ para que x_t seja estritamente exógeno.

Observe que a condição que devemos impor para que x_t seja exógeno fraco é $\sigma_{12} = 0$ e se esta condição não for satisfeita mas $\delta_1 = 0$ teremos que x_t não é exógeno farco mas é estritamente exógeno.

O último conceito a ser apresentado é o de super-exogeneidade que é definindo por:

Definição 6: Dizemos que x_t é super-exógeno para ψ se e somente se:

- (i) x_t é exógeno fraco para ψ
- (ii) λ_1 é invariante a intervenções que afetam λ_2

No exemplo anterior se $\sigma_{12} \neq 0$ e mesmo que β seja um parâmetro invariante a mudanças nos parâmetros do modelo marginal, isto é $(\delta_1, \delta_2, \sigma_{22})$, estas mudanças vão afetar os parâmetros do modelo condicional $(\beta, c_1, c_2, \sigma^2)$ violando a segunda condição de super-exogeneidade.

Neste caso a violação de exogeneidade fraca pode implicar em não constância do modelo condicional quando o modelo marginal muda. Temos então um teste indireto de exogeneidade fraca: devemos verificar se os parâmetros do modelo condicional são invariante no tempo e se a parte não sistemática do modelo, o erro, é um processo de inovação.

No exemplo anterior se $\sigma_{12} = 0$ a primeira condição de super-exogeneidade é satisfeita e se os parâmetros do modelo condicional, que neste caso são (β, σ_{11}) forem invariantes a mudanças que afetam os parâmetros do modelo marginal, isto é, $(\delta_1, \delta_2, \sigma_{22})$ então a segunda condição da definição acima também é satisfeita e, portanto, x_t é super-exógeno para ψ e isto vale mesmo que $\delta_1 \neq 0$, isto é, a variável pode ser super-exógena sem ser exógena forte.

Como mudanças que afetam o modelo marginal estão associadas a mudanças de política economica, o conceito de super-exogeneidade está associado a análise de política dentro da amostra.

2.4 Testes de Exogeneidade

A condição para exogeneidade fraca era que $\sigma_{12} = 0$. Esta condição para ser testada necessita da especificação não só do modelo condicional mas também do modelo marginal.

Mas sob a hipótese nula, isto é $H_0 : \sigma_{12} = 0$, (75) é o modelo condicional que tem distribuição independente do modelo marginal dado por (76) e, portanto, ambos podem ser estimados de forma eficiente por MQO. Denote por $\hat{\varepsilon}_y$ e $\hat{\varepsilon}_x$ os resíduos de (75) e (76) respectivamente o teste LM é construído da seguinte forma:

- (i) faça a regressão de $\hat{\varepsilon}_y$ em constante, x e $\hat{\varepsilon}_x$;
- (ii) sob H_0 , nR^2 onde R^2 é o coeficiente de determinação da regressão em (i), tem distribuição $\chi^2(1)$;
- (iii) rejeite H_0 se nR^2 exceder algum valor crítico preselecionado.

Uma outra forma, equivalente, de testar esta hipótese é usar y em vez de $\hat{\varepsilon}_y$ no item (i) acima e testar a significância do coeficiente de $\hat{\varepsilon}_x$.

Vamos mostrar que o coeficiente estimado para $\hat{\varepsilon}_x$ nas duas formas é o mesmo, que os desvios padrão estimados são os mesmos nas duas especificações e que o teste LM é equivalente ao teste t de significância do coeficiente de $\hat{\varepsilon}_x$.

Considere as duas especificações, isto é:

$$\hat{\varepsilon}_y = xb_1 + \hat{\varepsilon}_x b_2 + u \quad (89)$$

$$y = x\beta_1 + \hat{\varepsilon}_x \beta_2 + v \quad (90)$$

e desejamos mostrar primeiro que $\hat{b}_2 = \hat{\beta}_2$.

Multiplique (89) por $M_x = I - x(x'x)^{-1}x'$ e como $M_x x = 0$ temos:

$$M_x \hat{\varepsilon}_y = M_x \hat{\varepsilon}_x b_2 + M_x u = M_x \hat{\varepsilon}_x b_2 + u \quad (91)$$

e $M_x u = u$ porque $x'u = 0$.

Agora usando o fato que $\hat{\varepsilon}_y = M_x y$, o estimador de b_2 será dado por:

$$\hat{b}_2 = (\tilde{\varepsilon}_x M_x \hat{\varepsilon}_x)^{-1} \tilde{\varepsilon}_x M_x \hat{\varepsilon}_y = (\tilde{\varepsilon}_x M_x \hat{\varepsilon}_x)^{-1} \tilde{\varepsilon}_x M_x y \quad (92)$$

Agora o resíduo de (91) é dado por:

$$\begin{aligned}\hat{u} &= M_x \hat{\varepsilon}_y - M_x \hat{\varepsilon}_x (\tilde{\varepsilon}_x M_x \hat{\varepsilon}_x)^{-1} \tilde{\varepsilon}_x M_x \hat{\varepsilon}_y \\ &= N M_x \hat{\varepsilon}_y = N M_x y\end{aligned}\tag{93}$$

onde $N = I - M_x \hat{\varepsilon}_x (\tilde{\varepsilon}_x M_x \hat{\varepsilon}_x)^{-1} \tilde{\varepsilon}_x M_x$.

Agora pre-multiplicando (90) por M_x temos:

$$M_x y = M_x \hat{\varepsilon}_x \beta_2 + M_x v = M_x \hat{\varepsilon}_x \beta_2 + v\tag{94}$$

e o estimador de β_2 é dado por

$$\hat{\beta}_2 = (\tilde{\varepsilon}_x M_x \hat{\varepsilon}_x)^{-1} \tilde{\varepsilon}_x M_x y\tag{95}$$

que é identico a \hat{b}_2 obtido em (92). Observe que o resíduo de (94) é dado por:

$$\begin{aligned}\hat{v} &= M_x y - M_x \hat{\varepsilon}_x \hat{\beta}_2 = M_x y - M_x \hat{\varepsilon}_x (\tilde{\varepsilon}_x M_x \hat{\varepsilon}_x)^{-1} \tilde{\varepsilon}_x M_x y \\ &= N M_x y\end{aligned}\tag{96}$$

que é identico a (93).

Agora vamos mostrar que a estatística F para testar $\beta_2 = 0$ é equivalente a nR^2 de (89).

A estatística F para testar $\beta_2 = 0$ é dada por:

$$\begin{aligned}F &= \frac{(SQRes_R - SQRes_I)/r}{SQRes_I/(n-k)} = \frac{(y'M_x y - \hat{v}\hat{v})/r}{\hat{v}\hat{v}/(n-k)} \\ &= \frac{(y'M_x y - \hat{u}\hat{u})/r}{\hat{u}\hat{u}/(n-k)}\end{aligned}\tag{97}$$

e nR^2 de (89) é dado por:

$$nR^2 = \frac{SQReg}{SQTotal/n} = \frac{y'M_x y - \hat{u}\hat{u}}{y'M_x y/n}\tag{98}$$

observe que o denominador de (97) e de (98) são ambos estimadores consistentes para σ^2 , e, portanto nR^2 e rF tem distribuição $\aleph^2(r)$.

2.4.1 Testes de Exogeneidade de Durbin-Wu-Hausman

O modelo (75-76) escrito em forma matricial é dado por:

$$y = x\beta + \varepsilon_1 \quad (99)$$

$$x = x_{-1}\delta_1 + y_{-1}\delta_2 + \varepsilon_2 \quad (100)$$

onde $x_{-1} = [x_{n-1}, x_{n-2}, \dots, x_0]'$ e $y_{-1} = [y_{n-1}, y_{n-2}, \dots, y_0]'$.

Já vimos que se $\sigma_{12} \neq 0$ então ε_{2t} afeta x_t e ε_{1t} e logo x_t e ε_{1t} serão correlacionados na primeira equação e portanto a utilização de MQO nesta equação implica em estimadores viesados e inconsistentes para β .

Em vez de derivar um teste direto para σ_{12} o procedimento de Durbin-Wu-Hausman (veja Durbin (1954), Wu (1973) e Hausman (1978)) e concentra na primeira equação e testa

$$H_0 : p \lim \left(\frac{1}{n} x' \varepsilon_1 \right) = 0$$

contra a alternativa

$$H_1 : p \lim \left(\frac{1}{n} x' \varepsilon_1 \right) \neq 0$$

A idéia básica deste teste é comparar dois possíveis estimadores de β . Sob H_0 o estimador de MQO $\hat{\beta}_0 = (x'x)^{-1}x'y$ é consistente e assintoticamente eficiente. Sob H_1 este estimador será inconsistente. Suponha então que é possível obter z vetor de instrumentos para x , que satisfaz:

$$p \lim \left(\frac{1}{n} z' x \right) \neq 0 \text{ e } p \lim \left(\frac{1}{n} z' \varepsilon_1 \right) = 0$$

e o estimador de variáveis instrumentais $\hat{\beta}_1 = (z'x)^{-1}z'y$ pode ser construído e será um estimador consistente em ambas as hipóteses.

Como ambos os estimadores são consistentes sob H_0 a diferença entre eles será assintoticamente negligenciável.

Denote esta diferença por

$$\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$$

que sob H_0 terá distribuição

$$\frac{\widehat{q}}{d.p.(\widehat{q})} \stackrel{a}{\sim} N(0, 1) \text{ ou } \frac{\widehat{q}^2}{Var(\widehat{q})} \stackrel{a}{\sim} \aleph^2(1)$$

Como, sob H_0 , o estimador $\widehat{\beta}_0$ é mais eficiente do que $\widehat{\beta}_1$ temos que

$$Var(\widehat{q}) = Var(\widehat{\beta}_1) - Var(\widehat{\beta}_0)$$

e o teste pode ser escrito da seguinte forma:

$$\frac{\widehat{q}^2}{Var(\widehat{\beta}_1) - Var(\widehat{\beta}_0)} \stackrel{a}{\sim} \aleph^2(1) \quad (101)$$

O estimador de variáveis instrumentais pode ser escrito com um estimador de mínimos quadrados em dois estágios, a saber

$$\widehat{\beta}_1 = (\widehat{x}\widehat{x})^{-1}\widehat{x}y$$

onde

$$\widehat{x} = z(z'z)^{-1}z'x = P_z x$$

Observe que este estimador é obtido da seguinte forma.

No primeiro estágio pela regressão de x em z obtém-se o valor ajustado \widehat{x} e o resíduo deste primeiro estágio $v = x - \widehat{x}$ por construção é ortogonal a \widehat{x} e portanto temos que $\widehat{x}\widehat{x} = \widehat{x}x$.

No segundo estágio fazemos a regressão de y em \widehat{x} .

Desde modo a diferença entre os dois estimadores, \widehat{q} , pode ser escrita da seguinte forma:

$$\begin{aligned} \widehat{q} &= \widehat{\beta}_1 - \widehat{\beta}_0 = (\widehat{x}\widehat{x})^{-1}\widehat{x}y - (x'x)^{-1}x'y \\ &= (\widehat{x}\widehat{x})^{-1}[\widehat{x}y - (\widehat{x}\widehat{x})(x'x)^{-1}x'y] \\ &= (\widehat{x}\widehat{x})^{-1}[\widehat{x}y - (\widehat{x}x)(x'x)^{-1}x'y] \\ &= (\widehat{x}\widehat{x})^{-1}[\widehat{x}M_x y] \end{aligned} \quad (102)$$

A expressão acima está relacionada ao coeficiente de \widehat{x} na regressão auxiliar de y em x e \widehat{x} um vez que esta regressão é dada por:

$$y = x\beta + \widehat{x}\delta + u \quad (103)$$

e o estimador de δ é dado por

$$\widehat{\delta} = (\widehat{x}M_x\widehat{x})^{-1}[\widehat{x}M_xy] \quad (104)$$

com matriz de variância-covariância dada por:

$$Var(\widehat{\delta}) = \sigma^2(\widehat{x}M_x\widehat{x})^{-1} \quad (105)$$

e a hipótese nula agora é dada por $H_0 : \delta = 0$, cujo estatística de teste é dada por

$$\frac{\widehat{\delta}^2}{Var(\widehat{\delta})} \stackrel{a}{\sim} \mathbb{N}^2(1) \quad (106)$$

e vamos mostrar que esta expressão é idêntica (101).

Substituindo (104) e (105) em (106) temos

$$\frac{\widehat{\delta}^2}{Var(\widehat{\delta})} = \frac{1}{\sigma^2}(\widehat{x}M_x\widehat{x})^{-1}[\widehat{x}M_xy]^2 \quad (107)$$

Agora por (102) temos que $\widehat{q} = (\widehat{x}\widehat{x})^{-1}[\widehat{x}M_xy]$ e $Var(\widehat{q})$ é dada por:

$$\begin{aligned} Var(\widehat{q}) &= \sigma^2 \left[\frac{1}{\widehat{x}\widehat{x}} - \frac{1}{x'x} \right] \\ &= \sigma^2 \left[\frac{x'x - \widehat{x}\widehat{x}}{(\widehat{x}\widehat{x})(x'x)} \right] \\ &= \sigma^2 \frac{1 - (x'x)^{-1}(\widehat{x}\widehat{x})}{\widehat{x}\widehat{x}} \\ &= \sigma^2 \frac{\widehat{x}\widehat{x} - \widehat{x}\widehat{x}(x'x)^{-1}\widehat{x}\widehat{x}}{(\widehat{x}\widehat{x})^2} \\ &= \sigma^2 \frac{\widehat{x}\widehat{x} - \widehat{x}x(x'x)^{-1}x\widehat{x}}{(\widehat{x}\widehat{x})^2} \\ &= \sigma^2 \frac{\widehat{x}M_x\widehat{x}}{(\widehat{x}\widehat{x})^2} \end{aligned} \quad (108)$$

Substituindo (108) e (102) em (101) temos:

$$\begin{aligned}\frac{\widehat{q}^2}{Var(\widehat{q})} &= \frac{(\widehat{x}\widehat{x})^{-2}[\widehat{x}M_x y]^2}{\sigma^2 \frac{\widehat{x}M_x \widehat{x}}{(\widehat{x}\widehat{x})^2}} \\ &= \frac{1}{\sigma^2}(\widehat{x}M_x \widehat{x})^{-1}[\widehat{x}M_x y]^2\end{aligned}$$

que é exatamente a expressão (107).

2.4.2 Testando exogeneidade fraca de INC para CONS em DATAGIVE

Considere o seguinte modelo:

$$Cons_t = \beta_0 + \beta_1 Inc_t + \varepsilon_t \quad (109)$$

e suspeitamos que Inc não é exógeno fraco para β . Vamos usar o procedimento da seção (2.4), isto é, o modelo marginal para Inc será dado por

$$Inc_t = \pi_0 + \pi_1 Cons_{t-1} + \pi_2 Inc_{t-1} + v_t \quad (110)$$

O programa abaixo apresenta o teste

Teste de Exogeneidade

Temos os seguintes resultados para o primeiro estágio

Tabela 1: Teste de Exogeneidade primeiro estágio usando (109)

EQ(1) Teste de Exogeneidade primeiro estágio				
The estimation sample is: 1 - 158				
	Coefficient	Std. Error	t-Statistics	p-value
<i>const</i>	89.6142	23.0012	3.896	0.0001
<i>Const_{t-1}</i>	0.2490	0.0549	4.536	0.0000
<i>Inc_{t-1}</i>	0.6547	0.0691	9.481	0.0000
no of observations: 158				
no of parameters: 3				
$\sigma_{\varepsilon} = 3.131$				
$R^2 = 0.9157$		$\bar{R}^2 = 0.9146$		
$AIC = 363.6097$				
$BIC = 363.726$				
log-likelihood = -2879.16				

E para o segundo estágio estimamos o modelo:

$$Cons_t = \beta_0 + \beta_1 Inc_t + \hat{v}_t + \varepsilon_t \quad (111)$$

onde \hat{v}_t é o resíduo do primeiro estágio.

Temos então os seguintes resultados

Tabela 2: Teste de Exogeneidade usando (111)

EQ(2) Teste de Exogeneidade segundo estágio				
The estimation sample is: 1 - 158				
	Coefficient	Std. Error	t-Statistics	p-value
<i>const</i>	-242.4547	27.4835	-8.822	0.0000
<i>Inc_t</i>	1.2543	0.0308	40.693	0.0000
<i>\hat{v}_t</i>	-0.7584	0.1061	-7.146	0.0000
no of observations: 158				
no of parameters: 5				
$\sigma_\varepsilon = 3.959$				
$R^2 = 0.9155$		$\bar{R}^2 = 0.9144$		
$AIC = 437.7544$				
$BIC = 437.8707$				
log-likelihood = -34576.6				

Observe que como \hat{v}_t é significativo, temos que *Inc* não é exógeno fraco e portanto a inferência no modelo condicional é inválida.

2.5 Exercícios

1. As questões abaixo são do tipo falsa, verdadeira ou ambígua. Justifique as suas respostas.
 - (i) o conceito de exogeneidade fraca é fundamental para se fazer inferência no modelo marginal;
 - (ii) se os parâmetros de interesse são do modelo conjunto e estimamos somente o modelo condicional a inferência é inválida;
 - (iii) o conceito de exogeneidade forte é fundamental se quisermos fazer previsões para o modelo condicional;
 - (iv) o conceito de super-exogeneidade está relacionado ao conceito de invariância dos parâmetros do modelo condicional à intervenções nos parâmetros do modelo marginal.

2. Considere o seguinte sistema:

$$\begin{aligned}y_t &= \beta x_t + \varepsilon_t \\x_t &= \pi_1 y_{t-1} + \pi_2 x_{t-1} + v_t\end{aligned}$$

com

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \sim NI \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} : \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix} \right]$$

- (i) escreva o sistema na forma $Z_t = \Phi Z_{t-1} + u_t$ onde $Z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}$
- (ii) obtenha o modelo condicional, isto é, $E[y_t \mid x_t, Z_{t-1}]$
- (iii) podemos dizer que x_t é exógeno fraco para β ? Se não quais as condições que devem ser satisfeitas para que seja.
- (iv) podemos dizer que x_t é exógeno forte para β ? Se não quais as condições que devem ser satisfeitas para que seja.

3. Considere o seguinte sistema:

$$\begin{aligned}y_t &= \beta x_t + \varepsilon_t \\x_t &= \pi_1 y_{t-1} + \pi_2 x_{t-1} + v_t\end{aligned}$$

com

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \sim NI \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} : \begin{pmatrix} \sigma_\varepsilon^2 & \rho \sigma_\varepsilon \sigma_v \\ \rho \sigma_\varepsilon \sigma_v & \sigma_v^2 \end{pmatrix} \right]$$

- (i) escreva o sistema na forma $Z_t = \Phi Z_{t-1} + u_t$ onde $Z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}$
- (ii) obtenha o modelo condicional, isto é, $E[y_t \mid x_t, Z_{t-1}]$

- (iii) podemos dizer que x_t é exógeno fraco para β ? Se não quais as condições que devem ser satisfeitas para que seja.
 - (iv) podemos dizer que x_t é exógeno forte para β ? Se não quais as condições que devem ser satisfeitas para que seja.
4. Com os dados de DATAGIVE.XLSX teste causalidade de Granger entre CONS e INC e entre OUTPUT e INFLAT. O que você conclui em cada um dos casos. O número de defasagens utilizados tem alguma importância? Porque? Se você mudar a amostra usada nos testes, os resultados se modificam? Porque?

3 Tipologia de Modelos Dinâmicos

7

Vamos considerar o modelo ADL(1,1), isto é:

$$y_t = m + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \quad (112)$$

onde $\varepsilon_t \sim NI(0, \sigma_\varepsilon^2)$ e por simplicidade vamos assumir que $m = 0$. Existem 10 sub-modelos de (112) que são obtidos impondo restrições que devem ser testadas. Estes sub-modelos são apresentados na tabela 1 do apêndice.

Três observações importantes devem ser feitas antes de comentarmos cada um destes modelos. A primeira está relacionada ao status da variável x_t a segunda a dependência das propriedades do modelo as propriedades dos dados e a última se cada tipo de modelo está sendo tratado como corretamente especificado ou se é uma aproximação de um processo gerador dos dados mais geral. Todas estas três observações devem ser feitas uma vez que não se começou a análise do modelo mais geral e também porque o processo gerador dos dados é em geral desconhecido.

Para que a inferência nestes modelos seja válida devemos assumir que estes representam o modelo condicional válido e, portanto, estamos implicitamente assumindo que x_t é exógeno fraco para os parâmetros de interesse que

⁷Esta sessão é baseada em Hendry et al. (1984)

são $\theta = (\alpha_1, \beta_0, \beta_1, \sigma_\varepsilon^2)$. Assumindo que x_t é exógeno fraco para os parâmetros de interesse, se qualquer sub-modelo da tipologia for válido então todos os modelos menos restrito que aquele também serão.

Estamos assumindo que x_t é $I(1)$ e portanto Δx_t é $I(0)$. Isto resolve a segunda observação embora algumas séries temporais economicas sejam $I(0)$, por exemplo, a taxa de desemprego, e portanto o caso de x_t estacionário também é relevante. Se x_t e y_t cointegrarem então $u_t = y_t - \gamma x_t$ será $I(0)$, mas isto pode não acontecer e portanto u_t será $I(1)$.

A tipologia trata cada modelo em separado e pressupõe que esta especificação é correta e portanto a terceira observação não se aplica.

Quanto a estimação destes modelos, como estamos assumindo que o modelo está corretamente especificado e que a parte não sistemática é um ruído branco, quase todos podem ser estimados por M.Q.O., com a exceção do modelo de regressão estática com erros correlacionados. Neste modelo apresentaremos dois métodos de estimação, a saber, M.Q.G. e um procedimento iterativo.

Para todos os modelos não apresentaremos testes de especificação ou falta de especificação. Mas para o modelo de regressão estática com erros correlacionados serão apresentados testes para detectar autocorrelação nos erros, uma vez que estes testes também servem para verificar a especificação dinâmica do modelo.

3.1 Regressão Estática

O modelo é dado por:

$$y_t = m + \beta_0 x_t + \varepsilon_t \quad (113)$$

(em geral tanto β_0 quanto x_t são vetores) e tem importância em sistemas macroeconômicos uma vez que pode ser interpretado como uma equação estrutural.

Na prática a hipótese de que ε_t é ruído branco é pouco provável uma vez que falta de dinâmica induz correlação nos erros e correlação espúria entre as variáveis pode induzir alta correlação nos erros tornando a inferência usual inválida.

Recentemente regressão estática reapareceu na literatura como parte do procedimento em dois passos de Engle & Granger (veja Engle and Granger (1987)) para se testar cointegração.

Deste modo β_0 será um estimador de γ elasticidade de longo-prazo. Este tipo de procedimento para se obter a elasticidade de longo-prazo foi questionado na literatura uma vez que a relação estática impõe certas restrições no multiplicador médio que podem não ser válida, veja, por exemplo Ericsson and MacKinnon (2002).

Uma outra justificativa para o modelo estático está relacionada a abordagem da autocorrelação, isto é, estimasse o modelo estático e corrigisse a possível autocorrelação dos resíduos através de um modelo autorregressivo, mas como veremos mais a frente este tipo de procedimento pode impor restrições de fatores comuns que podem ser inválidas.

3.1.1 Exemplo usando DATAGIVE.XLSX

Considere o modelo

$$\text{Log}(Cons_t) = m + \beta_0 \text{Log}(Inc_t) + \varepsilon_t \quad (114)$$

o seguinte comando do R estima este modelo por M.Q.O.

Regressão Estática

e o resultado é apresentado abaixo:⁸

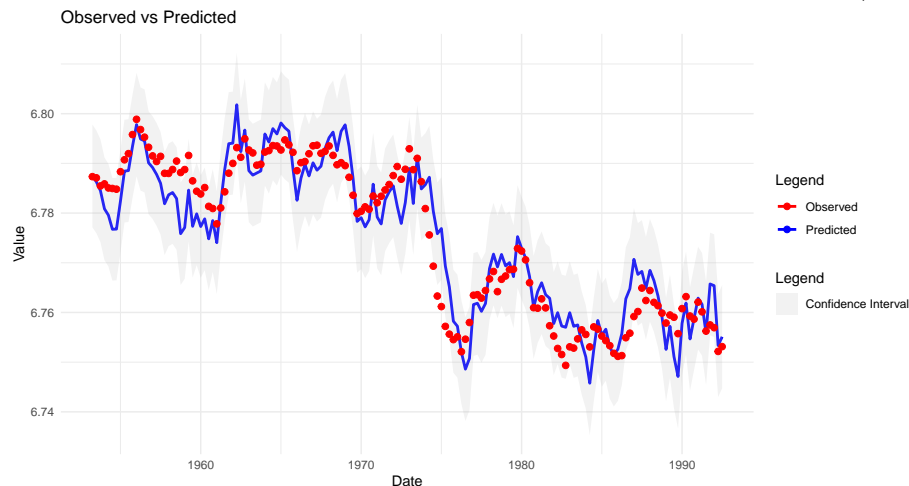
Tabela 3: Regressão Estática

EQ(3) Regressão Estática variável dependente $\log(Cons_t)$				
Desvios Padrões corrigidos por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
<i>Intercepto</i>	-1.4604	0.2930	-4.9844	0.0000
<i>log(Inc_t)</i>	1.2077	0.043	28.0826	0.0000
no of observations: 158				
no of parameters: 2				
$\sigma_\varepsilon = 0.005202$				
$R^2 = 0.8877$		$\bar{R}^2 = 0.8870$		
F-statistics(1,156)= 1233 p-value = 0.000				
$AIC = -7.6543$				
$BIC = -7.5962$				
log-likelihood = 606.6916				

⁸HAC é a correção dos desvios padrões das estimativas por heteroscedasticidade e autocorrelação de Newey-West veja Newey and West (1987) e, serão usadas em todas as estimativas.

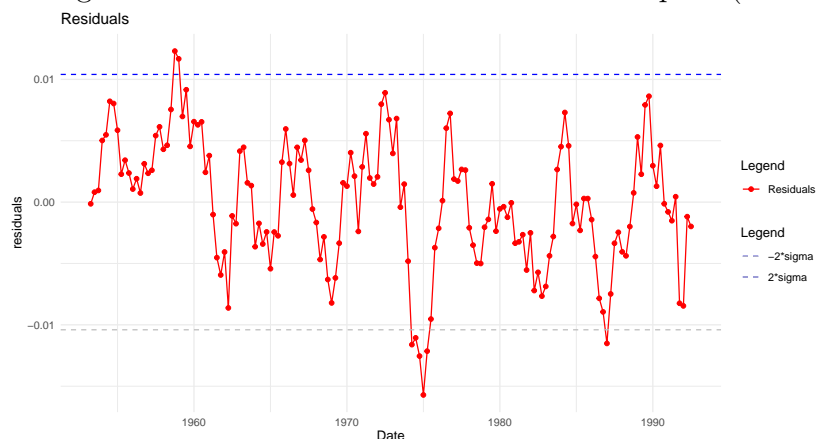
O gráfico abaixo do observado e previsto e os respectivos intervalos de confiança para a previsão. Observe que o observado está dentro da banda de previsão do modelo.

Figura 1: Observado, Previsto e Intervalo de Confiança para (114)



O gráfico abaixo apresenta os resíduos com as respectivas bandas. Observe que os resíduos indicam uma possível falta de especificação uma vez que eles não parecem ruído branco.

Figura 2: Resíduos e Intervalo de Confiança para (114)



3.2 Modelo Autorregressivo Univariado

O processo gerador dos dados é:

$$y_t = m + \alpha_1 y_{t-1} + \varepsilon_t \quad (115)$$

que é uma parametrização possível para modelos univariados de séries temporais.

Observe que no caso em que $\alpha_1 = 1$ este modelo é um passeio aleatório se ε_t for um ruído branco.

Modelo Autorregressivos são muito usados em previsão ex-ante tanto no caso univariado quanto no caso multivariado e competem com modelos econométricos em termos de poder preditivo.

No entanto em economia as decisões economicas são interdependentes induzindo modelos autorregressivos mas estes não são autonomos, porque mudanças no processo gerador de x_t podem afetar o processo gerador de y_t , uma vez que o processo autorregressivo para y_t pode ser obtido através de eliminação ou marginalização do processo gerador de x_t .

Por exemplo, suponha que $x_t = x_{t-1} + v_t$ com $v_t \sim NI(0, \sigma_t^2)$ e se as restrições $\alpha_1 = 1$ e $\beta_0 = -\beta_1$ forem válidas temos que o processo gerador de y_t é dado por $y_t = y_{t-1} + \varepsilon_t + \beta_0 v_t$ que tem variância não constante e dada por $t(\sigma_\varepsilon^2 + \beta_0^2 \sigma_v^2)$.

Deste modo o modelo econométrico deve ter melhor aderência do que o modelo autorregressivo, ou é menos mal especificado dinamicamente do que o modelo autorregressivo e, também as previsões do modelo econométrico devem ser melhores do que as do modelo autorregressivo ou então a não constância do modelo econométrico deve ser testada.

3.2.1 Continuação do Exemplo

Foi estimado um $AR(1)$ para $\log(Cons_t)$.

$$\log(Cons_t) = m + \alpha_1 \log(Cons_{t-1}) + \varepsilon_t \quad (116)$$

Os comandos no R são apresentados abaixo

$$AR(1)$$

tem-se os seguintes resultados:⁹

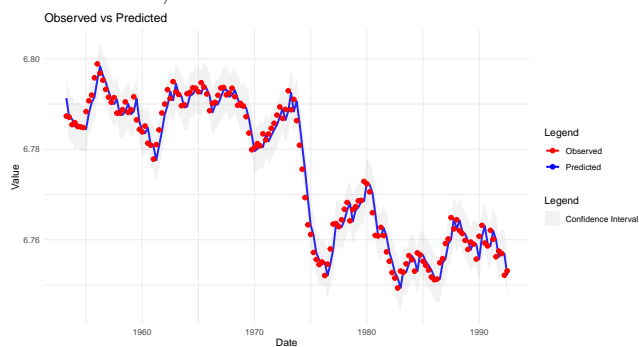
Tabela 4: $AR(1)$

EQ(4) Modelo $AR(1)$ para $\log(Cons_t)$				
Desvios Padrões sem correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
<i>Intercepto</i>	0.0712	0.0886	0.803	0.4230
$\log(Cons_{t-1})$	0.9895	0.0131	75.628	0.0000
no of observations: 158				
no of parameters: 2				
$\sigma_\varepsilon = 0.002529$				
$R^2 = 0.9734$ $\bar{R}^2 = 0.9733$				
F-statistics(1,156)= 5720 p-value = 0.000				
$AIC = -9.0965$				
$BIC = -9.0384$				
log-likelihood = 720.6236				

Observe que o coeficiente de $\log(Cons_{t-1})$ é próximo de 1 indicando que esta série é não estacionária

O gráfico abaixo do observado e previsto e os respectivos intervalos de confiança para a previsão. Observe que o observado está dentro da banda de previsão do modelo.

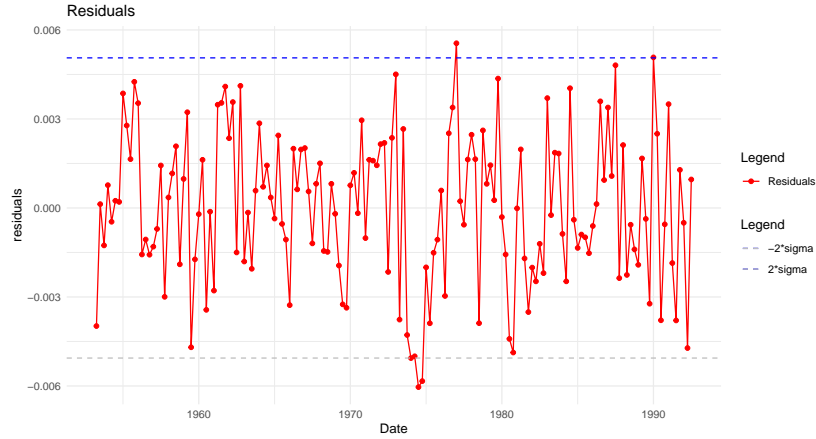
Figura 3: Observado, Previsto e Intervalo de Confiança para (116



⁹O modelo estimado foi uma regressão onde a variável independente foi $\log(Cons_{t-1})$. O R não calcula a correção para o coeficiente da variável defasada. Desde modo são apresentados os desvios padrões sem correção.

O gráfico abaixo apresenta os resíduos com as respectivas bandas. Observe que os resíduos neste caso parecem um ruído branco.

Figura 4: Resíduos e Intervalo de Confiança para (116)



Observe que o BIC é menor do que o da Regressão Estática, este modelo é escolhido.

3.3 Indicadores Antecedentes

O processo gerador dos dados é dado por:

$$y_t = m + \beta_1 x_{t-1} + \varepsilon_t \quad (117)$$

e pode ser usado para fazer previsões para y_t um vez que x_t antecede y_t .

Na ausência de uma teoria comportamental é pouco provável que β_1 seja constante. Deste modo as previsões com este modelo não serão boas em momentos de mudanças.

Observe que neste modelo a variável y_{t-1} foi excluída e a priori esta exclusão pode implicar em falta de especificação dinâmica.

3.3.1 Continuação do Exemplo

Foi estimado um modelo para $\log(Cons_t)$ usando $\log(Inc_t)$ como indicador antecedente, i.e. a variável independente é $\log(Inc_{t-1})$. Temos o seguinte modelo a ser estimado:

$$\log(Cons_t) = m + \beta_1 \log(Inc_{t-1}) + \varepsilon_t \quad (118)$$

O programa no R para estimar este modelo é apresentado abaixo

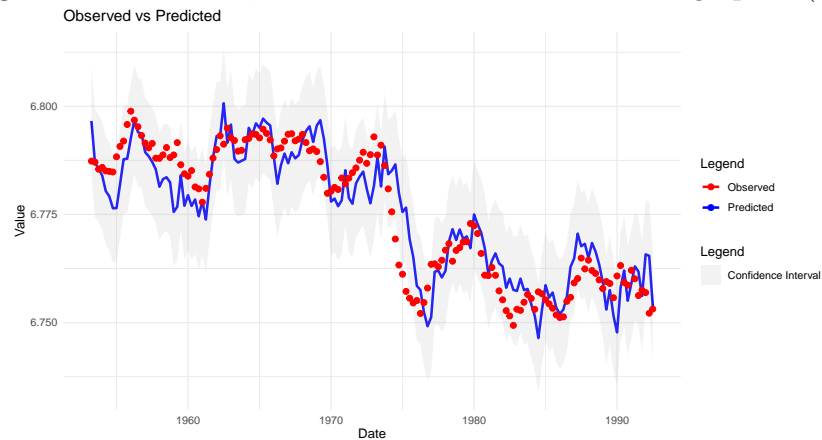
Indicador Antecedente

tem-se os seguintes resultados:

Tabela 5: Indicador Antecedente				
EQ(5) Modelo Indicador Antecedente				
Desvios Padrões com correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
<i>Intercepto</i>	-1.2058	0.3675	-3.2811	0.001276
<i>log(Cons_{t-1})</i>	1.1748	0.0541	21.7099	0.0000
no of observations: 158				
no of parameters: 2				
$\sigma_\varepsilon = 0.006286$				
$R^2 = 0.8360$ $\overline{R}^2 = 0.8349$				
F-statistics(1,156)= 5720 p-value = 0.000				
<i>AIC</i> = -7.2756				
<i>BIC</i> = -7.2174				
log-likelihood = 576.7722				

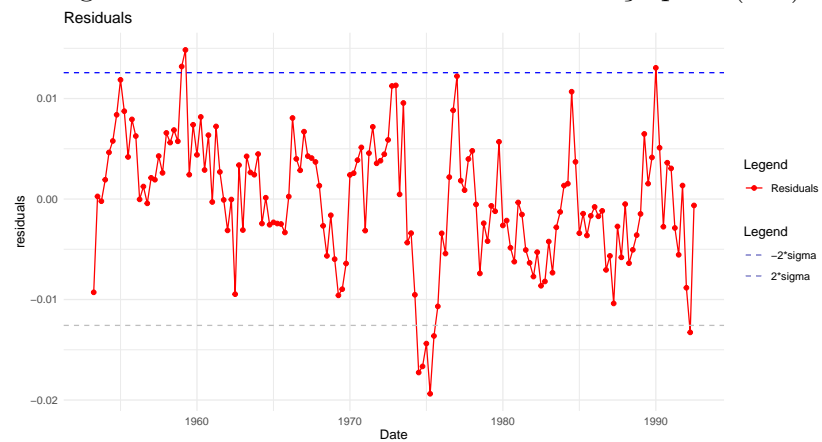
O gráfico abaixo do observado e previsto e os respectivos intervalos de confiança para a previsão. Observe que o observado está dentro da banda de previsão do modelo.

Figura 5: Observado, Previsto e Intervalo de Confiança para (118)



O gráfico abaixo apresenta os resíduos. Observe que estes resíduos não parecem ser ruído branco, indicando um problema de falta de especificação no modelo.

Figura 6: Resíduos e Intervalo de Confiança para (118)



O Schwarz deste modelo é maior do que do modelo $AR(1)$, portanto, o melhor modelo continua sendo o $AR(1)$,

3.4 Modelos de Taxa de Crescimento

Séries economicas tem tendência seja ela determinística ou estocástica e, portanto, antes de fazer análise estatística dos dados era sugerido diferenciar os dados.

Existem outras transformações que retiram tendência, por exemplo, razões entre variáveis, mas que não excluem a relação de longo prazo que pode existir entre as variáveis, e como veremos mais adiante, estas transformações estão relacionadas ao conceito de cointegração.

O modelo de taxa de crescimento é dado por:

$$\Delta y_t = \beta_1 \Delta x_t + \varepsilon_t \quad (119)$$

e este modelo não sofre do problema de regressão espúria que está presente em modelos com variáveis $I(1)$.

Por outro lado, se a variância de Δx_t é grande em relação a variância de Δy_t , β_0 será pequeno mesmo que y_t e x_t cointegrem com $\gamma = 1$. Além disto, mesmo que $y_t = \gamma x_t$ implique que $\Delta y_t = \gamma \Delta x_t$ a recíproca não é verdadeira.

A priori não há nenhuma razão para se excluir o nível das variáveis uma vez que desequilíbrios iniciais não podem ser assumidos como irrelevantes.

3.4.1 Continuação do Exemplo

Foi estimado um modelo de taxa de crescimento para CONS e INC. Temos o seguinte modelo a ser estimado:

$$\Delta(\log(Cons_t)) = \beta_1 \Delta(\log(Inc_{t-1})) + \varepsilon_t \quad (120)$$

O programa no R para estimar este modelo é apresentado abaixo

Taxa de Crescimento

tem-se os seguintes resultados:

Tabela 6: Taxa de Crescimento com $\Delta(\log(Cons_t))$ com variável dependente

EQ(6) Taxa de Crescimento com $\Delta(\log(Cons_t))$ com variável dependente				
Desvios Padrões com correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
$\Delta(\log(Inc_t))$	0.5034	0.0271	18.574	0.0000
no of observations: 158				
no of parameters: 1				
$\sigma_\varepsilon = 0.001685$				
$R^2 = 0.5592$ $\bar{R}^2 = 0.5564$				
F-statistics(1,157)= 199.2 p-value = 0.000				
AIC = -9.9149				
BIC = -9.8761				
log-likelihood =785.2737				

Observe que o critério de informação desta regressão não é comparável com as anteriores uma vez que a variável dependente neste caso é $D(\log(CONS))$ em vez de $\log(CONS)$. Uma forma de resolver este problema é estimar o modelo escrito da seguinte forma, que preservar o modelo de taxa de crescimento uma vez que o coeficiente de $CONS_{t-1}$ está restrito a ser um:

$$\log(Cons_t) = \log(Cons_{t-1}) + \beta_1 \Delta(\log(Inc_t)) + \varepsilon_t \quad (121)$$

obtem-se os seguintes resultados:

Tabela 7: Taxa de Crescimento com $\log(Cons_t)$ como variável dependente

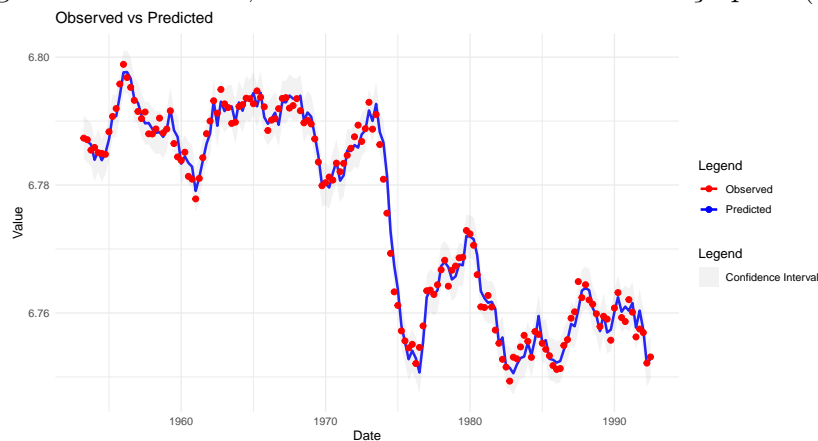
EQ(7) Taxa de Crescimento com $\log(Cons_t)$ como variável dependente				
Desvios Padrões com correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
$\Delta(\log(Inc_t))$	0.5034	0.0271	18.574	0.0000
no of observations: 158				
no of parameters: 1				
$\sigma_\varepsilon = 0.001685$				
$R^2 = 1.0000$ $\bar{R}^2 = 1.0000$				
F-statistics(1,157)= 2.544e+09 p-value = 0.000				
AIC = -9.9149				
BIC = -9.8761				
log-likelihood =785.2737				

e agora podemos comparar o critério de informação desta regressão com os anteriores. Observe que o Schwarz diminui o que implica que este modelo é

melhor do que o $AR(1)$. Observe também que todas as estimativas são iguais com exceção do R^2 e \bar{R}^2 , que agora são próximo de um enquanto que na estimação com $\Delta(\log(Cons_t))$ com variável dependente era próximo de 0.6.

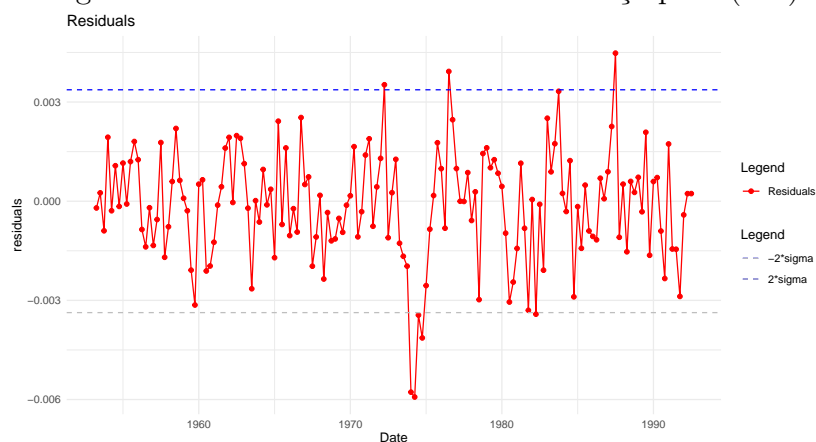
Abaixo é apresentado o gráfico do observado, previsto e Intervalo de Confiança para este modelo

Figura 7: Observado, Previsto e Intervalo de Confiança para (121)



O gráfico abaixo apresenta os resíduos. Observe que estes resíduos parecem ser ruído branco, não indicando um problema de falta de especificação no modelo.

Figura 8: Resíduos e Intervalo de Confiança para (121)



3.5 Defasagens Distribuídas

O processo gerador dos dados é dado por:

$$y_t = m + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \quad (122)$$

Observe que as críticas a regressão estática também se aplicam aqui.

A exclusão a priori de y_{t-1} pode mais uma vez implicar em falta de especificação dinâmica.

3.5.1 Continuação do Exemplo

Foi estimado um modelo de defasagens distribuídas para $\log(Cons_t)$ e $\log(Inc_t)$, isto é:

$$\log(Cons_t) = m + \beta_0 \log(Inc_t) + \beta_1 \log(Inc_{t-1}) + \varepsilon_t \quad (123)$$

Temos o seguinte programa no R:

Defasagens Distribuídas

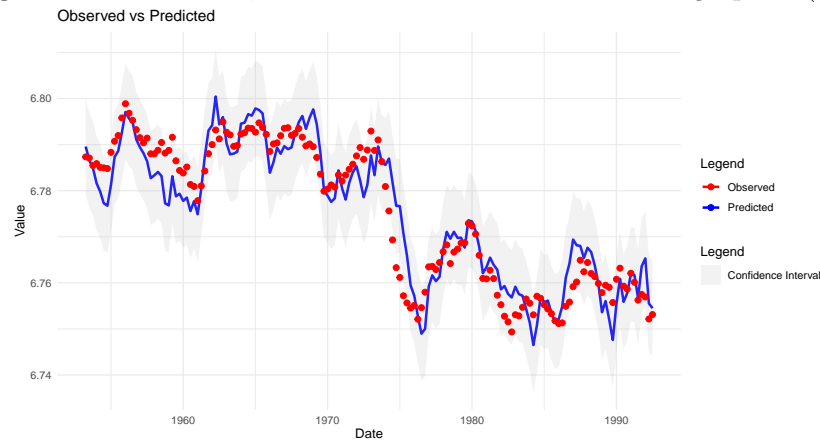
tem-se os seguintes resultados:

Tabela 8: Defasagens Distribuídas

EQ(8) Defasagens Distribuídas				
Desvios Padrões com correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
<i>Intercepto</i>	-1.5398	0.2913	-5.2863	0.0000
<i>log(Inc_t)</i>	0.9784	0.1500	6.5234	0.0000
<i>log(Inc_{t-1})</i>	0.2456	0.1553	1.5814	0.1158
no of observations: 158				
no of parameters: 3				
$\sigma_{\varepsilon} = 0.005137$				
$R^2 = 0.8912$		$\overline{R}^2 = 0.8898$		
F-statistics(2,155)= 634.7 p-value = 0.000				
$AIC = -7.67322$				
$BIC = -7.595686$				
log-likelihood =609.1844				

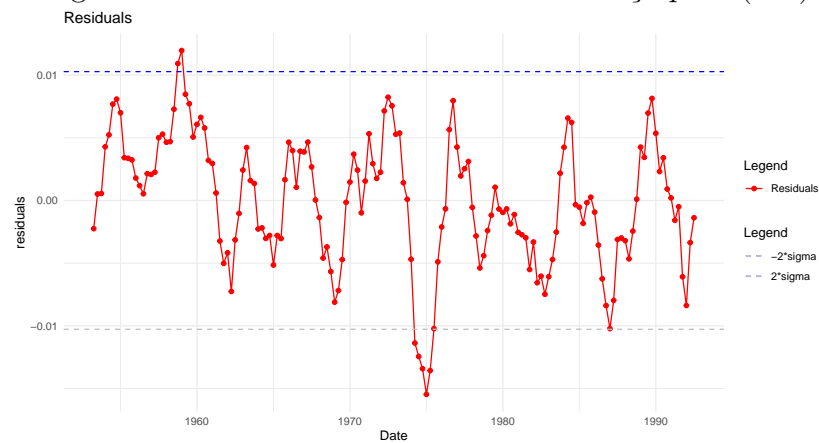
O gráfico dos observados, previstos e bandas é apresentado abaixo

Figura 9: Observado, Previsto e Intervalo de Confiança para (123)



O gráfico dos resíduos e respectivos Intervalos de Confianças é apresentado abaixo

Figura 10: Resíduos e Intervalo de Confiança para (123)



Comparando com os modelos anteriores este modelo é pior do que o modelo de taxa de crescimento.

3.6 Ajustamento Parcial

O processo gerador dos dados é dado por:

$$y_t = m + \beta_0 x_t + \alpha_1 y_{t-1} + \varepsilon_t \quad (124)$$

e este modelo pode ser obtido através de um processo de ajustamento onde a relação desejada de longo prazo é dada por $y_t^* = \frac{m}{(1-\alpha_1)} + \gamma x_t$ e o custo de ajustamento são quadráticos.

Uma outra forma de obter o modelo de ajustamento parcial é assumir que a variável y_t representa o realizado e que y_t^* representa o desejado de longo prazo e existe o fator x_t . A relação de equilíbrio de longo prazo é dada por:

$$y_t^* = \frac{m}{(1-\alpha_1)} + \gamma x_t. \quad (125)$$

O ajustamento entre observado e desejado é dado por:

$$y_t - y_{t-1} = \phi(y_t^* - y_{t-1}) + \varepsilon_t \quad (126)$$

Usando (126) é possível explicitar a variável y_t^* em função de y_t e y_{t-1} temos:

$$y_t^* = \frac{1}{\phi} y_t + \left(\frac{\phi-1}{\phi}\right) y_{t-1} - \frac{1}{\phi} \varepsilon_t \quad (127)$$

e substituindo (127) em (125) temos:

$$\begin{aligned} \frac{1}{\phi} y_t + \left(\frac{\phi-1}{\phi}\right) y_{t-1} - \frac{1}{\phi} \varepsilon_t &= \frac{m}{(1-\alpha_1)} + \gamma x_t \\ \Downarrow \\ y_t &= \phi \left(\frac{m}{(1-\alpha_1)} \right) + (1-\phi) y_{t-1} + \phi \gamma x_t + \varepsilon_t \end{aligned} \quad (128)$$

3.6.1 Continuação do Exemplo

Foi estimado um modelo de ajustamento parcial para $\log(Cons_t)$ e $\log(Inc_t)$, isto é:

$$\log(Cons_t) = m + \beta_0 \log(Inc_t) + \alpha_1 \log(Cons_{t-1}) + \varepsilon_t \quad (129)$$

Temos o seguinte programa no R:

Ajustamento Parcial

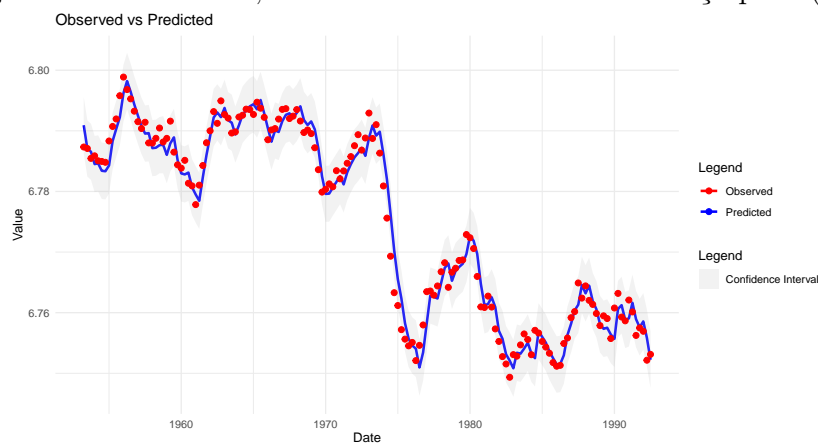
tem-se os seguintes resultados:

Tabela 9: Ajustamento Parcial

EQ(9) Ajustamento Parcial				
Desvios Padrões com correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
<i>Intercepto</i>	-0.3556	0.1610	-2.2095	0.0286
<i>log(Inc_t)</i>	0.2275	0.0681	3.3413	0.0010
<i>log(Const_{t-1})</i>	0.8244	0.0526	15.6572	0.1158
no of observations: 158				
no of parameters: 3				
$\sigma_\varepsilon = 0.00233$				
$R^2 = 0.9776$ $\bar{R}^2 = 0.9773$				
F-statistics(2,155)= 3383 p-value = 0.000				
$AIC = -9.254162$				
$BIC = -9.176628$				
log-likelihood = 734.0788				

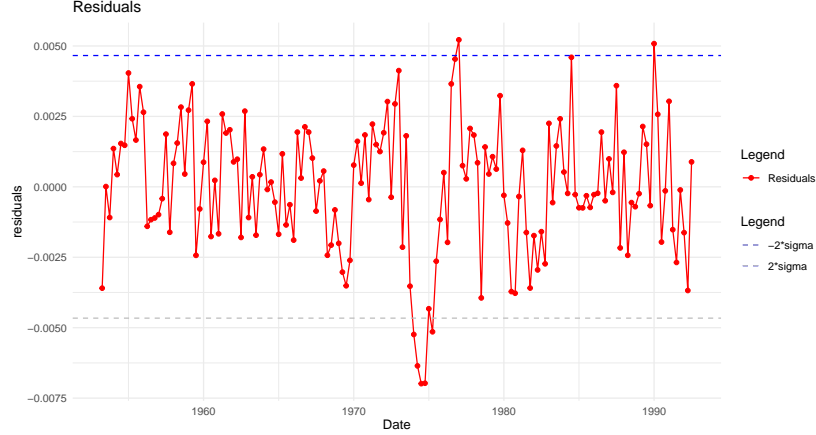
O gráfico dos observados, previstos e bandas é apresentado abaixo

Figura 11: Observado, Previsto e Intervalo de Confiança para (129)



O gráfico dos resíduos e respectivos Intervalos de Confianças é apresentado abaixo

Figura 12: Resíduos e Intervalo de Confiança para (129)



Comparando com os modelos anteriores este modelo é pior do que o modelo de taxa de crescimento.

3.7 Regressão Estática com Erros AR(1) ou Modelos de Fatores Comuns

Escrevendo (112) em operador defasagem temos:

$$(1 - \alpha_1 L)y_t = m + \beta_0(1 + \frac{\beta_1}{\beta_0}L)x_t + \varepsilon_t \quad (130)$$

Agora se e somente se $\alpha_1 = -\frac{\beta_1}{\beta_0}$ ou $\alpha_1\beta_0 + \beta_1 = 0$, os dois polinômios $(1 - \alpha_1 L)$ e $(1 + \frac{\beta_1}{\beta_0}L)$ tem a mesma raiz, isto é, tem um fator comum e, portanto, podemos dividir (130) por $(1 - \alpha_1 L)$ obtendo:

$$y_t = \frac{m}{(1 - \alpha_1)} + \beta_0 x_t + \frac{\varepsilon_t}{(1 - \alpha_1 L)} \quad (131)$$

e fazendo $\rho = \alpha_1$, podemos re-escrever (131) da seguinte forma:

$$\begin{aligned} y_t &= \frac{m}{(1 - \rho)} + \beta_0 x_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned} \quad (132)$$

que é o modelo de regressão estática com erros autorregressivos.

Observe que $(1 - \alpha_1 L)$ é um dos fatores do polinômio $A(L)$ e de forma semelhante $(1 + \frac{\beta_1}{\beta_0} L)$ é um dos fatores do polinômio $B(L)$, e quando eles são iguais os dois polinômios $A(L)$ e $B(L)$ tem um fator em comum.

Como (132) impõe uma restrição testável no modelo $ADL(1,1)$ se esta restrição for rejeita este sub-modelo é uma redução inválida do modelo geral. Desta forma os modelos ADL incluem os modelos de regressão estática com erros correlacionados.

Observe que a metodologia do particular para o geral, que parte da regressão estática e testa autocorrelação nos erros e caso esta hipótese não seja rejeitada corrigimos por autocorrelação os erros implica que a restrição de fatores comuns foi imposta e pode ser inválida. Mas observe que se estimarmos o modelo por M.Q.O. sem levar em consideração a autocorrelação dos erros os estimadores continuam consistente mas são ineficientes.

3.7.1 Estimação com Erros Correlacionados

Considere a estimação de (132) e observe que neste caso a matriz de variância-covariância dos erros é não escalar e será dada por:

$$Var(\mathbf{u}) = \sigma^2 \Omega \Rightarrow \Omega = \frac{1}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & . & . & . & \rho^{T-1} \\ \rho & 1 & \rho & . & . & . & \rho^{T-2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \rho^{T-2} & \rho^{T-3} & . & . & \rho & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & . & . & . & \rho & 1 \end{bmatrix} \quad (133)$$

e a estimação dos parâmetros é obtida por Mínimos Quadrados Generalizadas (M.Q.G.), isto é

$$\hat{\beta}_0 = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \quad (134)$$

Observe que em (134), implicitamente, estamos assumindo que ρ é conhecido o que não é verdadeiro na prática. Se $\hat{\rho}$ é um estimador para ρ o estimador $\hat{\beta}_0$ será dado por $(X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y$ onde $\hat{\Omega}$ é dada por:

$$\hat{\Omega} = \frac{1}{1 - \hat{\rho}^2} \begin{bmatrix} 1 & \hat{\rho} & \hat{\rho}^2 & . & . & . & \hat{\rho}^{T-1} \\ \hat{\rho} & 1 & \hat{\rho} & . & . & . & \hat{\rho}^{T-2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \hat{\rho}^{T-2} & \hat{\rho}^{T-3} & . & . & \hat{\rho} & 1 & \hat{\rho} \\ \hat{\rho}^{T-1} & \hat{\rho}^{T-2} & . & . & . & \hat{\rho} & 1 \end{bmatrix} \quad (135)$$

Um estimador possível para ρ é

$$\hat{\rho} = \frac{\sum_{t=1}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_{t-1}^2}$$

onde $\hat{u}_t = y_t - \hat{\beta}_0 x_t$.

Observe que o estimador de β_0 depende do estimador de ρ e vice-versa.

Uma outra forma de obter estes estimadores é através de Mínimos Quadrados Ponderados uma vez que a inversa da matriz Ω tem forma fechada e é dada por:

$$\Omega^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & \dots & 0 & 0 & 0 \\ . & . & . & \dots & . & . & . \\ . & . & . & \dots & . & . & . \\ 0 & 0 & 0 & \dots & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{bmatrix} \quad (136)$$

que pode ser escrita como $\Omega^{-1} = P'P$ onde P é dada por:

$$P = \begin{bmatrix} \sqrt{1 - \rho^2} & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 & 0 \\ . & . & . & \dots & . & . & . \\ . & . & . & \dots & . & . & . \\ 0 & 0 & 0 & \dots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{bmatrix} \quad (137)$$

que implica na seguinte transformação dos dados:

$$y^* = \begin{bmatrix} \sqrt{1 - \rho^2} \cdot y_1 \\ y_2 - \rho y_1 \\ \cdot \\ \cdot \\ y_T - \rho y_{T-1} \end{bmatrix} \quad (138)$$

e

$$x^* = \begin{bmatrix} \sqrt{1 - \rho^2} \cdot x_1 \\ x_2 - \rho x_1 \\ \cdot \\ \cdot \\ x_T - \rho x_{T-1} \end{bmatrix} \quad (139)$$

e usando Mínimos Quadrados Ordinários no modelo transformado temos o estimador equivalente ao M.Q.G.

Observe que se a primeira linha de P não for utilizada o estimador de β_0 é facilmente obtido, uma vez que será M.Q.O. para $t = 2, \dots, T$. Observe também que este estimador será aproximado uma vez que a primeira observação está sendo descartada.

Como observado acima ρ é desconhecido mas pode ser estimado por Mínimos Quadrados Não Lineares no seguinte modelo, que é uma transformação de (132):

$$y_t = m + \rho y_{t-1} + \beta_0 x_t - \rho \beta_0 x_t + \varepsilon_t \quad (140)$$

assumindo que $\varepsilon_t \sim NI(0, \sigma^2)$.

Atualmente é fácil estimar (140) por MQNL mas antigamente um procedimento iterativo era utilizado. Este procedimento consistia em:

- (i) assumindo que $\rho = 0$ o estimador consistente para β_0 era dado por MQO da regressão de y_t em constante e x_t ;
- (ii) obtinhamos o resíduo da regressão do item (i), denotando-se por \hat{u}_t ;
- (iii) fazendo-se a regressão de \hat{u}_t em \hat{u}_{t-1} , era obtido uma estimativa para ρ , i.e., $\hat{\rho}$;

- (iv) transformado-se as variáveis do modelo usando-se (138) e (139) obtemos uma nova estimativa para β_0 , i.e. $\hat{\beta}_0$;
- (v) repetimos os itens (ii-iv) até que a mudança nos parâmetros sejam pequenas.

3.7.2 Testes para Autocorrelação - o teste de Durbin & Watson

Consider o modelo dado por (132) e desejamos testar:

$$H_0 : \rho = 0 \text{ versus } H_0 : \rho > 0 \quad (141)$$

Como a hipótese é a respeito do processo gerador de u_t que é não observado podemos usar o resíduo de M.Q.O., isto é, $\hat{u}_t = y_t - \hat{\beta}_0 x_t$ onde $\hat{\beta}_0$ é o Estimador de M.Q.O. de β_0 .

Sabemos que:

- (i) $\hat{u}_t = M_X u_t$
- (ii) $Var(\hat{\mathbf{u}}) = \sigma^2 M_X$

Logo mesmo que se a hipótese de que $Var(\mathbf{u}) = \sigma^2 \mathbf{I}$ seja válida, os resíduos de M.Q.O. apresentarão autocorrelação.

Além do mais M_X depende de x_t fazendo com que um teste exato válido para qualquer X seja impossível de ser obtido.

Durbin and Watson (1950) e Durbin and Watson (1951) propõem a seguinte estatística de teste:

$$d = \frac{\sum_{t=1}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_{t-1}^2} \quad (142)$$

Observe que a estatística d está relacionada ao coeficiente de autocorrelação amostral de primeira ordem, pois:

$$d \approx 2(1 - \hat{\rho}) \quad (143)$$

onde $\hat{\rho}$ é o Estimador de M.Q.O. de ρ que é dado por

$$\hat{\rho} = \frac{\sum_{t=1}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_{t-1}^2} \quad (144)$$

Como o cálculo da estatística d depende da matriz de dados X , em Durbin and Watson (1950) e Durbin and Watson (1951) são obtidos limites inferiores (d_L) e superiores (d_U) para esta estatística de tal forma que:

- (i) se $d < d_L$ rejeitamos a nula;
- (ii) se $d > d_U$ não rejeitamos a nula; e
- (iii) se $d_L < d < d_U$ o teste é inconclusivo

O procedimento acima vale quando $d < 2$. Quando $d > 2$ a alternativa do teste é de autocorrelação negativa e o teste é feito usando-se $4 - d$.

O teste de Durbin & Watson só é válido para:

- (i) Correlação de primeira ordem do tipo autoregressiva;
- (ii) Os regressores não incluem variáveis endógenas defasadas;
- (iii) É necessário a inclusão da constante na regressão restrita, i.e. aquela que assume que não existe autocorrelação

Um dos problemas com este teste é que o resultado do teste pode indicar a região inconclusiva. Uma solução, conservadora é usar d_U como verdadeiro valor crítico, i.e. se $d < d_U$ então rejeitamos a nula.

3.7.3 Programa no R para gerar distribuição da estatística d de Durbin & Watson

O programa abaixo estima um modelo de regressão estática com erros $AR(1)$, i.e.

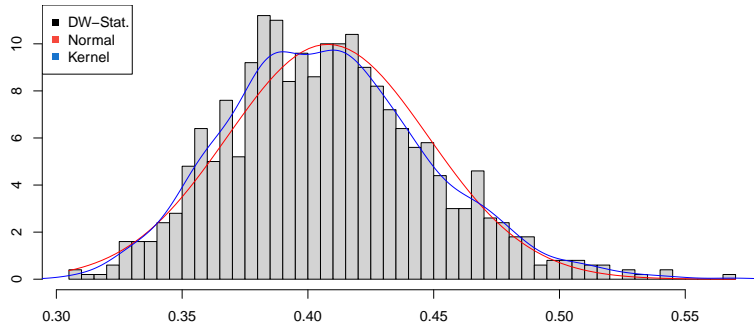
$$\begin{aligned}y_t &= 1.0 + 0.5x_t + ut \\u_t &= 0.8u_{t-1} + \varepsilon_t\end{aligned}\tag{145}$$

e obtém a distribuição da estatística d para 1000 replicações

Distribuição da Estatística d de Durbin & Watson

Abaixo é apresentada a densidade para a estatística d

Figura 13: Histograma da Estatística DW onde o P.G.D. é (145)



Observe que tanto a média, que é igual a 0.4058, quanto a mediana, que é igual a 0.4079 estão próximas do valor verdadeiro que é 0.4.

3.7.4 Teste de Autocorrelação de quarta ordem

Quando a frequência dos dados é trimestral, podemos ter correlação entre trimestres de diferentes anos, isto é,

$$u_t = \rho_4 u_{t-4} + \varepsilon_t\tag{146}$$

O teste neste caso é dado por:

$$H_0 : \rho_4 = 0 \text{ versus } H_0 : \rho_4 > 0\tag{147}$$

Wallis (1972) propôs a seguinte modificação da estatística de Durbin & Watson

$$d = \frac{\sum_{t=1}^T (\hat{u}_t - \hat{u}_{t-4})^2}{\sum_{t=1}^T \hat{u}_{t-4}^2} \quad (148)$$

e os valores críticos foram tabelados por Wallis (1972), sendo que estes valores são diferentes caso a especificação inclua ou não dummies sazonais.

3.7.5 Teste h de Durbin

Quando o modelo inclui uma variável endógena defasada como regressor o teste de Durbin & Watson não é válido.

Durbin (1970) propôs a seguinte modificação que foi chamada na literatura de teste h de Durbin.

Considere a seguinte especificação

$$y_t = \beta_1 y_{t-1} + \dots + \beta_r y_{t-r} + \beta_{r+1} x_{1t} + \dots + \beta_{r+s} x_{st} + u_t \quad (149)$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (150)$$

Sob a hipótese nula $H_0 : \rho = 0$, Durbin (1970) mostrou que a estatística dada por:

$$h = \hat{\rho} \sqrt{\frac{T}{1 - TVar(\hat{\beta}_1)}} \quad (151)$$

tem distribuição $N(0, 1)$, onde T é o tamanho da amostra, $Var(\hat{\beta}_1)$ é o estimador da variância do coeficiente de y_{t-1} em (149), $\hat{\rho}$ é o E.M.Q.O. de ρ dado por $\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_{t-1}^2}$ e \hat{u}_t é o resíduo de M.Q.O. de (149).

A implementação do teste é feita usando-se o seguinte procedimento:

- (i) Estime (149) por M.Q.O. e obtenha $Var(\hat{\beta}_1)$;
- (ii) Usando os resíduos de M.Q.O. estime $\hat{\rho}$ ou se a estatística DW foi calculada use aproximação $\hat{\rho} = 1 - d/2$;

- (iii) Calcule h por (151) e se $h > 1.65$ rejeita a nula usando um nível de significância de 5%.
- (iv) Caso h seja negativo o valor crítico será -1.65 .

Quando $TVar(\hat{\beta}_1) > 1$, Durbin (1970) propôs a seguinte modificação:

- (i) estime (149) por M.Q.O. e obtenha o resíduo \hat{u}_t ;
- (ii) estime por M.Q.O. a seguinte regressão de \hat{u}_t em $\hat{u}_{t-1}, y_{t-1}, \dots, y_{t-r}, x_{1t}, \dots, x_{st}$;
- (iii) se o coeficiente de \hat{u}_{t-1} é significativamente diferente de zero, usando o teste t usual, rejeite a nula H_0

3.7.6 Teste de autocorrelação de Breusch-Godfrey

Considere a seguinte especificação:

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad (152)$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (153)$$

com $|\rho| < 1$ e $\varepsilon_t \sim NI(0, \sigma^2)$.

Substituindo (153) em (152) temos:

$$y_t = \beta_0(1 - \rho) + \beta_1 x_t + \rho y_{t-1} - \beta_1 \rho x_{t-1} + \varepsilon_t \quad (154)$$

e desejamos testar

$$H_0 : \rho = 0 \text{ versus } H_0 : \rho \neq 0 \quad (155)$$

Observe que (154) é não linear nos parâmetros, mas sob H_0 reduz-se a um modelo linear nos parâmetros. Devemos, então, usar um teste do tipo *LM*.

Observe que o termo da soma dos quadrados na função de log-verossimilhança é dada por:

$$-\frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2 \quad (156)$$

O escore é dado por:

$$s(\beta) = \frac{\partial \ell}{\partial \beta} = -\frac{1}{\sigma^2} \sum_{t=1}^T \varepsilon_t \frac{\partial \varepsilon_t}{\partial \beta} = \frac{1}{\sigma^2} \sum_{t=1}^T \varepsilon_t w_t \quad (157)$$

onde $w_t = -\frac{\partial \varepsilon_t}{\partial \beta}$.

Temos que a matriz de informação é dada por:

$$\begin{aligned} I(\beta) &= E[s(\beta)s'(\beta)] \\ &= E \left[\frac{1}{\sigma^4} \left(\sum_{t=1}^T \varepsilon_t w_t \right) \left(\sum_{t=1}^T \varepsilon_t w_t \right)' \right] \\ &= E \left[\frac{1}{\sigma^4} \left(\sum_{t=1}^T \varepsilon_t^2 w_t w_t' + \sum_{t \neq s} \varepsilon_t \varepsilon_s w_t w_s' \right) \right] \\ &= \frac{1}{\sigma^2} E \left[\left(\sum_{t=1}^T w_t w_t' \right) \right] \end{aligned}$$

A estatística LM é dada por:

$$LM = \frac{1}{\tilde{\sigma}^2} \left(\sum_{t=1}^T \tilde{\varepsilon}_t \tilde{w}_t \right)' \left(\sum_{t=1}^T \tilde{w}_t \tilde{w}_t' \right) \left(\sum_{t=1}^T \tilde{\varepsilon}_t \tilde{w}_t \right) \quad (158)$$

onde o \sim indica estimador restrito.

Sabemos que $LM = nR^2$ onde R^2 é o coeficiente de determinação da regressão de $\tilde{\varepsilon}_t$ em \tilde{w}_t e a distribuição desta estatística é \aleph_1 .

3.7.7 Continuação do Exemplo

Foi estimado um modelo de regressão com erros correlacionados para $\log(Cons_t)$ e $\log(Inc_t)$, isto é:

$$\log(Cons_t) = \beta_0 + \beta_1 \log Inc_t + u_t \quad (159)$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (160)$$

Os comandos no R estão apresentados abaixo

Regressão Estática com Erros $AR(1)$

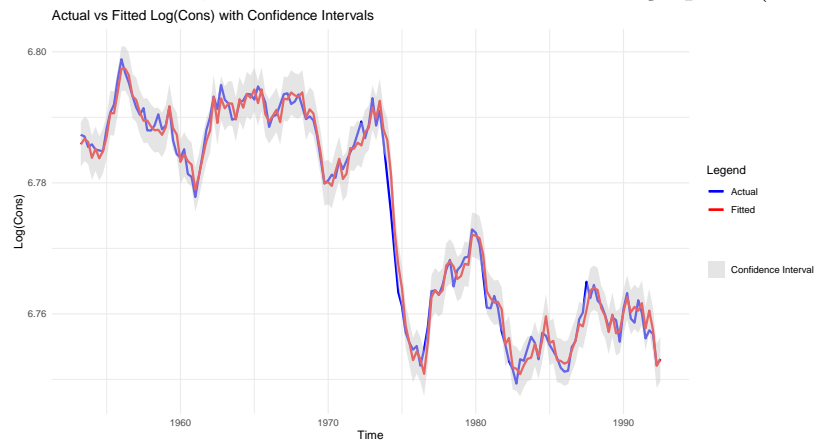
tem-se os seguintes resultados:

Tabela 10: Regressão Estática com Erros $AR(1)$

EQ10) Regressão Estática com Erros $AR(1)$				
Desvios Padrões com correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
$Intercepto$	3.3372	0.2477	13.7400	0.0000
$log(Inc_t)$	0.5058	0.0365	13.8661	0.0000
$AR(1)$	0.9864	0.0111	88.52377	0.0000
no of observations: 158				
no of parameters: 3				
$\sigma_\varepsilon = 2.875504e - 06$				
$R^2 = 0.9879896$		$\overline{R}^2 = 0.9882176$		
F-statistics(2,155)= 6242 p-value = 0.000				
$AIC = -9.8797$				
$BIC = -9.7254$				
log-likelihood = 783.5				

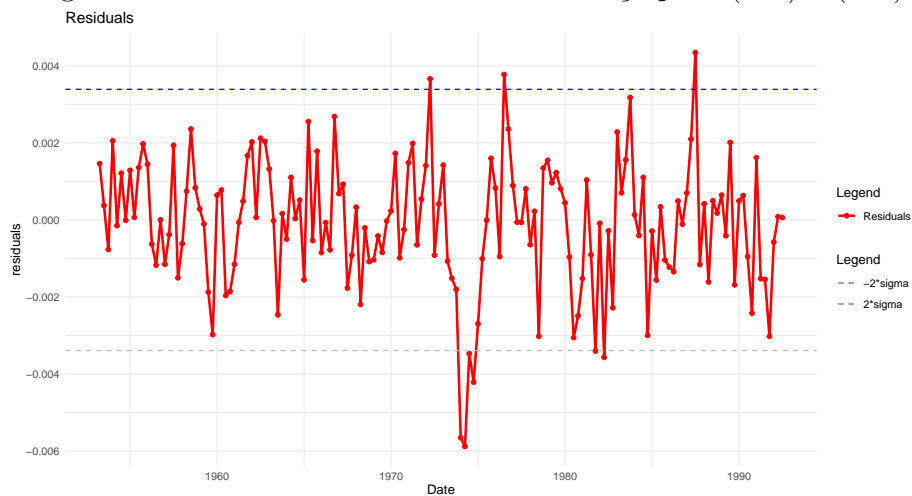
O gráfico dos observados, previstos e bandas é apresentado abaixo

Figura 14: Observado, Previsto e Intervalo de Confiança para (159) e (160)



O gráfico dos resíduos e respectivos Intervalos de Confianças é apresentado abaixo

Figura 15: Resíduos e Intervalo de Confiança para (159) e (160)



Observe que a estimativa do parâmetro autorregressivo é muito próxima da unidade. Pelo critério de Schwarz este modelo é melhor do que o de taxa de crescimento.

3.8 Forma Reduzida

O processo gerador dos dados é dado por:

$$y_t = m + \alpha_1 y_{t-1} + \beta_1 x_{t-1} + \varepsilon_t \quad (161)$$

e neste modelo a informação corrente de x é excluída.

Observe que (161) pode ser interpretado como uma das equações de um Vetor Autorregressivo e não estamos especificando a outra equação. Deste modo se o processo gerador de x_t não tiver parâmetros constantes, os parâmetros de (161) também não ser ao constantes.

3.8.1 Continuação do Exemplo

Foi estimado um modelo de forma reduzida para $\log(Cons_t)$ e $\log(Inc_t)$, isto é:

$$\log(Cons_t) = m + \alpha_1 \log(Cons_{t-1}) + \beta_1 \log(Inc_{t-1}) + \varepsilon_t \quad (162)$$

Os comandos no R estão apresentados abaixo:

Forma Reduzida

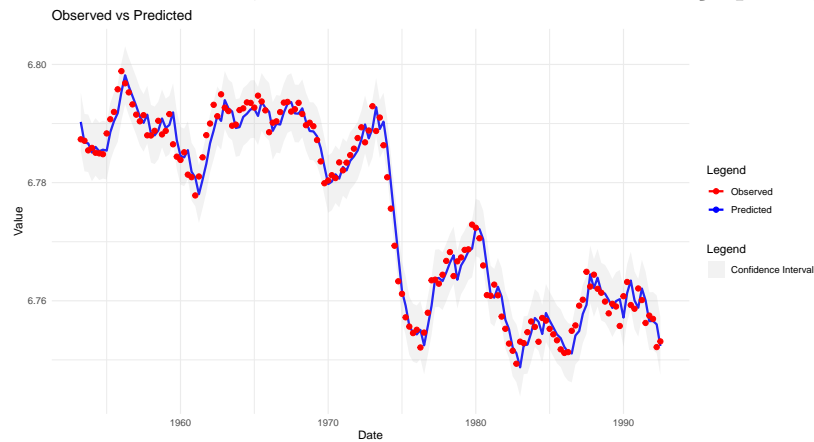
tem-se os seguintes resultados:

Tabela 11: Forma Reduzida

EQ11) Forma Reduzida				
Desvios Padrões com correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
<i>Intercepto</i>	0.3688	0.1467	2.5139	0.0130
<i>log(Cons_{t-1})</i>	1.1099	0.0502	22.1095	0.0000
<i>log(Inc_{t-1})</i>	+0.1640	0.0647	-2.5350	0.01224
no of observations: 158				
no of parameters: 3				
$\sigma_\varepsilon = 0.002447$				
$R^2 = 0.9753$ $\overline{R}^2 = 0.9750$				
F-statistics(2,155)= 3060 p-value = 0.000				
AIC =-9.156165				
BIC = -9.078631				
log-likelihood = 726.337				

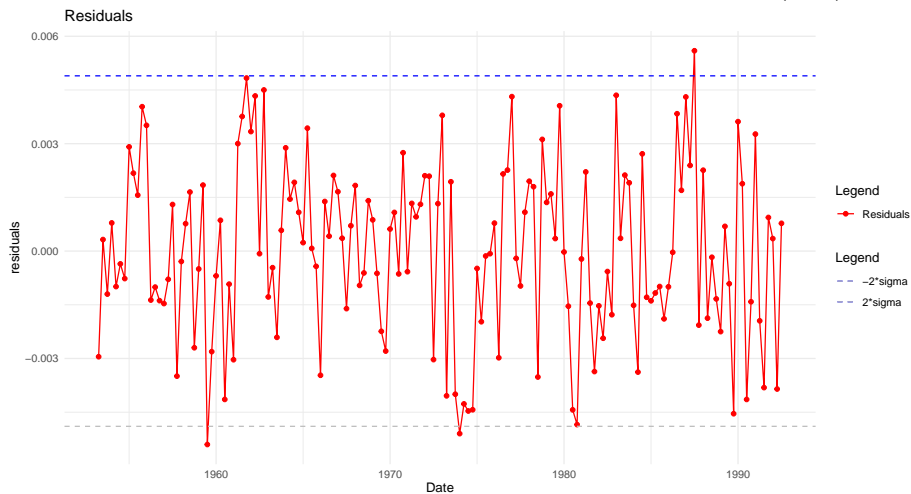
O gráfico dos observados, previstos e bandas é apresentado abaixo

Figura 16: Observado, Previsto e Intervalo de Confiança para (162)



O gráfico dos resíduos e respectivos Intervalos de Confianças é apresentado abaixo

Figura 17: Resíduos e Intervalo de Confiança para(162)



Pelo critério de Schwarz este modelo é pior do que a regressão estática com erros correlacionados.

3.9 Modelo de Correção de Erros Homogeneo

O processo gerador dos dados é:

$$\Delta y_t = m + \beta_0 \Delta x_t + (\alpha_1 - 1)(y_{t-1} - x_{t-1}) + \varepsilon_t \quad (163)$$

Observe que neste modelo temos variáveis $I(0)$, Δy_t e Δx_t e variáveis $I(1)$, y_{t-1} e x_{t-1} .

Em geral combinação linear de variáveis $I(1)$ resulta numa variável $I(1)$, mas sob certas condições como veremos no exemplo abaixo esta combinação pode resultar numa variável $I(0)$ e o termo $(y_{t-1} - \kappa x_{t-1})$ pode ser $I(0)$ e portanto todas as variáveis em (163) são $I(0)$ valendo inferência usual.

3.9.1 Exemplo - Cointegração

Considere o caso em que y é combinação linear de duas variáveis $I(1)$, isto é

$$\begin{aligned} y_t &= ax_t + bz_t \\ x_t &= x_{t-1} + \eta_t \\ z_t &= z_{t-1} + v_t \end{aligned}$$

onde a e b são constantes arbitrárias, η_t e v_t são ruídos brancos.

Substituindo a segunda e terceira equação na primeira temos

$$y_t = y_{t-1} + (a\eta_t + bv_t)$$

e portanto y também é $I(1)$.

Agora considere o modelo *ADL* na sua forma MCE irrestrito, isto é (11)

$$\Delta y_t = \beta_0 \Delta x_t - (1 - \alpha_1)[y_{t-1} - a - \gamma x_{t-1}] + \varepsilon_t$$

e como estamos assumindo que tanto y quanto x são $I(1)$ temos que Δy_t e Δx_t são $I(0)$.

Agora seja $z_t = y_t - a - \gamma x_t$, e subtraindo de ambos os lados de (112) a quantidade $a + \gamma x_t$, obtém-se:

$$\begin{aligned}
y_t - a - \gamma x_t &= m + \alpha_1 y_{t-1} - a - \gamma x_t + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \\
z_t &= \alpha_1 y_{t-1} - \frac{m\alpha_1}{1 - \alpha_1} - \alpha_1 \frac{\beta_0 + \beta_1}{1 - \alpha_1} x_{t-1} + \alpha_1 \frac{\beta_0 + \beta_1}{1 - \alpha_1} x_{t-1} \\
&\quad - \frac{\beta_0 + \beta_1}{1 - \alpha_1} x_t + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \\
z_t &= \alpha_1 z_{t-1} - \frac{\beta_0 \alpha_1 + \beta_1}{1 - \alpha_1} x_t + \frac{\beta_0 \alpha_1 + \beta_1}{1 - \alpha_1} x_{t-1} + \varepsilon_t \\
z_t &= \alpha_1 z_{t-1} + \varepsilon_t - \frac{\beta_0 \alpha_1 + \beta_1}{1 - \alpha_1} \eta_t
\end{aligned}$$

onde η_t é o erro do passeio aleatório para x_t . Se $|\alpha_1| < 1$ então z_t será estacionário e neste caso dizemos que y_t e x_t cointegram com vetor de cointegração dado por $\theta = (1, -a, -\gamma)$.

Observe que variáveis $I(1)$ tem tendência estocástica, isto é, tendem a crescer mas se as duas variáveis co-integram elas tendem a crescer junto uma vez que uma combinação linear delas é um processo estacionário com média zero, isto é um processo que reverte para zero frequentemente.

Usando a definição de z_t podemos reescrever o modelo de correção de erros da seguinte forma:

$$\Delta y_t = \beta_0 \Delta x_t - (1 - \alpha_1) z_{t-1} + \varepsilon_t \quad (164)$$

Observe que na especificação acima todas as variáveis são $I(0)$ e inferência usual vale e podemos recuperar os parâmetros β_0 e α_1 . Para recuperar o parâmetro β_1 é necessário reescrever (164) de uma outra forma, mas não será possível recuperar todos os parâmetros uma vez que não é possível especificar todas as variáveis como $I(0)$.

3.9.2 Estimação e Teste da equação de cointegração

Dada a especificação do modelo ADL, como se pode estimar a relação de cointegração? Como se poderia testar se existe uma relação de cointegração?

A primeira sugestão existente na literatura, o Procedimento de Engle and Granger (1987) que consiste em estimar a regressão de y_t em constante e x_t e os resíduos desta equação seriam uma estimativa para o termo de correção de erros z_t .

Para se verificar se existe relação de cointegração devemos testar se \hat{z}_t é $I(0)$, Os testes convencionais de raiz unitária não podem ser usados uma vez que estamos usando uma variável estimada. Valores críticos para o teste de raiz unitária neste caso foram obtidos por MacKinnon (1991):

Tabela 12: Valor Críticos para teste de raiz unitária dos resíduos do primeiro passo do procedimento de Engle and Granger (1987)

k'	<i>Estatística de Teste</i>	1%	5%	10%
2	c	-3.90	-3.34	-3.04
	ct	-4.32	-3.78	-3.50
3	c	-4.29	-3.74	-3.45
	ct	-4.66	-4.12	-3.84
4	c	-4.64	-4.10	-3.81
	ct	-4.97	-4.43	-4.15
5	c	-4.96	-4.42	-4.13
	ct	-5.25	-4.72	-4.43
6	c	-5.25	-4.71	-4.42
	ct	-5.52	-4.98	-4.70
Nota: k' é o número de variáveis na relação de cointegração; c e ct indicam se constante e tendência determinística estão presentes na relação de cointegração.				

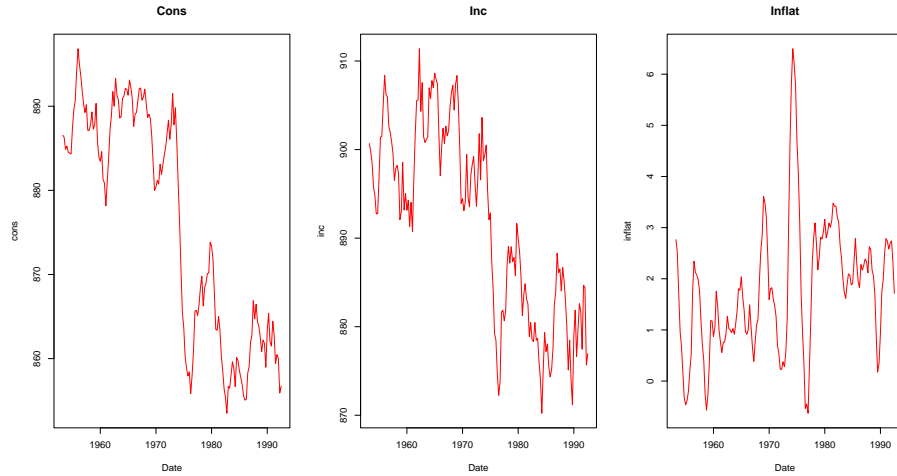
3.10 Exemplo - DATAGIVE

O arquivo DATAGIVE.XLS contém informações das variáveis:

- (i) Cons - consumo para o período de 1953:1 até 1992:3 para a economia inglesa.
- (ii) Inc - renda para o mesmo período
- (iii) Inflat - inflação
- (iv) Output - produto

Os gráficos abaixo apresentam as três primeiras variáveis

Figura 18: Cons, Inc e Inflat



Observe que todas as três séries tem uma quebra no final de 1973, devido ao primeiro choque do petróleo.

A tabela abaixo apresenta a matriz de correlação entre as três variáveis e as defasagens de ordem 1. Observe que existem várias correlações superiores a 0.90 o que pode implicar em multicolinearidade

Tabela 13: Matriz de Correlação entre Cons, Inc e Inflat e suas defasagens

CONS	CONS1	INC1	INC	INFLAT	INFLAT1
1.0000	0.9866	0.9143	0.9421	-0.3511	-0.4173
0.9866	1.0000	0.9414	0.9309	-0.2756	-0.3404
0.9143	0.9414	1.0000	0.9510	-0.0733	-0.1124
0.9421	0.9309	0.9510	1.0000	-0.1270	-0.1953
-0.3511	-0.2756	-0.0733	-0.1124	1.0000	0.9265
-0.4173	-0.3404	-0.1124	-0.1953	0.9265	1.0000

Suponha que você deseja começar usando um ADL(1,1) entre $\log(Cons_t)$ e $\log(Inc_t)$, isto é:

$$\log(Cons_t) = m + \alpha_1 \log(Cons_{t-1}) + \beta_0 \log(Inc_t) + \beta_1 \log(Inc_{t-1}) + \varepsilon_t \quad (165)$$

O programa abaixo estima o modelo (165):

$$\text{MCE } \log(Cons) \text{ e } \log(Inc)$$

A tabela abaixo apresenta os resultados deste ADL(1,1)

Tabela 14: ADL(1,1)

EQ12) ADL(1,1)				
Desvios Padrões com correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
<i>Intercepto</i>	0.0212	0.1377	0.154	0.8778
<i>log(Cons_{t-1})</i>	0.9863	0.0441	22.357	0.0000
<i>log(Inc_t)</i>	0.5046	0.0308	16.384	0.01224
<i>log(Inc_{t-1})</i>	-0.4941	0.0488	-10.127	0.01224
no of observations: 158				
no of parameters: 4				
$\sigma_{\varepsilon} = 0.001693$				
$R^2 = 0.9883$		$\bar{R}^2 = 0.988$		
F-statistics(3,154)= 4319 p-value = 0.000				
$AIC = -9.886751$				
$BIC = -9.789834$				
log-likelihood =785.0534				

Através do gráfico do observado e do previsto assim como do gráfico dos resíduos, vide abaixo, podemos observar que a especificação acompanha os movimentos de $\log(Cons_t)$ e que os resíduos parecem adequados embora a quebra da primeira crise do petróleo não tenha sido totalmente controlada.

Figura 19: Observado, Previsto e Intervalos de Confianças para (165)

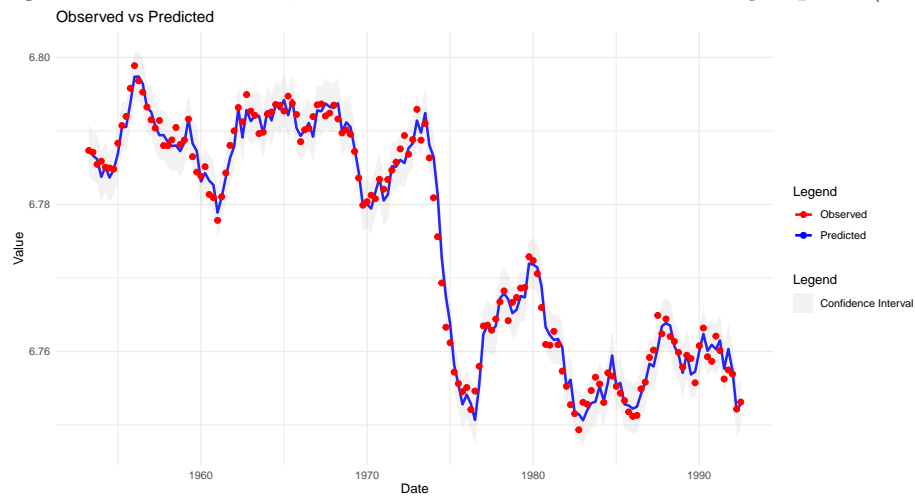
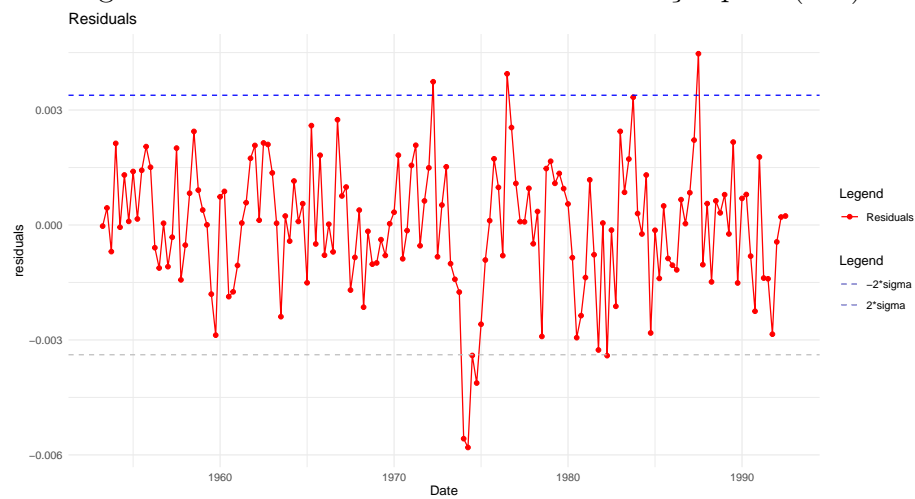


Figura 20: Resíduos e Intervalos de Confianças para (165)



Mas ao se testar se a soma dos coeficientes é um usando um teste de Wald obtém-se o seguinte resultado:

Tabela 15: Elasticidade de Longo Prazo bem definida

Wald Test	Value	p-value
$H_0 : \alpha_1 = 1, \beta_0 = -\beta_1$		
F-statistics: ($F(2, 152)$)	0.2764	0.7589
Chi-square: (χ^2_2)	0.5528	0.7585

que implica em não rejeição da nula.

Observe que esta hipótese implica que existe um fator comum nos polinômios autorregressivo e de defasagens distribuídas e este fator comum é igual a um, isto é, ambos os polinômios tem raiz unitária.

Desta forma a solução de longo-prazo não está definida uma vez que $\frac{\beta_0 + \beta_1}{1 - \alpha_1} = \frac{0}{0}$.

Observe que como a elasticidade de longo prazo não está bem definida, não faz sentido testar homogeneidade, isto é, $\alpha_1 + \beta_0 + \beta_1 = 1$, uma vez que a não rejeição da nula do teste anterior induz a não rejeição da nula no teste de homogeneidade.

Mas os testes apresentados acima são válidos? A especificação omite variáveis relevantes? Caso afirmativo sabemos que os estimadores dos coeficientes são inconsistentes e todos os testes são inválidos.

A seguir é apresentado o teste de omissão da variável *Inflat* e *Inflta(-1)* cujo resultado indica que esta variável é importante.

Vamos primeiro estimar o modelo irrestrito, isto é:

$$\begin{aligned} \log(Cons_t) &= m + \alpha_1 \log(Cons_{t-1}) + \beta_0 \log(Inc_t) + \beta_1 \log(Inc_{t-1}) \\ &+ \gamma_0 Inflat_t + \gamma_1 Inflat_{t-1} + \varepsilon_t \end{aligned} \quad (166)$$

O programa abaixo estima o modelo (166):

$$\text{MCE } \log(Cons), \log(Inc) \text{ e } Inflat$$

A tabela abaixo apresenta os resultados deste ADL(1,1,1) dado por (166):

Tabela 16: ADL(1,1,1)

EQ13) ADL(1,1,1)				
Desvios Padrões com correção por HAC				
The estimation sample is: 2 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
<i>Intercepto</i>	-0.1718	0.0621	-2.7641	0.0064
<i>log(Cons_{t-1})</i>	0.7992	0.02652	30.1258	0.0000
<i>log(Inc_t)</i>	0.5079	0.0264	19.2089	0.0000
<i>log(Inc_{t-1})</i>	-0.2820	0.0364	-7.7466	0.0000
<i>Inflat_t</i>	-0.00089	0.00018	-4.8816	0.0000
<i>Inflat_{t-1}</i>	-0.00030	0.00021	-1.4256	0.1560
no of observations: 158				
no of parameters: 6				
$\sigma_{\varepsilon} = 0.001237$				
$R^2 = 0.9938$		$\bar{R}^2 = 0.9936$		
F-statistics(5,151)= 4886 p-value = 0.000				
$AIC = -10.50308$				
$BIC = -10.3674$				
log-likelihood = 835.7434				

Mas ao se testar se os coeficientes $\gamma_0 = 0$ e $\gamma_1 = 0$, que corresponde ao teste de omissão das variáveis $Inflat_t$ e $Inflat_{t-1}$ obtém-se o seguinte resultado:

Tabela 17: Omissão das variáveis $Inflat_t$ e $Inflat_{t-1}$

Wald Test	Value	p-value
$H_0 : \gamma_0 = \gamma_1 = 0$		
F-statistics: ($F(2, 152)$)	68.37	0.0000
Chi-square: (χ^2_2)	136.74	0.0000

Implicando que a especificação correta é o modelo (166).

O gráfico do observado e previsto a especificação acompanha ops movimentos de $log(Cons_t)$ e pelo gráfico dos resíduos o problema do grande outlier no primeira crise do petróleo desaparece.

Figura 21: Observado, Previsto e Intervalos de Confianças para (166)

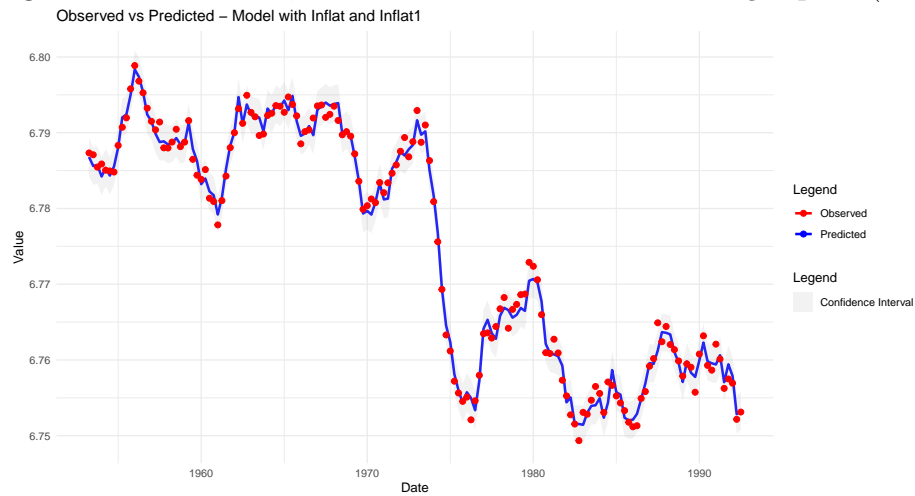
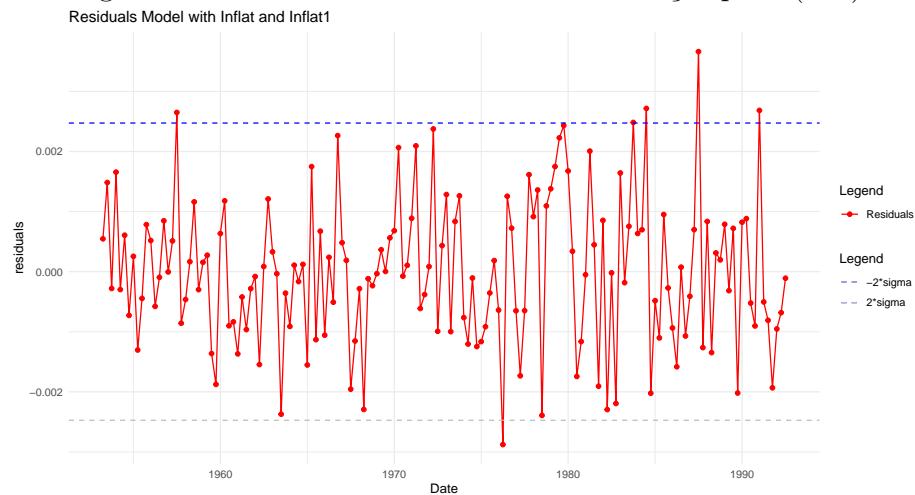


Figura 22: Resíduos e Intervalos de Confianças para (166)



Devemos testar se as coeficiente de longo prazo de $\log(Inc_t)$ e de $INFLAT_t$ está bem definido. Usando um teste de Wald para cada um dos coedicientes obtém-se os seguintes resultados:

Tabela 18: Elasticidade de Longo Prazo bem definida no ADL(1,1,1) dado por (166)

Wald Test	Value	p-value
$H_0 : \alpha_1 = 1, \beta_0 = -\beta_1$		
F-statistics: $(F(2, 152))$	28.125	0.0000
Chi-square: (χ^2_2)	56.25	0.0000

Tabela 19: Elasticidade de Longo Prazo bem definida no ADL(1,1,1) dado por (166)

Wald Test	Value	p-value
$H_0 : \alpha_1 = 1, \gamma_0 = -\gamma_1$		
F-statistics: $(F(2, 152))$	61.21	0.0000
Chi-square: (χ^2_2)	122.42	0.0000

que rejeita estas hipóteses, isto é, embora as variáveis seja não estacionárias, a relação de longo-prazo, agora, estará bem definida e será dada por

$$\log(Cons_t) = -0.8553073 + 1.12475 * \log(Inc_t) - 0.00593094 * Infla_t$$

E fazendo um teste de raíz unitária para

$$CI = \log(Cons_t) + 0.8553073 - 1.12475 * \log(Inc_t) + 0.00593094 * Infla_t \quad (167)$$

obtem-se os seguintes resultados (usou-se tendência determinística na especificação)

Tabela 20: ADF Test for CI

ADF test para CI = -4.62		
1% Critical Value	5% Critical Value	10% Critical Value
-3.99(-4.66)	-3.43(-4.12)	-3.13(-3.84)

Onde os valores entre-parenteses são os valores críticos corretos conforme tabela acima. Observe que se rejeita a 5% e a 10% mas não a 1%.

Estimou-se então o seguinte modelo de correção do equilíbrio, onde o termo de correção de equilíbrio é dado em (167):

$$\Delta \log(Cons_t) = \beta_0 \Delta \log(Inc_t) + \gamma_0 \Delta(Inflat_t) - (1 - \alpha_1) * CI_{t-1} + \varepsilon_t \quad (168)$$

Obtendo-se os seguintes resultados

Tabela 21: Modelo EqCM dado por (168)

EQ14) Modelo EqCM dado por (168)				
Desvios Padrões com correção por HAC				
The estimation sample is: 3 - 159				
	Coefficient	Std. Error	t-Statistics	p-value
$\Delta \log(Inc_t)$	0.4163	0.03727	11.170	0.0000
$\Delta(Inflat_t)$	-4.2159e-05	2.4718e-05	-0.1706	0.8648
$C I_{t-1}$	-0.1620	0.003088	-5.2457	0.0000
no of observations: 157				
no of parameters: 3				
$\sigma_\varepsilon = 0.001481$				
$R^2 = 0.6596$		$\bar{R}^2 = 0.6529$		
F-statistics(3,154)= 99.45 p-value = 0.000				
$AIC = -10.09615$				
$BIC = -10.01878$				
log-likelihood = 800.5959				

O gráfico do observado e previsto é apresentado abaixo assim como o dos resíduos. O previsto acompanha o observado e os resíduos parecem satisfazer as propriedades usuais.

Figura 23: Observado, Previsto e Intervalos de Confianças para (168)

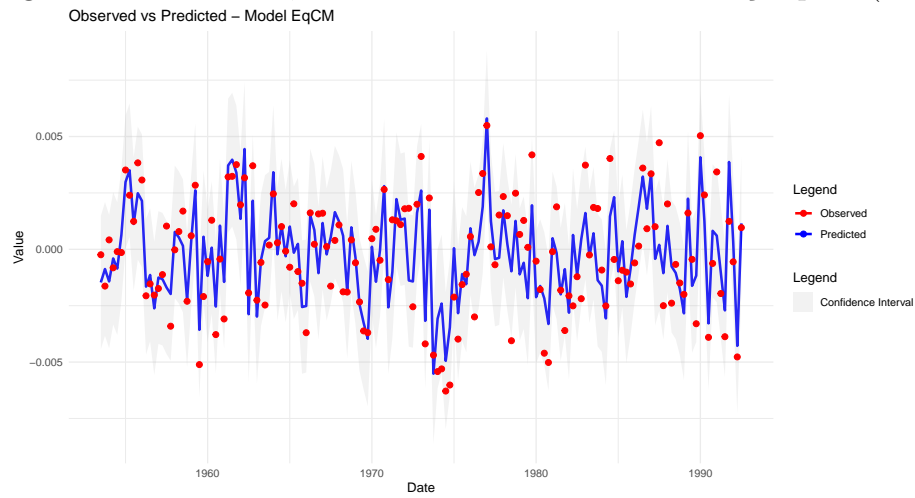
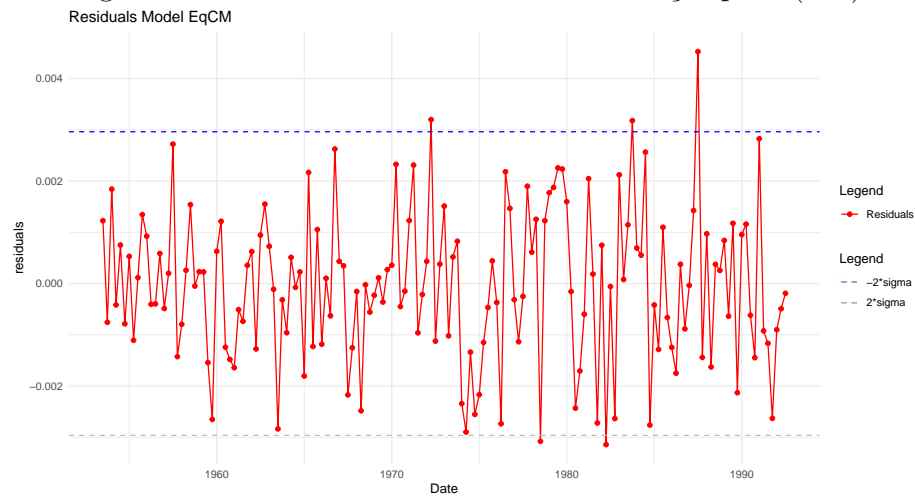


Figura 24: Resíduos e Intervalos de Confianças para (168)



O teste Breusch-Godfrey para correlação serial, não rejeita a hipótese nula de não correlação

Tabela 22: Teste de Autocorrelação dos Resíduos até ordem 5

Breusch - Godfrey Test for AR 1-5			
Statistics	Distribution	Value	p-value
$F - statistics =$	$F(4, 148)$	1.081	0.3683
$Obs * R^2 =$	χ^2_4	4.5531	0.3363

O teste ARCH também não rejeita a nula de variância constante

Tabela 23: Teste de Heteroscedasticidade Condicional dos Resíduos até ordem 4

ARCH Test for order 1-4			
Statistics	Distribution	Value	p-value
$F - statistics =$	$F(4, 148)$	0.5388	0.7074
$Obs * R^2 =$	χ^2_4	2.1960	0.6998

O teste de White para Heteroscedasticidade também não rejeita a nula de variância constante

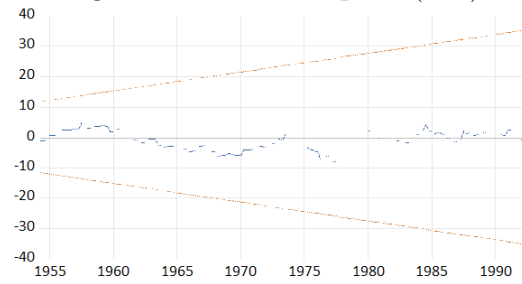
Tabela 24: Teste de Heteroscedasticidade Não Condicional dos Resíduos até ordem 4

White Test for order 1-4			
Statistics	Distribution	Value	p-value
$F - statistics =$	$F(4, 148)$	1.081	0.3683
$Obs * R^2 =$	χ^2_4	4.314	0.3618

O gráfico do CUSUM está dentro das bandas indicando que não houve nenhuma quebra na média.¹⁰

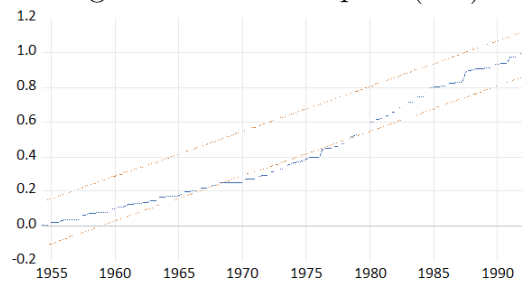
¹⁰Os testes CUSUM e CUSUMSQ assim como as estimativas recursivas dos coeficientes foram obtidas no OxMetrics.

Figura 25: CUSUM para (168)



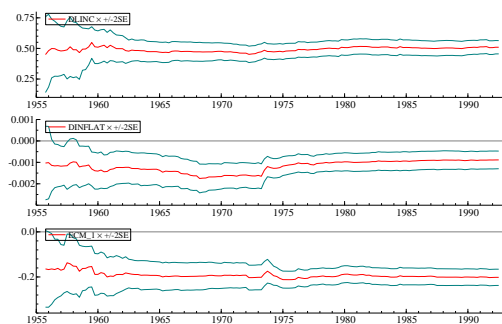
O gráfico do CUSUMSQ apresenta uma quebra no início dos anos 70, indicando que a especificação pode ser melhorada.

Figura 26: CUSUM para (168)



Por outro lado os gráficos dos estimadores recursivos dos parâmetros indicam que estes parâmetros são constantes para toda a amostra.

Figura 27: Coeficientes Recursivos USUM para (168)



3.11 Exercícios

1. Faça um programa no EVIEWS para gerar o seguinte modelo:

$$\begin{aligned} y_t &= 1.0 + 10 * x_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned}$$

com $\varepsilon_t \sim NI(0, 1)$, $\rho = 0.99$, $x_t \sim NI(0, 1)$ e o tamanho da amostra igual a 100.

- (i) estime os parâmetros do modelo por M.Q.O. sem levar em consideração o fato de que o modelo tem autocorrelação;
- (ii) re-estime os parâmetros do modelo assumindo que tem correlação nos erros e é do tipo $AR(1)$;
- (iii) comente os resultados dos itens anteriores;
- (iv) usando 1000 replicações obtenha a distribuição de frequência da constante, do coeficiente de x_t e do coeficiente de autocorrelação dos resíduos e comente os seus resultados.

2. Usando o arquivo ppp.wf1 estime todos os modelos da classe *ADL* para as seguintes especificações:

$$\ln(x_t) = \alpha + \beta_1 \ln(p_t) + \varepsilon_t$$

$$\ln(x_t) = \alpha + \beta_1 \ln(fr_t) + \beta_2 \ln(it_t) + \varepsilon_t$$

comente as implicações estatísticas e econômicas para os resultados obtidos.

3. Usando o arquivo ppp.wf1 estime os seguintes modelos:

$$\ln(x_t) = \alpha + \beta_1 \ln(p_t) + \varepsilon_t$$

$$\ln(x_t) = \alpha + \beta_1 \ln(fr_t) + \beta_2 \ln(it_t) + \varepsilon_t$$

- (i) qual dos dois modelos você escolheria?
 - (ii) teste autocorrelação de primeira ordem para ambos os modelos usando a estatística de Durbin & Watson e comente.
 - (iii) teste autocorrelação usando o correlograma dos resíduos para ambas as especificações e comente
 - (iv) teste autocorrelação de primeira ordem usando o teste de Breusch-Godfrey para ambas as especificações e comente.
 - (v) dado que existe autocorrelação, proponha uma forma de corrigir e implemente o procedimento proposto.
4. As questões abaixo são do tipo falsa, verdadeira ou ambígua. Justifique as suas respostas:
- (i) se existir autocorrelação nos resíduos, os estimadores de M.Q.O. dos parâmetros continuam consistentes;

- (ii) se existir autocorrelação nos resíduos é sinal de falta de especificação dinâmica no modelo;
- (iii) o teste de Durbin & Watson pode ser sempre usado se desejarmos testar autocorrelação de primeira ordem;
- (iv) se o correlograma dos resíduos não decair rapidamente é sinal de que o modelo é inadequado;
- (v) a relação de longo prazo num modelo ADL está sempre bem definida;

4 Métodos de Seleção de Variáveis - Eliminação Reversa, Métodos de Penalização e Auto-metrics

Considere um grande número, N , de variáveis candidatas a serem explicativas, onde o modelo linear geral é postulado ser:

$$Y_t = \sum_{j=1}^N \beta_j X_{j,t} + \varepsilon_t \quad (169)$$

A idéia é começar com um grande número de regressores para garantir que o modelo esteja bem especificado ou, pelo menos, que os resíduos sejam não informativos sobre os regressores.

Dessa forma, garantimos que as inferências sejam válidas e que evitamos o tipo de problema que teríamos se usássemos um modelo estático com erros autorregressivos - modelagem específica para geral.

Suspeitamos então que apenas alguns dos regressores importam. Ou seja, para a maioria dos j , então $\beta_j = 0$ e dizemos que essas variáveis são irrelevantes.

4.1 Eliminação Reversa

A técnica de *backward elimination*, ou eliminação reversa em português, é um método utilizado em análises estatísticas para a seleção de variáveis em um modelo de regressão.

O processo busca simplificar o modelo removendo, passo a passo, as variáveis menos significativas com base em critérios estatísticos, como o valor da estatística t de cada variável. Aqui está um resumo de como funciona a eliminação regressiva:

1. **Iniciar com o Modelo Completo:** Começa-se com um modelo que inclui todas as variáveis candidatas (regressores);
2. **Avaliação da Significância:** O modelo é estimado, e a significância de cada variável é avaliada, geralmente através de suas estatísticas t (ou p-valores);

3. **Remoção da Variável Menos Significativa:** A variável com a estatística t de menor valor absoluto (ou o maior p -valor, indicando a menor significância) é removida do modelo.
4. **Repetição:** O modelo é reestimado sem a variável removida, e os passos 2 e 3 são repetidos até que todas as variáveis remanescentes sejam significativas de acordo com o critério estabelecido, ou até que se atinja o número desejado de variáveis no modelo.

O objetivo da eliminação regressiva é chegar a um modelo mais simples e parcimonioso que ainda forneça uma boa representação das relações entre as variáveis explicativas e a variável dependente. Este método é particularmente útil quando se tem um grande número de variáveis explicativas e deseja-se identificar quais têm o maior impacto ou contribuição para explicar a variável dependente.

É importante notar, no entanto, que a eliminação reversa tem limitações, especialmente quando há alta correlação entre as variáveis explicativas (multicolinearidade), o que pode levar à remoção de variáveis importantes ou à retenção de variáveis menos importantes no modelo final.

Portanto, é sempre recomendável complementar a análise com outros métodos de seleção de variáveis ou conhecimento teórico específico do campo de estudo.

4.2 Seleção Para Frente - *Forward Selection*(FWD)

FWD (Forward Selection) consiste em regressar Y em X_1 , armazenando o residual ($\hat{\varepsilon}_1$).

Em seguida, procurar a covariável no conjunto de informações \mathbf{X} com a maior correlação com o residual, digamos X_2 .

O residual $\hat{\varepsilon}_1$ é projetado em X_2 e um novo residual $\hat{\varepsilon}_2$ é armazenado, e a covariável mais correlacionada com $\hat{\varepsilon}_2$ é identificada.

O procedimento continua até que todas as variáveis no conjunto de informações tenham sido classificadas ou quando um critério específico é satisfeito, por exemplo, o R^2 ajustado ou mínimo de algum critério de informação, por exemplo BIC.

FWD é exatamente o oposto do que está por trás do *hard-thresholding* (em português limiar rígido), enquanto este último pode selecionar um grande número de regressores muito correlacionados entre si, o **FWD** tende a manter menos variáveis o mais ortogonais possível entre si.

Outros métodos de seleção de variáveis são abordados em Kapetanios et al. (2014).

4.3 Regressão de Menor Ângulo - *Least angle regression* - (LARS)

O **LARS** começa como **FWD** identificando a covariável que tem a maior correlação com a série alvo, isto é, a variável dependente, Y .

O maior passo na direção desta covariável X_1 , é dado até que um novo preditor X_2 tenha a maior correlação com o residual atual.

Após este passo, o **LARS** prossegue equiangularmente entre X_1 e X_2 , ao invés de ortogonalmente como no **FWD**.

Após k passos, haverá k variáveis no conjunto ativo.

O algoritmo é interrompido aqui e os coeficientes dos $N - k$ regressores restantes são todos definidos como zero.

4.4 Métodos de Penalização

Os métodos de penalização são populares na literatura estatística atual, veja James et al. (2009)

A **Regressão Ridge** adiciona uma penalidade L^2 ao problema dos mínimos quadrados, de modo que

$$\hat{\beta}_{\text{RIDGE}} = \arg \min_{\beta} \left\{ \sum_{t=1}^T (Y_t - X_t' \beta)^2 + \lambda \left(\sum_{j=1}^N \beta_j^2 \right)^{1/2} \right\}, \quad (170)$$

onde $\lambda > 0$ é uma penalidade escolhida pelo usuário, ou através de algum critério de informação. Quanto maior o λ , maior a penalização para se ter mais regressores.

O **LASSO** (*Least Absolute Shrinkage Selection Operator*) pode ser obtido no algoritmo **LARS** impondo, em cada etapa do algoritmo, uma restrição sobre o sinal da correlação entre o novo candidato a regressor e a projeção na direção equiangular da etapa anterior.

Para obter alguma intuição sobre como o procedimento funciona, vamos começar novamente da etapa 1, quando a variável mais correlacionada com o alvo entra no conjunto ativo de regressores.

Suponha que essa correlação seja positiva.

Ao selecionar a segunda variável para o conjunto ativo, o **LARS** é agnóstico quanto ao sinal da correlação entre o alvo e a nova variável.

Se for imposta a condição de que o sinal dessa correlação não deve mudar, obtém-se a regressão **LASSO**.

Isso significa que, no contexto do **LASSO**, ao expandir o conjunto ativo de regressores, mantém-se uma consistência no sinal da correlação entre os regressores já selecionados e a variável alvo.

Isso é fundamental para a característica de seleção de variáveis do **LASSO**, que não apenas penaliza os coeficientes dos regressores para encolhê-los em direção a zero, mas também pode efetivamente reduzir o coeficiente de alguns regressores para zero, removendo-os do modelo.

Essa propriedade faz do **LASSO** uma ferramenta valiosa para a seleção de variáveis, especialmente em contextos com um grande número de variáveis explicativas, muitas das quais podem ser irrelevantes para a previsão da variável dependente.

A regressão **LASSO** adiciona uma penalidade L^1 , em vez da penalidade L^2 que foi usada na Regressão Ridge, de modo temos que:

$$\hat{\beta}_{\text{LASSO}} = \arg \min_{\beta} \left\{ \sum_{t=1}^T (Y_t - X_t' \beta)^2 + \lambda \sum_{j=1}^N |\beta_j| \right\}. \quad (171)$$

LASSO pode ser obtido no algoritmo **LARS** impondo que em cada passo a restrição do sinal da correlação entre o novo candidato a regressor e a

A idéia desses métodos é que, para uma grande penalidade, os coeficientes das variáveis irrelevantes são reduzidos a zero.

Esses métodos funcionam melhor com regressores não correlacionados.

Para regressores altamente correlacionados, eles podem fornecer resultados enganosos.

Esses métodos requerem que os regressores sejam medidos na mesma escala.

A alteração da escala muda a importância relativa dos regressores, implicando que o estimador pode mudar.

O estimador *Elastic Net* (**NET**) é um refinamento do **LASSO** e é a solução do seguinte problema de minimização:

$$\hat{\beta}_{\text{NET}} = \arg \min_{\beta} \left\{ \sum_{t=1}^T (Y_t - X_t' \beta)^2 + \lambda_1 \sum_{j=1}^N |\beta_j| + \lambda_2 \sum_{j=1}^N |\beta_j| \right\}. \quad (172)$$

Nesta formulação:

- λ_1 e λ_2 são parâmetros de penalização escolhidos pelo usuário, onde λ_1 controla a força da penalidade L^1 (**LASSO**) e λ_2 controla a força da penalização L^2 (**RIDGE**).

O Elastic Net combina as penalidades L^1 e L^2 , aproveitando as vantagens de ambos os métodos. A penalidade L^1 permite que o Elastic Net realize seleção de variáveis, reduzindo alguns coeficientes a zero, como o **LASSO**. A penalidade L^2 ajuda a lidar com o problema de multicolinearidade entre as variáveis explicativas e a estabilizar as estimativas dos coeficientes, uma característica do **RIDGE**.

Este método é particularmente útil em situações onde há muitas variáveis explicativas que são altamente correlacionadas. O Elastic Net pode selecionar grupos de variáveis correlacionadas, algo que o **LASSO** pode falhar em fazer devido à sua natureza de selecionar uma variável de um grupo de variáveis altamente correlacionadas.

Além disso, ao ajustar os parâmetros λ_1 e λ_2 , os usuários podem controlar o equilíbrio entre a seleção de variáveis e a regularização, tornando o Elastic Net uma ferramenta flexível e poderosa para a modelagem preditiva.

4.5 Avaliando todos os sub-modelos

O modelo (169) é chamado de Modelo Geral Não-Restrito (em inglês *General Unrestricted Model* (**GUM**)).

O GUM possui N variáveis. Cada variável possui um coeficiente que é ou zero ou não-zero.

Isso dá 2^N possibilidades para selecionar variáveis.

A ideia é usar todas as 2^N possibilidades, ou, se 2^N for um número grande, pelo menos um grande número desses submodelos.

Um submodelo é considerado 'terminal' ou às vezes 'pivotal' se os seguintes dois critérios forem satisfeitos:

- 1) o teste LR contra o modelo geral não restrito é aceito;

- 2) para cada uma das variáveis retidas, o teste t de um coeficiente zero é rejeitado.

Haverá pelo menos um modelo 'terminal', possivelmente mais.

Encontrar vários modelos 'terminais' é um sinal para o usuário de que esses submodelos são todas reduções estatisticamente equivalentes do ponto de partida, mas os dados não fornecem nenhum motivo para escolher entre eles.

4.6 Busca Multicaminho - *Multipath search* e *Autometrics*

Hoover and Perez (1999) introduziram uma busca multicaminho. A idéia deles era executar o seguinte algoritmo:

- (1) Execute a regressão com $m = N$ regressores;
- (2) Encontre os 10 regressores menos significativos;
- (3) Para cada um desses 10 regressores, elimine esse regressor e execute a eliminação reversa.

Este algoritmo pode lidar muito melhor com multicolinearidade do que uma simples eliminação reversa.

É um passo viável em direção aos algoritmos de todos os submodelos.

Isso foi implementado no PcGets, veja Hendry and Krolzig (2005), e no Autometrics, veja Doornik (2009) e Hendry and Doornik (2014).

Quando o número de regressores N é maior que o número de observações T , a regressão simplesmente fornecerá um ajuste perfeito, o que não é tão interessante.

Alguns dos métodos descritos anteriormente podem ser usados. Vamos discutir a abordagem Autometrics.

O Autometrics é uma metodologia avançada em econometria que automatiza o processo de seleção de modelos, procurando identificar o modelo mais adequado dentre um conjunto potencialmente vasto de modelos candidatos. Essa abordagem é particularmente útil em contextos onde o número de regressores (N) é muito grande, possivelmente maior que o número de observações (T), o que torna a seleção manual de modelos impraticável e sujeita a erros.

O processo do Autometrics envolve várias etapas chave:

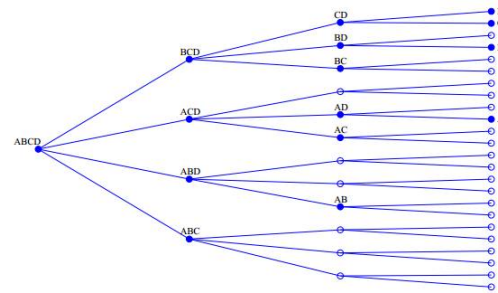
- (i) **Divisão dos Regressores em Blocos:** Para tornar o processo de seleção de modelo mais gerenciável, os regressores são divididos em blocos. Isso permite uma análise mais detalhada e focada de grupos de variáveis.
- (ii) **Realização de Eliminações Reversas em Grande Escala:** Utilizando um procedimento de eliminação reversa, o Autometrics executa uma série de regressões, removendo sistematicamente os regressores que não atendem a critérios de significância pré-estabelecidos. Esse processo é iterativo e pode envolver uma grande quantidade de modelos candidatos.
- (iii) **Testes de Significância e Especificação Conjuntos:** Em cada etapa, o Autometrics realiza testes para avaliar a significância conjunta dos regressores no modelo, bem como testes de especificação para verificar se o modelo está corretamente especificado, isto é, se ele adequadamente representa as relações subjacentes entre as variáveis.
- (iv) **Avaliação de Critérios de Informação:** O Autometrics utiliza critérios de informação, como o Critério de Informação de Akaike (AIC) ou o Critério de Informação Bayesiano (BIC), para comparar os modelos e selecionar aquele que oferece o melhor equilíbrio entre ajuste do modelo e parcimônia (complexidade do modelo).
- (v) **Saturação de Indicadores de Impulso e Degrau:** O Autometrics pode incluir técnicas de Saturação de Indicadores de Impulso (*Impluse Indicator Saturation* - IIS, veja Santos et al. (2008)) e Saturação de Indicadores de Degrau (*Step Indicaator Saturation* - SIS, veja Castle et al. (2015)) para identificar e modelar outliers e mudanças estruturais, respectivamente. Essas técnicas permitem que o modelo se ajuste flexivelmente a mudanças súbitas nos dados, melhorando sua capacidade de capturar a dinâmica subjacente.

Ao final do processo, o Autometrics identifica o modelo que, segundo os testes de significância, especificação e critérios de informação, melhor representa os dados.

Essa abordagem fornece uma maneira sistemática e objetiva de tratar a seleção de modelos em situações de alta dimensionalidade e complexidade, reduzindo o risco de sobre ajuste e melhorando a qualidade das inferências estatísticas.

O item(i) acima pode ser representado pela seguinte árvore de escolha entre modelos:

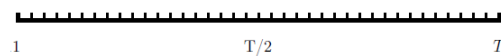
Figura 28: Autometrics Arvore



Nesta árvore estamos exemplificando com $N = 4$ regressores. O ramo mais à esquerda contém todos os quatro regressores. O próximo ramo, quando se anda para à direita agora contém modelos com três regressores. No próximo ramo estão os modelos com dois regressores e no último ramo modelos com um regressor.

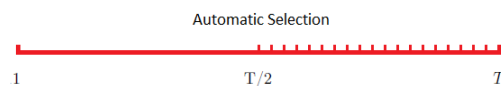
No item (v) acima o procedimento adotado pelo Autometrics consiste em usar todas as observações, isto é

Figura 29: Todas as observações



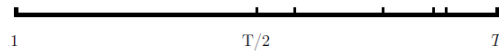
e colocar dummies em todas as últimas $T/2$ observações, isto é:

Figura 30: Dummies nas últimas observações



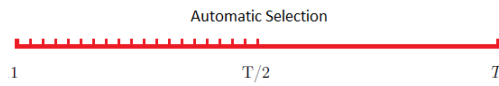
Os itens (ii)-(iii) são aplicados nestas últimas observações e, por exemplo, ficamos com as dummies nas seguintes observações:

Figura 31: Dummies selecionadas nas últimas observações



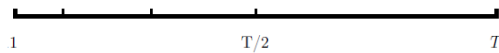
Agora coloca-se dummies em toas as primeiras $T/2$ observações, isto é:

Figura 32: Dummies nas primeiras observações



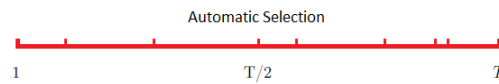
Os itens (ii)-(iii) são aplicados nestas primeiras observações e, por exemplo, ficamos com as dummies nas seguintes observações:

Figura 33: Dummies selecionadas nas primeiras observações



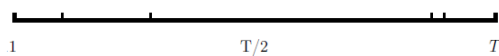
Juntando as dummies selecionadas na última parte, figura (32) com as selecionadas na primeira parte, figura (33), temos:

Figura 34: Junta as Dummies das duas partes



Os itens (ii)-(iii) são aplicados nestas dummies e, por exemplo, ficamos com as dummies nas seguintes observações:

Figura 35: Dummies selecionadas



O algoritmo Autometrics tem uma implementação no R que é o Gets , veja Pretis et al. (2018) ou Sucarrat (2021). O GUM será um ADL(5,5,5) escrito como EqCM, isto é:

$$\begin{aligned}\Delta \log(Cons_t) &= \sum_{i=1}^5 \alpha_i \Delta \log(Cons_{t-i}) + \sum_{i=0}^5 \beta_i \Delta \log(Inc_{t-i}) \\ &+ \sum_{i=0}^5 \gamma_i \Delta Inflat_{t-i} - \delta CI_{t-1} + \epsilon_t\end{aligned}\tag{173}$$

onde CI_{t-1} é o termo de correção do equilíbrio.

Temos os seguintes resultados:¹¹

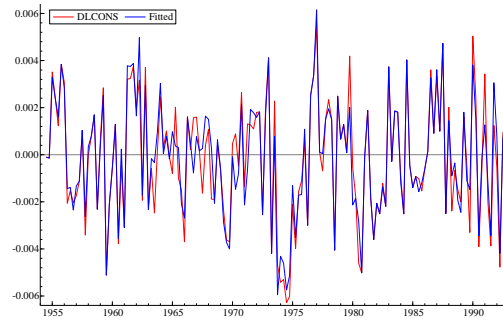
¹¹Estes resultados foram obtidos com o Autometrics dp OxMetrics, porque os resultados do Gets do R não conseguiram indentificar e listas as IIS só dois SIS. Outro problema é que o coeficiente de CI que deveria ser negativo é positivo.

Tabela 25: Modelo Final para (173)

EQ15) Modelo Final para (173) com IIS e SIS				
Desvios Padrões com correção por HAC				
The estimation sample is: 1954(3) - 1992(3)				
	Coefficient	Std. Error	t-Statistics	p-value
$DLConst-4$	0.1140	0.0377	3.02	0.0032
$DLInc_t$	0.5022	0.0242	20.7	0.0000
$DInflat_t$	-0.0011	0.0002	-4.48	0.0000
$CI_{new,t-1}$	-0.1959	0.0220	-8.93	0.0000
$IIS_{1954(3)}$	0.0005	0.0001	4.31	0.0000
$IIS_{1954(4)}$	-0.0006	0.0001	-5.86	0.0000
$IIS_{1955(2)}$	-0.0011	0.0002	-6.80	0.0000
$IIS_{1955(4)}$	0.0009	0.0002	4.64	0.0000
$IIS_{1957(2)}$	0.0007	0.0002	4.47	0.0000
$IIS_{1957(3)}$	0.0027	0.0002	18.7	0.0000
$IIS_{1958(2)}$	0.0002	6.275e - 05	3.91	0.0002
$IIS_{1958(3)}$	0.0009	0.0002	5.29	0.0000
$IIS_{1959(3)}$	-0.0014	0.0002	-5.86	0.0000
$IIS_{1959(4)}$	-0.0016	0.0002	-9.58	0.0000
$IIS_{1960(1)}$	0.0005	0.0001	4.83	0.0000
$IIS_{1960(2)}$	0.0009	0.0001	7.55	0.0000
$IIS_{1961(1)}$	-0.0014	0.0001	-10.2	0.0000
$IIS_{1972(2)}$	0.0022	0.0002	14.0	0.0000
$IIS_{1972(3)}$	-0.0011	7.259e - 05	-15.5	0.0000
$IIS_{1973(1)}$	0.0012	0.0002	6.84	0.0000
$IIS_{1973(2)}$	-0.0011	0.0002	-5.30	0.0000
$IIS_{1976(2)}$	-0.0025	0.0002	-9.96	0.0000
$IIS_{1976(4)}$	0.0009	0.0002	4.54	0.0000
$IIS_{1977(2)}$	-0.0012	0.0004	-2.64	0.0095
$IIS_{1977(4)}$	0.00144	0.0002	5.88	0.0000
$IIS_{1978(2)}$	0.0014	0.0001	120	0.0000
$IIS_{1978(3)}$	-0.0024	0.0002	-15.9	0.0000
$IIS_{1978(4)}$	0.0008	0.0002	4.57	0.0000
$IIS_{1979(1)}$	0.0012	0.0001	10.7	0.0000
$IIS_{1979(2)}$	0.0016	8.854e - 05	18.6	0.0000
$IIS_{1979(3)}$	0.0026	0.0002	12.6	0.0000
$IIS_{1980(4)}$	-0.0017	0.0002	-9.88	0.0000
$IIS_{1981(2)}$	0.0022	0.0001	19.5	0.0000
$IIS_{1981(3)}$	0.0010	0.0002	3.89	0.0002
$IIS_{1981(4)}$	-0.0014	0.0002	-5.48	0.0000
$IIS_{1982(1)}$	0.0008	0.0001	6.58	0.0000
$IIS_{1982(2)}$	-0.0026	0.0002	-12.7	0.0000
$IIS_{1982(4)}$	-0.0019	0.0001	-12.9	0.0000
$IIS_{1983(1)}$	0.0019	8.896e - 05	20.9	0.0000
$IIS_{1983(3)}$	0.0009	8.631e - 05	9.98	0.0000
$IIS_{1983(4)}$	0.0027	0.0001	18.4	0.0000
$IIS_{1984(2)}$	0.0007	0.0001	5.11	0.0000
$IIS_{1984(3)}$	0.0025	0.0003	9.76	0.0000
$IIS_{1984(4)}$	-0.0023	0.0002	-10.1	0.0000
$IIS_{1985(1)}$	-0.0004	9.374e - 05	-4.05	0.0001
$IIS_{1985(2)}$	-0.0007	0.0002	-4.58	0.0000
$IIS_{1986(1)}$	-0.0008	0.0001	-7.86	0.0000
$IIS_{1986(2)}$	-0.0015	6.246e - 05	-23.6	0.0000
$IIS_{1986(4)}$	-0.0009	0.0001	-7.86	0.0000
$IIS_{1987(2)}$	0.0007	0.0002	3.48	0.0007
$IIS_{1987(3)}$	0.0033	0.0002	19.3	0.0000
$IIS_{1987(4)}$	-0.0014	8.415e - 05	-17.1	0.0000
no of observations: 153				
no of parameters: 53				
$\sigma_\varepsilon = 0.0009528$				
$R^2 = 0.9112$		$\overline{R}^2 = 0.8671$		
$AIC = -10.8068$				
$BIC = -9.75704$				
log-likelihood = 879.720				

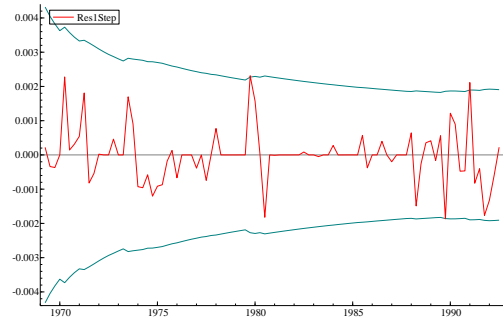
O gráfico do observado e previsto para o modelo acima

Figura 36: Observado e Previsto para o Modelo Final



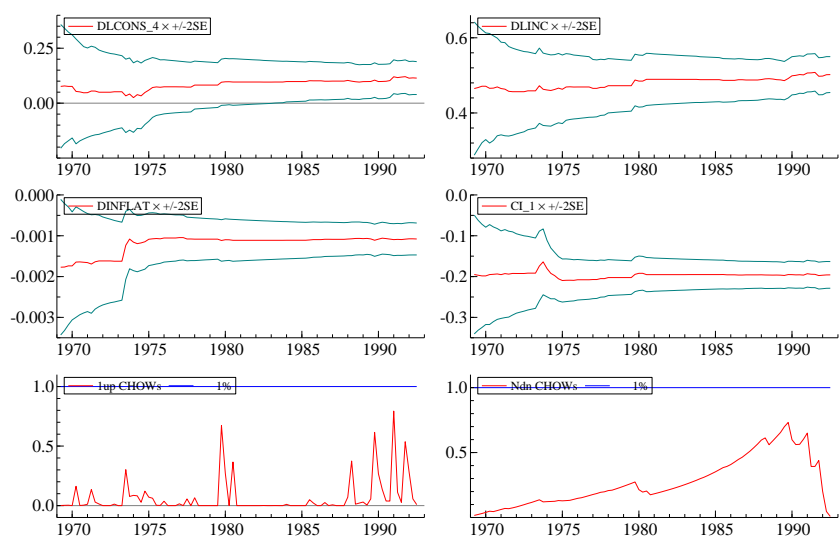
O gráfico do resíduos recursivos para o modelo acima é apresentado abaixo:

Figura 37: Resíduos Recursivos para o Modelo Final



O gráfico abaixo apresenta as estimativas recursivas dos coeficientes de $DLCons_{t-4}$, $DLInc_t$, $DInflat_t$ e CI_{t-1} , assim como os testes de Chow para constância da média condicional.

Figura 38: Estimativa Recursiva dos Coeficientes e teste de Chow para o Modelo Final



5 Apêndice 1

Tabela 26: Tipologia de Modelos Dinâmicos

Tipo de Modelo	Parâmetros $\theta = (m, \alpha_1, \beta_0, \beta_1)$	Modelo	Restrição	Tipo
Regressão Estática	$(m, 0, \beta_0, 0)$	$y_t = m + \beta_0 x_t + \varepsilon_t$	$\alpha_1 = \beta_1 = 0$	sem dinâmica
AR(1)	$(m, \alpha_1, 0, 0)$	$y_t = m + \beta_1 y_{t-1} + \varepsilon_t$	$\beta_0 = \beta_1 = 0$	sem covariável
Indicador Antecedente	$(m, 0, 0, \beta_1)$	$y_t = m + \beta_1 x_{t-1} + \varepsilon_t$	$\alpha_1 = \beta_0 = 0$	sem contemporaneidade
Taxa de Crescimento	$(m, 1, \beta_0, -\beta_1)$	$\Delta y_t = \Delta \beta_0 x_t + \varepsilon_t$	$\alpha_1 = 1, \beta_0 = -\beta_1$	sem nível
Defasagens Distribuídas	$(m, 0, \beta_0, \beta_1)$	$y_t = m + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$	$\alpha_1 = 0$	defasagens finitas
Ajustamento Parcial	$(m, \alpha_1, \beta_0, 0)$	$y_t = m + \beta_1 y_{t-1} + \beta_0 x_t + \varepsilon_t$	$\beta_1 = 0$	sem defasagens em x
Regressão Estática com Erros Correlacionados	$(m, \alpha_1, \beta_0, -\alpha_1 \beta_0)$	$y_t = m + \frac{m}{1-\alpha_1} + \beta_0 x_t + u_t$ $u_t = \alpha_1 u_{t-1} + \varepsilon_t$	$\beta_1 = -\alpha_1 \beta_0$	restrição de fatores comuns
Forma Reduzida	$(m, \alpha_1, 0, \beta_1)$	$y_t = m + \beta_1 y_{t-1} + \beta_1 x_{t-1} + \varepsilon_t$	$\beta_0 = 0$	só variáveis defasadas
Modelo de Correção do Equilíbrio	$(m, \alpha_1, \beta_0, \gamma)$	$\Delta y_t = \Delta \beta_0 x_t - (1 - \alpha_1)(y_{t-1} - \frac{m}{1-\alpha_1} - \gamma x_{t-1}) + \varepsilon_t$	$\gamma = \frac{\beta_0 + \beta_1}{1-\alpha_1}$	resposta de longo prazo
Modelo de Correção do Equilíbrio Homogêneo	$(m, \alpha_1, \beta_0, \beta_1)$	$\Delta y_t = \Delta \beta_0 x_t - (1 - \alpha_1)(y_{t-1} - \frac{m}{1-\alpha_1} - x_{t-1}) + \varepsilon_t$	$1 = \beta_0 + \beta_1 + \alpha_1$	sem restrição

Referências

- J. Castle, J. Doornik, D. Hendry, and F. Pretis. Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, 3(2): 240–264, 2015. ISSN 2225-1146. doi: 10.3390/econometrics3020240.
- J. A. Doornik. Autometrics. In J. Castle and N. Shephard, editors, *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, chapter 4, pages 88–121. Oxford University Press, 2009.
- J. Durbin. Errors in variables. *Revue de Institut International de Statistique / Review of the International Statistical Institute*, 22(1/3):23–32, 1954. doi: 10.2307/1401917.
- J. Durbin. Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica*, 38(3):410, 1970. ISSN 0012-9682. doi: 10.2307/1909547.
- J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4):409, 1950. ISSN 0006-3444. doi: 10.2307/2332391.
- J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression. ii. *Biometrika*, 38(1/2):159, 1951. ISSN 0006-3444. doi: 10.2307/2332325.
- R. F. Engle and C. W. J. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276, mar 1987. doi: 10.2307/1913236.
- N. R. Ericsson. Cointegration, exogeneity, and policy analysis: An overview. *Journal of Policy Modeling*, 14(3):251–280, jun 1992. doi: 10.1016/0161-8938(92)90001-s.
- N. R. Ericsson. The et interview: Professor david f. hendry: Interviewed by neil r. ericsson. *Econometric Theory*, 20(04), 2004-08. ISSN 1469-4360. doi: 10.1017/s0266466604204078.
- N. R. Ericsson and J. G. MacKinnon. Distributions of error correction tests for cointegration. *The Econometrics Journal*, 5(2):285–318, 2002.

- C. L. Gilbert. Use and the british approach to time series econometrics. *Oxford Economic Papers*, 41(1):108–128, 1989. ISSN 0030-7653. doi: 10.1093/oxfordjournals.oep.a041887.
- T. Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1):1, 1943. ISSN 0012-9682. doi: 10.2307/1905714.
- J. A. Hausman. Specification tests in econometrics. *Econometrica*, 46(6): 1251–1271, nov 1978. doi: 10.2307/1913827.
- D. Hendry. *Dynamic Econometrics*. Oxford University Press, 1995. ISBN 0198283164. URL https://www.ebook.de/de/product/3257383/david_hendry_dynamic_econometrics.html.
- D. F. Hendry and J. A. Doornik. *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics*. MIT PR, 2014. ISBN 0262028352. URL https://www.ebook.de/de/product/22394850/david_f_hendry_jurgen_a_doornik_empirical_model_discovery_and_theory_evaluation_automatic_selection_methods_in_econometrics.html.
- D. F. Hendry and H.-M. Krolzig. The properties of AutomaticGetsModelling. *The Economic Journal*, 115(502):C32–C61, mar 2005. doi: 10.1111/j.0013-0133.2005.00979.x.
- D. F. Hendry and B. Nielsen. *Econometric Modeling*. Princeton University Press, 2007. URL https://www.ebook.de/de/product/19215268/david_f_hendry_econometric_modeling.html.
- D. F. Hendry, A. R. Pagan, and J. Sargan. Chapter 18 dynamic specification. volume 2 of *Handbook of Econometrics*, pages 1023–1100. Elsevier, 1984. doi: [https://doi.org/10.1016/S1573-4412\(84\)02010-9](https://doi.org/10.1016/S1573-4412(84)02010-9). URL <https://www.sciencedirect.com/science/article/pii/S1573441284020109>.
- K. D. Hoover and S. J. Perez. Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *The Econometrics Journal*, 2(2):167–191, dec 1999. doi: 10.1111/1368-423x.00025.
- G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. SPRINGER NATURE, 2nd edition edition, 2009. ISBN 1461471370.

URL https://www.ebook.de/de/product/20292548/gareth_james_daniela_witten_trevor_hastie_an_introduction_to_statistical_learning_with_applications_in_r.html.

- J. Johnston and J. DiNardo. *Econometric Methods*. McGraw-Hill/Irwin, 1996. ISBN 978-0079131218. URL <https://www.amazon.com/Econometric-Methods-Jack-Johnston/dp/0079131212?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0079131212>.
- K. Juselius. Models and relations in economics and econometrics. *Journal of Economic Methodology*, 6(2):259–290, 1999. ISSN 1469-9427. doi: 10.1080/13501789900000017.
- G. Kapetanios, M. G. Marcellino, and F. Papailias. Variable selection for large unbalanced datasets using non-standard optimisation of information criteria and variable reduction methods. *SSRN Electronic Journal*, 2014. doi: 10.2139/ssrn.2444418.
- J. G. MacKinnon. Critical values for cointegration tests. In R. F. Engle and C. W. J. Granger, editors, *Long- Run Economic Relationships*, pages 267–276. Oxford University Press, 1991.
- W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, may 1987. doi: 10.2307/1913610.
- F. Pretis, J. J. Reade, and G. Sucarrat. Automated general-to-specific (gets) regression modeling and indicator saturation for outliers and structural breaks. *Journal of Statistical Software*, 86(3):1–44, 2018. doi: 10.18637/jss.v086.i03. URL <https://www.jstatsoft.org/index.php/jss/article/view/v086i03>.
- C. Santos, D. F. Hendry, and S. Johansen. Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, 23(2):317–335, apr 2008. doi: 10.1007/s00180-007-0054-z.
- G. Sucarrat. garchx: Flexible and Robust GARCH-X Modeling. *The R Journal*, 13(1):276–291, 2021. doi: 10.32614/RJ-2021-057. URL <https://doi.org/10.32614/RJ-2021-057>.

- K. F. Wallis. Testing for fourth order autocorrelation in quarterly regression equations. *Econometrica*, 40(4):617–636, jul 1972. doi: 10.2307/1912957.
- D.-M. Wu. Alternative tests of independence between stochastic regressors and disturbances. *Econometrica*, 41(4):733–750, jul 1973. doi: 10.2307/1914093.