

**FUNDAÇÃO GETULIO VARGAS
ESCOLA DE PÓS-GRADUAÇÃO EM
ECONOMIA**

Diogo Vinícius Menezes Saraiva

**Previsão de preços de ativos usando restrições do
modelo de Valor Presente**

Rio de Janeiro
2012

Diogo Vinícius Menezes Saraiva

**Previsão de preços de ativos usando restrições do
modelo de Valor Presente**

Dissertação submetida a Escola de Pós-Graduação em Economia como requisito parcial para a obtenção do grau de Mestre em Economia.

Área de Concentração: Econometria

Orientador: João Victor Issler

Rio de Janeiro
2012

Saraiva, Diogo Vinícius Menezes

Previsão de preços de ativos usando restrições do modelo de valor presente / Diogo Vinícius Menezes Saraiva. – 2012.
33 f.

Dissertação (mestrado) - Fundação Getulio Vargas, Escola de Pós-Graduação em Economia.

Orientador: João Victor Issler.

Inclui bibliografia.

1. Modelos econométricos. 2. Previsão econômica – Modelos econométricos. 3. Ciclos econômicos – Modelos econométricos. 4. Cointegração – Modelos econométricos. I. Issler, João Victor. II. Fundação Getulio Vargas. Escola de Pós-Graduação em Economia. III. Título.

CDD – 330.015195



**FUNDAÇÃO
GETULIO VARGAS**

DIOGO VINICIUS MENEZES SARAIVA

**PREVISÃO DE PREÇOS DE ATIVOS UTILIZANDO RESTRIÇÕES DO MODELO DE
VALOR PRESENTE .**

Dissertação apresentada ao Curso de Mestrado em Economia da Escola de Pós-Graduação em Economia para obtenção do grau de Mestre em Economia.

Data da defesa: 21/12/2012

Aprovada em:

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

João Victor Issler
Orientador (a)

Osmani Teixeira de Carvalho Guillén
Co-Orientador (a)

Marco Antonio Cesar Bonomo

Resumo

Motivados pelo debate envolvendo modelos estruturais e na forma reduzida, propomos nesse artigo uma abordagem empírica com o objetivo de ver se a imposição de restrições estruturais melhoram o poder de previsibilidade vis-a-vis modelos irrestritos ou parcialmente restritos. Para respondermos nossa pergunta, realizamos previsões utilizando dados agregados de preços e dividendos de ações dos EUA. Nesse intuito, exploramos as restrições de cointegração, de ciclo comum em sua forma fraca e sobre os parâmetros do *VECM* impostas pelo modelo de Valor Presente. Utilizamos o teste de igualdade condicional de habilidade de previsão de Giacomini e White (2006) para comparar as previsões feitas por esse modelo com outros menos restritos. No geral, encontramos que os modelos com restrições parciais apresentaram os melhores resultados, enquanto o modelo totalmente restrito de VP não obteve o mesmo sucesso.

PALAVRAS CHAVE: *Valor Presente; Ciclos Comuns; Previsão; Cointegração.*

Abstract

Given the debate in econometrics about the use of structural models versus reduced form models, we ask here whether the straight use of the former helps in forecasting. To answer our question, we performed out-of-sample forecasts using monthly observations on aggregate prices and dividends of U.S. stocks. We show that PV relationships entail cointegration, weak-form common feature relationship and coefficient restrictions on *VECM* form. We compare forecasting models which deal with these restrictions in different layers. For this comparison, we use the equality test of conditional predictive ability of Giacomini e White (2006). Overall, we found that models with partial restrictions had the best results, while the model fully restricted by Presente Value Model did not work well.

KEYWORDS: *Present Value; Common Cycle; Forecasting; Cointegration.*

Sumário

1	Introdução	6
2	O modelo de valor presente	8
3	Restrições de Ciclos Comuns	11
4	Seleção do modelo e estimação	12
5	Dados dos preços agregados e dividendos	15
6	Teste fora da amostra	17
6.1	Resultados da Estimação	18
7	Conclusão	25
A	Apêndice	28

1 Introdução

Relações de cointegração entre o nível de duas variáveis Y_t , y_t , têm sido um tema recorrente na literatura de modelos de Valor Presente (VP) desde o paper de Campbell e Shiller (1987). A necessidade de cointegração para a validação do modelo de Valor Presente, com a consequente imposição da condição de transversalidade sobre os dados, tem sido largamente pesquisada na academia.

Entretanto, Hecq e Issler (2012) trabalharam em uma outra condição também necessária e pouco discutida para a existência do modelo de VP. Essa se refere a existência de uma combinação linear das séries que é imprevisível ao conjunto de informação passado. Se essa condição falhar, a equação de valor presente não será válida, uma vez que irá conter um termo adicional capturando o valor esperado condicional de erros futuros.

Baseado nessas restrições impostas pelo modelo de Valor Presente, e motivados pelo debate existente em econometria sobre a utilização de modelos estruturais *versus* na forma reduzida, propomos nesse artigo uma abordagem empírica para ver se a imposição de restrições estruturais melhoram o poder de previsibilidade vis-a-vis modelos irrestritos ou parcialmente restritos.

Athanasopoulos et al. (2011) desenvolveram uma estratégia ótima para selecionar conjuntamente séries com restrições de longo prazo e de curto prazo em sua forma fraca, estimando conjuntamente seus parâmetros. Eles mostraram os benefícios obtidos dessa nova estratégia para propósitos de previsão, quando comparado aos critérios de seleção existentes. Esse resultado se mostra bastante interessante para a nossa aplicação, tendo em vista que os modelos de Valor Presente implicam nesses tipos de restrição. Logo, a estratégia de Athanasopoulos et al. (2011) serve como ponto de partida para construirmos modelos de previsão com restrições de valor presente.

No entanto, nos deparamos com o problema de que o modelo de Valor Presente impõem restrições não-lineares sobre os parâmetros de longo prazo, de modo que uma solução *second best* para a estimação do modelo restrito é utilizada.

Para respondermos nossa pergunta inicial, sobre a utilidade de equações estruturais do modelo de Valor Presente para a previsibilidade, comparamos esse modelo totalmente restrito com outros três modelos: $Var(p+1)$, $VECM(HQ-PIC)$ e $VECM(AIC-J)$. O primeiro é um modelo totalmente irrestrito, enquanto os demais envolvem restrições parciais de cointegração e/ou de ciclos comuns em sua forma fraca. Para seleção e estimação desses modelos utilizamos tanto o procedimento recente de Athanasopoulos et al. (2011) quanto os mais tradicionais bayesiano (BIC), Akaike (AIC) e Johansen (1988, 1991).

Para analisar a performance dos nossos modelos, nós examinamos os erros de previsão para dois períodos específicos fora da amostra de estimação. O primeiro, até 1994, e o segundo até 2012. Nós optamos por essa estratégia pelas crises mundiais de 2000 e 2010 que poderiam alterar os resultados. Nós concentramos nossas observações para a usual raiz dos erros quadráticos médios de previsão (RMSFE), calculadas em relação ao modelo *benchmarking* $Var(p)$. No entanto, a medida RMSFE não fornece evidências se a diferença é estatisticamente significativa.

Desse modo, um teste mais formal para decidir qual modelo é mais preciso se faz necessário. No nosso caso, temos uma complicação no fato que os modelos são *nested*, impossibilitando alguns testes como o de Diebold e Mariano (1995). Apoiados no fato que utilizamos uma estimação por janela rolante (*rolling window*), escolhemos o teste de igualdade condicional de habilidade em previsões de Giacomini e White (2006). Esse teste possui uma série de vantagens, como permitir a comparação tanto entre modelos *nested*

como também não *nested*, independente do procedimento de estimação utilizado para a realização das previsões.

Utilizando a metodologia descrita acima, os modelos VECM(HQ-PIC) e VECM(AIC-J) apresentaram os melhores resultados no computo geral. Podemos dizer que os resultados do último foram mais impactantes, por conseguir resultados estatisticamente mais significativos. O modelo com todas as restrições de VP, mostrou-se o menos preciso no aspecto geral, apesar de em determinados momentos superar o *benchmarking* $Var(p+1)$. Uma característica comum a todos, foi o melhor desempenho na amostra restrita até 1994, o que indica que sob períodos de estabilidade as previsões dos modelos são mais confiáveis.

O restante do artigo é organizado da seguinte maneira: Na seção II apresentamos o modelo de Valor Presente e suas derivações, na seção III falamos das restrições impostas pelo modelo, na seção IV do procedimento de seleção dos modelos e estimação, na seção V descrevemos os dados utilizados, na seção VI os resultados da estimação fora da amostra. A seção VII conclui.

2 O modelo de valor presente

Nesta seção, apresentaremos o modelo de valor presente (VP) com suas notações e representações.

Considere a seguinte equação de valor presente:

$$Y_t = \theta(1 - \delta) \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}, \quad (1)$$

onde Y_t é uma função linear do valor presente descontado do fluxo esperado futuro y_t , onde $\mathbb{E}_t(\cdot)$ é o operador do valor esperado condicional, com conjunto de informações formado por aquelas disponíveis até t . Alguns dos principais exemplos de Y_t e y_t incluem: taxas de juros de longo e curto prazo, preços de ações e dividendos, consumo pessoal e renda disponível. Na maioria dos casos Y_t e y_t são variáveis $I(1)$ e no nosso em especial eles representarão preços e dividendos, respectivamente. Seguindo o modelo exposto em Campbell e Shiller (1987), assumimos ao longo do artigo que o preço da ação Y_t é medido no início do período, enquanto os dividendos são recebidos ao final do período. Ou seja, a compra de uma ação ao preço Y_t hoje dá direito ao comprador de receber o dividendo y_t desse mesmo período. Além disso, temos retornos esperados constantes com fator de desconto dado por $\delta = \frac{1}{1+R}$. Sendo assim, temos a seguinte relação com o retorno esperado R da ação τ passos a frente:

$$\mathbb{E}_t[R_{t+\tau}] = R \quad (2)$$

onde $R_{t+\tau} \equiv \frac{Y_{t+\tau} + y_t}{Y_t} - 1$. Usando essa definição com a igualdade expressa na equação (2), rearranjando os termos, obtemos a equação relacionando o preço corrente da ação com o valor esperado do dividendo e do preço no próximo período da ação:

$$Y_t = \mathbb{E}_t \left[\frac{Y_{t+1} + y_t}{1 + R} \right] \quad (3)$$

Essa equação diferencial em expectativa pode ser resolvida recursivamente por substituições repetitivas dos preços futuros e usando a Lei das Expectativas Iteradas. Resolvendo para K períodos a frente nós temos:

$$Y_t = \mathbb{E}_t \left[\frac{1}{1+R} \sum_{i=0}^K \left(\frac{1}{1+R} \right)^i y_{t+i} \right] + \mathbb{E}_t \left[\left(\frac{1}{1+R} \right)^K Y_{t+K} \right] \quad (4)$$

O segundo termo no lado direito da equação (4) é o valor esperado descontado do preço da ação K períodos a frente do presente. Assumindo que esse termo converge para zero conforme o horizonte K cresce:

$$\lim_{K \rightarrow \infty} \mathbb{E}_t \left[\left(\frac{1}{1+R} \right)^K Y_{t+K} \right] = 0 \quad (5)$$

A hipótese estabelecida pela equação (5) será satisfeita ao menos que seja esperado um aumento no preço da ação para sempre a taxa R ou maior.

Fazendo K tender ao infinito e assumindo (5), obtemos a fórmula que expressa o preço da ação como valor presente esperado dos dividendos futuros, descontado a uma

taxa constante:

$$Y_t = \mathbb{E}_t \left[\frac{1}{1+R} \sum_{i=0}^{\infty} \left(\frac{1}{1+R} \right)^i y_{t+i} \right] \quad (6)$$

O modelo de valor presente para ações exposto na equação (6) acima é um caso especial da equação (1), em que o coeficiente de proporcionalidade θ é igual a $\delta/(1-\delta)$. Diferentes medidas desse coeficiente se refletem apenas em diferentes valores para o vetor de cointegração. Sendo assim, de agora em diante trabalharemos com a seguinte equação:

$$Y_t = \delta \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i}. \quad (7)$$

Seguindo Campbell e Shiller (1987), o *spread* efetivo é definido como:

$$S_t \equiv Y_t - \frac{\delta}{1-\delta} y_t, \quad (8)$$

onde S_t é $I(0)$ com vetor de cointegração dado por $(1, \frac{-\delta}{1-\delta})$ se Y_t e y_t são cointegráveis. Subtraindo $\frac{\delta}{1-\delta} y_t$ de ambos os lados de (7), encontramos a seguinte relação para o valor do spread teórico S'_t :

$$S'_t = \frac{\delta}{1-\delta} \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t \Delta y_{t+i}. \quad (9)$$

Da equação acima temos que as séries teoricamente devem ser cointegradas pois o lado direito é uma função de termos $I(0)$ com pesos exponenciais decrescentes. Ainda, subtraindo $\delta \mathbb{E}_t Y_{t+1} = \delta \sum_{i=0}^{\infty} \delta^i \mathbb{E}_t y_{t+i+1}$ de Y_t em (7), obtemos:

$$Y_t = \delta \mathbb{E}_t Y_{t+1} + \delta y_t. \quad (10)$$

De (10), se adicionarmos e subtrairmos δY_t , levando a $Y_t = \delta \mathbb{E}_t \Delta Y_{t+1} + \delta Y_t + \delta y_t$, finalmente obtemos:

$$S''_t = \frac{\delta}{1-\delta} \mathbb{E}_t \Delta Y_{t+1}. \quad (11)$$

A equação (11) explicita o *spread* em função da previsão um passo a frente de ΔY_{t+1} . Podemos sempre performar a seguinte decomposição:

$$\Delta Y_{t+1} = \mathbb{E}_t \Delta Y_{t+1} + \underbrace{(\Delta Y_{t+1} - \mathbb{E}_t \Delta Y_{t+1})}_{u_{t+1}}. \quad (12)$$

Hecq e Issler (2012) mostram que substituindo (12) em (11), e defasando toda a equação por um período:

$$S_{t-1} = \frac{\delta}{1-\delta} \Delta Y_t + u_t \quad (13)$$

ou alternativamente,

$$\Delta Y_t = \frac{1-\delta}{\delta} S_{t-1} + v_t \quad (14)$$

onde u_t (or $v_t = -\frac{1-\delta}{\delta} u_t$) é ortogonal ao passado. De (14) também obtemos:

$$\begin{aligned} (1-\delta)S_{t-1} &= \delta\Delta Y_t + (1-\delta)u_t \\ S_{t-1} - \delta Y_{t-1} + \delta \frac{\delta}{1-\delta} y_{t-1} &= \delta Y_t - \delta Y_{t-1} + \left\{ \delta \frac{\delta}{1-\delta} y_t - \delta \frac{\delta}{1-\delta} y_t \right\} + (1-\delta)u_t \\ S_{t-1} &= \delta S_t + \delta \frac{\delta}{1-\delta} \Delta y_t + (1-\delta)u_t \end{aligned} \quad (15)$$

que resulta em

$$S_t = \frac{1}{\delta} S_{t-1} - \frac{\delta}{1-\delta} \Delta y_t + \varepsilon_t \quad (16)$$

com $\varepsilon_t = \frac{(1-\delta)}{\delta} u_t$.

Como explorado por Campbell (1987), no contexto de renda permanente, a equação diferencial estocástica de primeira ordem acima gera o modelo de valor presente. Hecq e Issler (2012) ressaltam que duas importantes condições podem ser derivadas das equações (16) à (9): Cointegração implica na condição de transversalidade $\lim_{k \rightarrow \infty} \delta^k \mathbb{E}_t(S_{t+k}) = 0$, enquanto a imprevisibilidade do ε_t em relação ao passado, isto é, $\mathbb{E}_t(\varepsilon_{t+1}) = 0$ garante que não haverá nenhum termo no lado direito da equação (9) invalidando-o. A primeira implica em uma restrição de longo prazo entre Y_t e y_t . A segunda, restringe a dinâmica da representação estacionária do sistema, fazendo S_t e Δy_t funções específicas do seu passado apenas. Então, pode ser vista como restrição de curto prazo no comportamento de S_t e Δy_t . Esses são os tipos de restrições estudados na literatura de ciclos comuns, que serão exploradas na próxima seção.

3 Restrições de Ciclos Comuns

Assumimos que o sistema bivariado $I(1)$ da série $(Y_t, y_t)'$ segue um $VAR(p)$ em nível, e que $S_t = Y_t - \theta y_t$ é o termo de correção de erro estacionário. Como já mencionado, no caso de preço e dividendo $\theta = \frac{\delta}{1-\delta}$. Logo, temos que o modelo de correção de erro (VECM) é representado por:

$$\begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \Gamma_1 \begin{pmatrix} \Delta Y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \dots + \Gamma_{p-1} \begin{pmatrix} \Delta Y_{t-p+1} \\ \Delta y_{t-p+1} \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} S_{t-1} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \quad (17)$$

onde assumimos que os distúrbios são ruídos brancos e que condições necessárias para evitar que a série seja $I(2)$ são atendidas. Os Γ_i s são os coeficientes matriciais de curto prazo, e α_1 e α_2 são os coeficientes que acompanham o termo de correção de erro.

Na literatura, é de conhecimento comum que relações de valor presente impõem restrições sobre modelos dinâmicos. Campbell e Shiller (1987) e outros exploraram o fato do VAR possuir restrições entre suas equações. No entanto, Hecq e Issler (2012), exploraram restrições de natureza diferente. Eles focaram em restrições de posto reduzido sobre modelos na forma VECM, permitindo a aplicação do ferramental existente na literatura de ciclos comuns ao lidarmos com modelos de valor presente.

Hecq e Issler (2012) estabeleceram a seguinte proposição:

Proposition 1. *Se os elementos de $(Y_t, y_t)'$ obedecem uma relação de VP como em (13), i.e., $S_{t-1} = \frac{\delta}{1-\delta} \Delta Y_t + u_t$, então, sua representação VECM obedece a forma fraca de ciclo comum (WSCCF) (veja Hecq et al. (2006) e Athanasopoulos et al. (2011)): existe um vetor 1×2 γ' tal que $\gamma \Gamma_1 = \gamma \Gamma_2 = \dots = \gamma \Gamma_{p-1} = 0$, mas $\gamma' \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \neq 0$. Mais ainda, $\gamma' = (1 : 0)$, a primeira linha de cada Γ_i , $i = 1 \dots p-1$, deve ser zero, e as seguintes restrições também devem ser atendidas: $\alpha_1 = \frac{1-\delta}{\delta}$.*

Para exemplificar, vamos tratar do caso mais simples, com $p = 1$:

$$\begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ * & * \end{pmatrix} \begin{pmatrix} \Delta Y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \begin{pmatrix} \frac{1}{\theta} \\ \alpha_2 \end{pmatrix} (Y_{t-1} - \theta y_{t-1}) + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (18)$$

Restrições: $(1 : 0) \begin{pmatrix} 0 & 0 \\ * & * \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ e $(1 : 0) \begin{pmatrix} \frac{1}{\theta} \\ \alpha_2 \end{pmatrix} = \frac{1}{\theta} \neq 0$. Então,

$$(1 : 0) \begin{pmatrix} \Delta Y_t \\ \Delta y_t \end{pmatrix} = \Delta Y_t = \frac{1}{\theta} S_{t-1} + u_{1t}.$$

Combinando essa com $S_t = Y_t - \frac{\delta}{1-\delta} y_t$, depois de alguma álgebra:

$$S_t = \frac{1}{\delta} S_{t-1} - \frac{\delta}{1-\delta} \Delta y_t + u_{1t}. \quad (19)$$

Sendo essa a equação de primeira ordem que gera $S_t = \frac{\delta}{1-\delta} \sum_{i=1}^{\infty} \delta^i \mathbb{E}_t \Delta y_{t+i}$.

Portanto, o modelo de Valor Presente implica cointegração e condições adicionais de ortogonalidade associado a restrições de posto em modelos na forma VECM.

4 Seleção do modelo e estimação

Nesta seção iremos tratar dos critérios utilizados para a seleção dos modelos e dos procedimentos de estimação empregados.

Athanasopoulos et al. (2011) conceberam uma estratégia ótima para a seleção conjunta do posto de cointegração q , posto das matrizes de curto prazo r e de defasagem p de um VAR contendo restrições de longo prazo e de curto prazo em sua forma fraca, estimando conjuntamente seus parâmetros. Adicionalmente eles mostraram como esse novo critério pode ser útil na obtenção de resultados mais precisos em previsões.

Na medida em que modelos de Valor Presente implicam em restrições de longo e curto prazo em sua forma fraca, a estratégia utilizada por Athanasopoulos et al. (2011) pode ser usada como ponto inicial para construir modelos de previsão com restrições de VP.

Primeiramente, vamos tratar da estimação por máxima verossimilhança de um VAR escrito na forma de correção de erro

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_p \Delta y_{t-p} + \eta_t \quad (20)$$

sob a restrição de longo prazo que o posto de Π é q , de curto prazo que o posto de $[\Gamma_1 \ \Gamma_2 \ \dots \ \Gamma_p]$ é r , e sob a hipótese de normalidade. Assim, podemos reescrever o modelo acima da seguinte forma:

$$\Delta y_t = \gamma \alpha' y_{t-1} + C [D_1 \Delta y_{t-1} + D_2 \Delta y_{t-2} + \cdots + D_p \Delta y_{t-p}] + \eta_t, \quad (21)$$

onde α é uma matriz $K \times q$ de posto q e C é uma matriz $K \times r$ de posto r .

Athanasopoulos et al. (2011) desenvolveram o seguinte procedimento para computar a estimação por máxima verossimilhança dos parâmetros da equação (21) acima:

- Passo 0. Estimar $[\hat{D}_1, \hat{D}_2, \dots, \hat{D}_p]$ de uma regressão na forma reduzida de Δy_t sobre $(\Delta y_{t-1}, \dots, \Delta y_{t-p})$ controlando para y_{t-1} . Essas estimações são os coeficientes das variáveis canônicas correspondentes as r maiores correlações canônicas parciais (PCCs) ao quadrado entre Δy_t e $(\Delta y_{t-1}, \dots, \Delta y_{t-p})$, controlando para y_{t-1} .
- Passo 1. Computar as PCCs entre Δy_t e y_{t-1} condicional a $[\hat{D}_1 \Delta y_{t-1} + \hat{D}_2 \Delta y_{t-2} + \cdots + \hat{D}_p \Delta y_{t-p}]$. Tomar as q variáveis canônicas $\hat{\alpha}' y_{t-1}$ correspondentes as q maiores PCCs ao quadrado como estimações das relações de cointegração. Regrida Δy_t em $\hat{\alpha}' y_{t-1}$ e $[\hat{D}_1 \Delta y_{t-1} + \hat{D}_2 \Delta y_{t-2} + \cdots + \hat{D}_p \Delta y_{t-p}]$, e compute $\ln |\hat{\Omega}|$, o logaritmo do determinante da matriz de variância do resíduo.
- Passo 2. Computar as PCCs entre Δy_t e $(\Delta y_{t-1}, \dots, \Delta y_{t-p})$ condicional a $\hat{\alpha}' y_{t-1}$. Tome as r variáveis canônicas $[\hat{D}_1 \Delta y_{t-1} + \hat{D}_2 \Delta y_{t-2} + \cdots + \hat{D}_p \Delta y_{t-p}]$ correspondentes as r maiores PCCs como estimações de $[D_1 \Delta y_{t-1} + D_2 \Delta y_{t-2} + \cdots + D_p \Delta y_{t-p}]$. Regrida Δy_t em $\hat{\alpha}' y_{t-1}$ e $[\hat{D}_1 \Delta y_{t-1} + \hat{D}_2 \Delta y_{t-2} + \cdots + \hat{D}_p \Delta y_{t-p}]$, e compute $\ln |\hat{\Omega}|$, o logaritmo do determinante da matriz de variância dos resíduos. Se for diferente do valor computado no passo 1, volte para o passo 1. Caso contrário, pare.

O valor de $\ln |\hat{\Omega}|$ se torna menor até atingir seu mínimo, o qual denotamos por $\ln |\hat{\Omega}_{p,r,q}|$. Os valores de $\hat{\alpha}$ e $[\hat{D}_1, \hat{D}_2, \dots, \hat{D}_p]$ no estágio final serão os estimadores de máxima verossimilhança de α e $[D_1, D_2, \dots, D_p]$. Os estimadores de máxima verossimilhança de outros parâmetros serão simplesmente os coeficientes da regressão final.

Entretanto, a estratégia de Athanasopoulos et al. (2011) só se aplica a restrições lineares, de modo que quando estamos tratando do modelo de Valor Presente, estamos falando de restrições não lineares. Sendo assim, uma solução *first best* de calcular todos os parâmetros conjuntamente não é possível. Optamos assim por uma solução *second best* condicionando no fato que os parâmetros de longo prazo são T -consistentes, comparados a usual \sqrt{T} consistência dos parâmetros de curto prazo. Assim, estimamos primeiramente os parâmetros de longo prazo pelo procedimento usual de Johansen (1988, 1991), posteriormente estimando os demais parâmetros de curto prazo por máxima verossimilhança condicional restrita.

Tratando agora da estratégia para seleção de modelos, Athanasopoulos et al. (2011), desenvolveram um procedimento de escolha em dois passos para q , r e p em contraponto aos critérios de seleção mais comuns. Esses, no contexto de VECM, consideravam escolher (p, r) por um dos critérios tradicionais e então usar uma sequência de testes de razão de verossimilhança para determinar o posto de cointegração q .

Os principais critérios tradicionais são os de Akaike (AIC), Hannan e Quinn (HQ) e o Bayesiano (BIC), definidos como:

$$AIC(p, r) = T \sum_{i=K-r+1}^K \ln \left(1 - \hat{\lambda}_i^2(p) \right) + 2(r(K-r) + rKp) \quad (22)$$

$$HQ(p, r) = T \sum_{i=K-r+1}^K \ln \left(1 - \hat{\lambda}_i^2(p) \right) + 2(r(K-r) + rKp) \ln \ln T \quad (23)$$

$$SC(p, r) = T \sum_{i=K-r+1}^K \ln \left(1 - \hat{\lambda}_i^2(p) \right) + (r(K-r) + rKp) \ln T, \quad (24)$$

onde K é a dimensão (números de série) do sistema, r é o posto de $[\Gamma_1 \ \Gamma_2 \ \dots \ \Gamma_p]$, p é o número de defasagens em diferença no VECM, T é o número de observações e $\hat{\lambda}_i^2(p)$ são as PCCs amostrais ao quadrado entre Δy_t e o conjunto de regressores $(\Delta y_{t-1}, \dots, \Delta y_{t-p})$ após a influência linear y_{t-1} ser retirada. Critérios de seleção tradicionais, que não consideram as restrições de curto prazo, são casos especiais em que os postos são cheios, isto é, r é igual a K .

Athanasopoulos et al. (2011) também consideram um outro critério de seleção baseado no contexto bayesiano. Esse procedimento se chama *Posterior Information Criteria* (PIC) e foi proposto por Phillips (1996) e Phillips e Ploberger (1996). Ele se apóia no fato que o processo gerador dos dados (DGP) não pertence a nenhuma família de modelos, e pode evoluir no tempo. Phillips (1996) apresenta uma aproximação assintótica da densidade *a priori*, que depois ele utiliza para aproximar o fator de Bayes, que será utilizado no critério em questão. Esse combina a qualidade do ajuste e a parcimônia dos modelos propostos na hora de escolher o modelo ótimo. Sua grande diferença em relação aos tradicionais AIC, BIC e HQ, reside no fato que ele penaliza os modelos maiores baseado na comparação da matriz de informação de Fisher, usando as informações redundantes adicionadas para punir o excesso de parametrização. Enquanto os demais, usam simplesmente o número de parâmetros a mais. Entretanto, existem dificuldades práticas em trabalhar com o PIC que motiva uma simplificação dessa metodologia.

Desse modo, Athanasopoulos et al. (2011) propõem um novo procedimento, realizado em duas etapas. Na primeira etapa, a influência linear de y_{t-1} é removida de Δy_t e

$(\Delta y_{t-1}, \dots, \Delta y_{t-p})$, então $HQ(p, r)$, como definido em (23), é usado para determinar p e r . Depois, o PIC é calculado para os valores escolhidos de p e r , para todos os valores de q , de 0 a K , assim chegando a escolha ótima de p , r e q .

Baseado nos procedimentos mencionados acima, os seguintes modelos foram selecionados para a comparação de previsões fora da amostra:

1. **Var(p+1) em nível:** Selecionar p por BIC, estimar por máxima verossimilhança condicional e realizar as previsões.
2. **VECM(HQ-PIC):** É um Var(p+1) possivelmente restrito por cointegração e/ou por WSCCF: Selecionar (p, r, q) conjuntamente por HQ-PIC Athanasopoulos et al. (2011), estimar todos os parâmetros pela máxima verossimilhança condicional em dois passos e realizar as previsões.
3. **VECM(AIC-J):** VECM(p) irrestrito: Escolher o p pelo AIC, estimar q e θ consistentemente pelo procedimento usual de Johansen (1988, 1991), e os demais parâmetros por máxima verossimilhança condicional com \hat{S}_t , e realizar as previsões.
4. **Modelo de VP:** VECM(p) restrito pelo modelo de VP ($q = 1, r = 1$): Selecionar p pelo procedimento HQ-PIC Athanasopoulos et al. (2011), estimar todos os parâmetros por máxima verossimilhança condicional restrita, com $\hat{\alpha}_1 = \frac{1-\hat{\delta}}{\hat{\delta}} = 1/\hat{\theta}$ estimado por Johansen (1988, 1991), $\Gamma_i = \begin{pmatrix} 0 & 0 \\ * & * \end{pmatrix} \forall i$, e realizar as previsões.

5 Dados dos preços agregados e dividendos

Os dados utilizados no artigo para a comparação entre os modelos foram extraídos da base de dados de Robert Shiller, cujos dados nominais consistem em observações do índice trimestral de preços composto *Standard & Poor*, utilizando interpolação linear para valores mensais. Os dados começam em Janeiro de 1871 extendendo-se até Março de 2012.

A série de dividendos correspondente a série de preços foram compiladas de duas fontes distintas. Começando em 1926, os dividendos nominais originais são dividendos por ação, ajustados ao índice. Entretanto, *Standard & Poor* não disponibiliza os dividendos anteriores ao ano de 1926. Para esses anos, Robert Shiller interpolou os dados anuais presentes em Cowles (1938), que fornece a série de dividendos correspondente ao índice, a qual ele multiplicou pela taxa da série em 1926 para ajustar a mudança de ano base.

Nossa estimação e previsão fora da amostra foi feita com esses dados ajustados pelo CPI (*Consumer Price Index*), tomando como base o seu valor em março de 2012, último mês observado. Assim, a série de preços e dividendos em valores reais utilizada para o mês i do ano x foi computada da seguinte forma:

$$P_{x(i)} = \frac{p_{x(i)} \times CPI_{2012(3)}}{CPI_{x(i)}} \quad (25)$$

onde $p_{x(i)}$ é o valor do preço ou dividendo nominal do índice.

Na figura (1) abaixo, podemos observar o comportamento dos dados utilizados. Notamos que ambas as séries arbitrariamente se afastam do seu valor inicial, dando indícios de um padrão não-estacionário. Entretanto, podemos perceber que ambas se movem juntas, mantendo-se uma distância relativamente fixa entre elas, mostrando que existe uma tendência comum, característica de processos cointegrados.

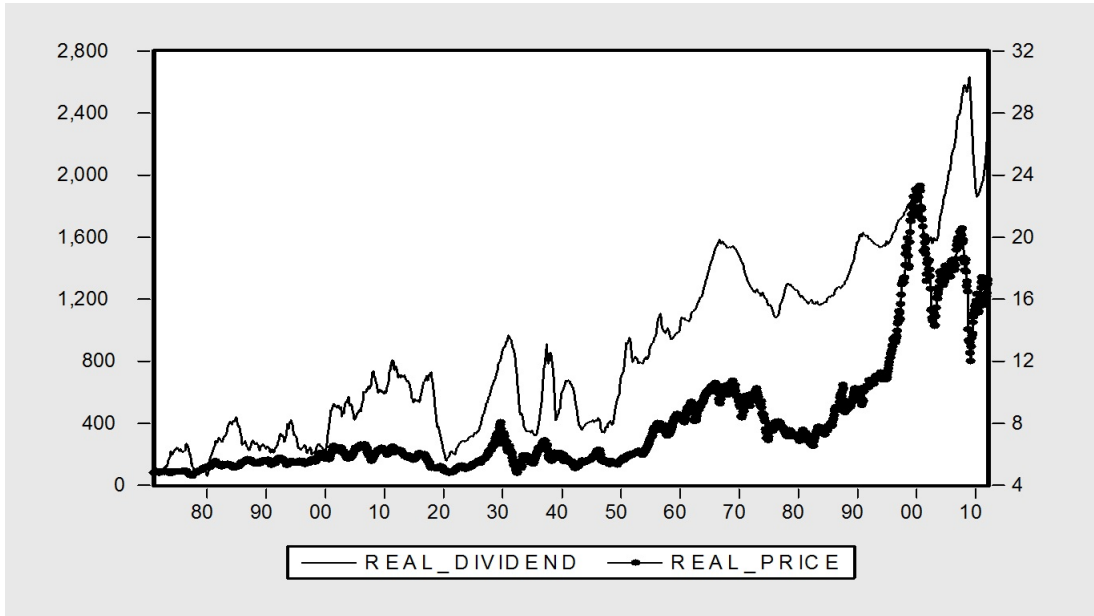


Figura 1: Preço Real e Dividendo Real (01.1871 - 03.2012)

Na figura (2), temos o comportamento do termo $S_t = Y_t - \theta y_t$, uma combinação linear de Y_t e y_t que oscila em torno de um valor médio, natureza comum a processos $I(0)$.

É pertinente observarmos que o parâmetro θ do vetor de cointegração apresentou uma variação bem grande, aproximadamente entre -520 até 1400 dependendo do intervalo de observações analisado.

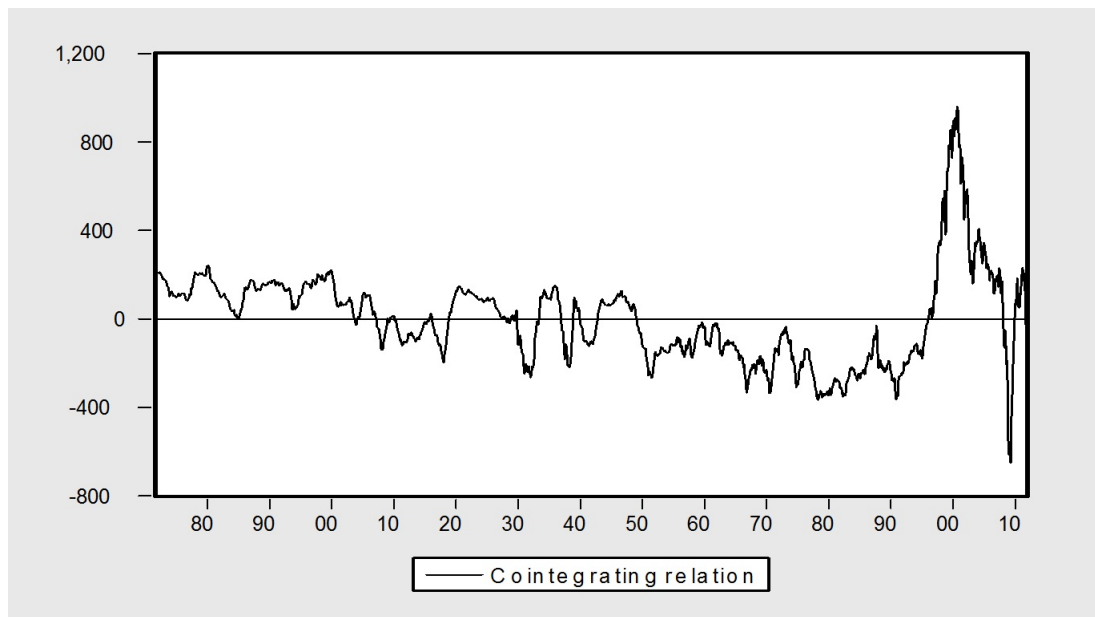


Figura 2: Relação de cointegração (01.1871 - 03.2012)

6 Teste fora da amostra

Giacomini e White (2006) desenvolveram um teste fora da amostra para comparar a precisão entre diferentes modelos de previsão. No nosso estudo iremos explorar as inovações metodológicas provenientes desse teste em relação aos existentes até então.

A primeira inovação significativa é que o teste considera previsões baseadas em estimadores de memória limitada, os quais se caracterizam pelas propriedades amostrais finitas serem preservadas assintoticamente. As principais vantagens desses estimadores são sua maior robustez a heterogeneidade dos dados e ao seu freqüente melhor desempenho em previsões acometidas pelo problema de mal especificação. Outra novidade relevante é que o teste em estudo considera também a questão de previsibilidade como um problema de inferência sobre expectativas condicionais e de erros de previsão.

Essa abordagem permite considerar os efeitos da incerteza da estimação na performance das previsões, de modo que não apenas o modelo será avaliado mas também o procedimento de estimação e os dados usados para tal. Para exemplificar, considere que queiramos comparar a precisão de modelos de previsão $f_t(\beta_1)$ e $g_t(\beta_2)$ para τ passos a frente da variável $Y_{t+\tau}$, usando a função de perda $L_{t+\tau}(\cdot)$. Pelo teste de Diebold e Mariano (1995),

$$H_0 : E[L_{t+\tau}(Y_{t+\tau}, f_t(\beta_1^*)) - L_{t+\tau}(Y_{t+\tau}, g_t(\beta_2^*))] = 0 \quad (26)$$

onde β_1^* e β_2^* são valores populacionais. Sua hipótese nula diz que os modelos são igualmente precisos em média.

No teste de Giacomini e White (2006), a função perda depende dos parâmetros estimados $\hat{\beta}_{1t}$ e $\hat{\beta}_{2t}$, ao invés dos seus limites de probabilidade; e a função expectativa é condicional a algum conjunto de informação G_t :

$$H_0 : E[L_{t+\tau}(Y_{t+\tau}, f_t(\hat{\beta}_{1t})) - L_{t+\tau}(Y_{t+\tau}, g_t(\hat{\beta}_{2t})) | G_t] = 0 \quad (27)$$

O foco nos parâmetros estimados faz o teste se referir aos modelos, bem como aos procedimentos de estimação e às possíveis escolhas das janelas de estimação. Sua hipótese nula diz que não é possível dizer qual modelo é mais preciso na previsão $t + \tau$ usando as informações em G_t .

No caso em que $\tau = 1$, Giacomini e White (2006) exploram o fato de que a hipótese nula é equivalente a estabelecer que $\Delta L_{m_1, m_2, t} = L_{t+\tau}(Y_{t+\tau}, f_t(\hat{\beta}_{1t})) - L_{t+\tau}(Y_{t+\tau}, g_t(\hat{\beta}_{2t}))$ é uma sequência martingale em diferenças, implicando que para todas as funções mensuráveis g_t no conjunto de informação no instante t , temos $E[g_t \cdot \Delta L_{m_1, m_2, t}] = 0$ sob H_0 .

Elaborando a primeira previsão de τ passos a frente no instante m , utilizando o procedimento de estimação por janela (*rolling-window*), Giacomini e White (2006) mostram que dado um conjunto q de funções mensuráveis h_t , a hipótese nula de igualdade de habilidade de previsão condicional para uma par de modelos M_1 e M_2 , pode ser testada usando a estatística:

$$T_{m, n, \tau}^h = n \left(n^{-1} \sum_{t=m}^{T-\tau} h_t \Delta L_{m, t+\tau} \right)' \tilde{\Omega}_n^{-1} \left(n^{-1} \sum_{t=m}^{T-\tau} h_t \Delta L_{m, t+\tau} \right) \quad (28)$$

onde h_t é uma função teste $q \times 1$ mensurável, $n \equiv T - \tau - m + 1$ e

$$\tilde{\Omega}_n \equiv n^{-1} \sum_{t=m}^{T-\tau} h_t \Delta L_{m,t+\tau} \left(\sum_{t=m}^{T-\tau} h_t \Delta L_{m,t+\tau} \right)' + n^{-1} \sum_{j=1}^{\tau-1} w_{n,j} \times \sum_{t=m+j}^{T-\tau} [h_t \Delta L_{m,t+\tau} (h_t \Delta L_{m,t+\tau-j})' + h_t \Delta L_{m,t+\tau-j} (h_t \Delta L_{m,t+\tau})']$$

onde $w_{n,j}$ é uma função de ponderação tal que $w_{n,j} \rightarrow 1$ quando $n \rightarrow \infty$ para cada $j = 1, \dots, \tau - 1$.

No caso especial de $\tau = 1$, temos que $\tilde{\Omega}_n$ simplifica para $\hat{\Omega}_n \equiv n^{-1} \sum_{t=m}^{T-1} h_t \Delta L_{m,t+1} \times (h_t \Delta L_{m,t+1})'$. Um teste α rejeita a hipótese nula sempre que $T_{m,n,\tau}^h > \chi_{q,1-\alpha}^2$, onde $\chi_{q,1-\alpha}^2$ é o quantil $(1 - \alpha)$ da distribuição χ_q^2 .

Uma última vantagem do teste de Giacomini e White (2006) é permitir um tratamento único para modelos *nested* e *nonnested*, independentemente do procedimento de estimação usado para derivação dos modelos usados na previsão, enquanto muitos outros testes, como Diebold e Mariano (1995), não se aplicam para modelos *nested*.

6.1 Resultados da Estimação

Para facilitar a comparação, os resultados são expostos em função da raiz do erro quadrático médio de previsão (RMSFE) do preço de um determinado modelo em relação ao modelo VAR em nível:

$$RRMSFE_h^M = \frac{RMSFE_h^M}{RMSFE_h^{VAR}} \quad (29)$$

onde $RMSFE_h^M = \sum_{t=T_1}^{T_2-h} (Y_{t+h}^h - \hat{Y}_{M,t+h|t}^h)^2$. Um valor de RRMSFE menor que 1 indica que o modelo em questão performou melhor que o VAR em nível na precisão da previsão fora da amostra de estimação. Os resultados estão resumidos nas tabelas (6.1) e (6.1) abaixo.

Para garantir um resultado mais robusto, nós dividimos a amostra de previsão em um período anterior às crises “com” e financeira de 2008-2009, com números até 1994; e outro que compreende dados até 2012, de modo a investigar se houve alguma alteração nas performances relativas de previsão dos modelos.

Na tabelas (1), (2), (3), (4) encontramos os resultados para o teste de igual capacidade preditiva condicional do agregado de preços de ações negociadas nos Estados Unidos, para a amostra até 2012 e 1994 respectivamente. Seguindo Giacomini e White (2006), as entradas principais correspondem ao p-valor do teste. O número entre parênteses corresponde ao percentual das vezes que o modelo na linha foi escolhido em detrimento ao da coluna, isto é, apresentou menor valor da função perda. O sinal corresponde se a média dos erros de previsão quadráticos foram maiores para o modelo na linha (sinal positivo) ou na coluna (sinal negativo). Finalmente, os números com * são aqueles significativos a 10%.

Observando a amostra com dados até 2012, os resultados apontam que em geral não há diferença estatisticamente significativa entre os modelos em questão. Diante disso, pode-se dizer que fica difícil distinguir a performance dos modelos em relação a média dos erros

	Passo=1			Passo=2			Passo=3			Passo=4		
	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP
VAR	0.731 [0.497]−	0.788 [0.444]+	0.360 [0.521]−	0.527 [0.491]+	0.227 [0.465]+	0.797 [0.510]−	0.361 [0.500]+	0.387 [0.474]+	0.545 [0.528]−	0.238 [0.495]+	0.098* [0.472]+	0.235 [0.512]−
HQ-PIC		0.345	0.476		0.372	0.608		0.871	0.464		0.502	0.271
AIC-J		[0.355]+	[0.512]−		[0.359]−	[0.516]−		[0.369]−	[0.531]−		[0.366]−	[0.537]−
			0.399			0.538			0.466			0.213
			[0.523]−			[0.516]−			[0.530]−			[0.519]−

Nota: Resultados do teste de igualdade condicional de habilidade para previsão realizados dois a dois. As entradas são os p valores do teste de Gacomini e White de igualdade condicional de habilidade para previsão entre os métodos da respectiva linha e coluna. A perda é quadrática e a função teste é $h_t = (1, \Delta L_{m,t})'$. O número entre parentêses corresponde ao percentual de vezes que o modelo na linha foi escolhido em detrimento ao modelo da coluna. Os números com * indicam que o resultado foi significativo a 10%. O sinal corresponde se a média dos erros de previsão ao quadrados foram maiores para o modelo na linha (sinal positivo) ou na coluna (sinal negativo).

Tabela 1: Teste condicional de habilidade de previsão Giacomini-White até 2012

	Passo=5			Passo=6			Passo=7			Passo=8		
	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP
VAR	0.817 [0.470] ⁺	0.629 [0.455] ⁺	0.301 [0.521] ⁻	0.140 [0.469] ⁺	0.103 [0.463] ⁺	0.348 [0.519] ⁻	0.551 [0.460] ⁺	0.441 [0.463] ⁺	0.463 [0.510] ⁻	0.173 [0.455] ⁺	0.128 [0.449] ⁺	0.358 [0.503] ⁻
HQ-PIC		0.733 [0.369] ⁺	0.397 [0.538] ⁻		0.160 [0.387] ⁻	0.418 [0.531] ⁻		0.834 [0.378] ⁻	0.492 [0.523] ⁻		0.266 [0.380] ⁻	0.510 [0.512] ⁻
AIC-J			0.342 [0.521] ⁻			0.379 [0.524] ⁻			0.437 [0.517] ⁻			0.545 [0.512] ⁻

Nota: Resultados do teste de igualdade condicional de habilidade para previsão realizados dois a dois. As entradas são os p valores do teste de Giacomini e White de igualdade condicional de habilidade para previsão entre os métodos da respectiva linha e coluna. A perda é quadrática e a função teste é $h_t = (1, \Delta L_{m,t})'$. O número entre parentêses corresponde ao percentual de vezes que o modelo na linha foi escolhido em detrimento ao modelo da coluna. Os números com * indicam que o resultado foi significativo a 10%. O sinal corresponde se a média dos erros de previsão ao quadrados foram maiores para o modelo na linha (sinal positivo) ou na coluna (sinal negativo).

Tabela 2: Teste condicional de habilidade de previsão Giacomini-White até 2012

	Passo=1			Passo=2			Passo=3			Passo=4		
	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP
VAR	0.385 [0.472] ⁺	0.137 [0.469] ⁺	0.085* [0.524] ⁻	0.338 [0.495] ⁺	0.404 [0.486] ⁺	0.385 [0.526] ⁻	0.232 [0.495] ⁺	0.189 [0.495] ⁺	0.767 [0.526] ⁻	0.576 [0.500] ⁺	0.194 [0.507] ⁺	0.358 [0.535] ⁻
HQ-PIC		0.523	0.014*		0.850	0.034*		0.297	0.035*		0.821	0.002*
AIC-J		[0.415] ⁻	[0.521] ⁻ 0.030*		[0.401] ⁻	[0.516] ⁻ 0.021*		[0.404] ⁻	[0.523] ⁻ 0.102		[0.418] ⁺	[0.538] ⁻ 0.005*
			[0.535] ⁻			[0.538] ⁻			[0.535] ⁻			[0.531] ⁻

Nota: Resultados do teste de igualdade condicional de habilidade para previsão realizados dois a dois. As entradas são os p valores do teste de Giacomini e White de igualdade condicional de habilidade para previsão entre os métodos da respectiva linha e coluna. A perda é quadrática e a função teste é $h_t = (1, \Delta L_{m,t})'$. O número entre parentêses corresponde ao percentual de vezes que o modelo na linha foi escolhido em detrimento ao modelo da coluna. Os números com * indicam que o resultado foi significativo a 10%. O sinal corresponde se a média dos erros de previsão ao quadrados foram maiores para o modelo na linha (sinal positivo) ou na coluna (sinal negativo).

Tabela 3: Teste condicional de habilidade de previsão Giacomini-White até 1994

	Passo=5			Passo=6			Passo=7			Passo=8		
	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP	HQ-PIC	AIC-J	VP
VAR	0.494 [0.479] ⁺	0.177 [0.476] ⁺	0.358 [0.533] ⁻	0.325 [0.474] ⁺	0.098* [0.476] ⁺	0.158 [0.549] ⁻	0.351 [0.474] ⁺	0.267 [0.470] ⁺	0.089* [0.540] ⁻	0.567 [0.481] ⁺	0.403 [0.463] ⁺	0.306 [0.535] ⁻
HQ-PIC		0.911 [0.423] ⁺	0.016* [0.542] ⁻		0.730 [0.418] ⁺	0.006* [0.563] ⁻		0.723 [0.425] ⁻	0.005* [0.547] ⁻		0.390 [0.390] ⁺	0.003* [0.551] ⁻
AIC-J			0.014* [0.544] ⁻			0.011* [0.542] ⁻			0.004* [0.557] ⁻			0.003* [0.556] ⁻

Nota: Resultados do teste de igualdade condicional de habilidade para previsão realizados dois a dois. As entradas são os p valores do teste de Giacomini e White de igualdade condicional de habilidade para previsão entre os métodos da respectiva linha e coluna. A perda é quadrática e a função teste é $h_t = (1, \Delta L_{m,t})'$. O número entre parênteses corresponde ao percentual de vezes que o modelo na linha foi escolhido em detrimento ao modelo da coluna. Os números com * indicam que o resultado foi significativo a 10%. O sinal corresponde se a média dos erros de previsão ao quadrados foram maiores para o modelo na linha (sinal positivo) ou na coluna (sinal negativo).

Tabela 4: Teste condicional de habilidade de previsão Giacomini-White até 1994

quadráticos de previsão fora da amostra. Entretanto, o modelo AIC-J demonstrou uma superioridade estatisticamente significativa em relação ao VAR em nível para a previsão quatro passos a frente, e praticamente significativa para seis passos a frente. Nessas observações, o VAR em nível obteve um maior erro quadrático médio e alcançou um maior erro quadrático de previsão entre 53 e 54% das vezes, evidenciando a superioridade do AIC-J.

Considerando agora as observações até 1994, notamos diferenças no padrão dos resultados. O modelo de valor presente (VP) se mostrou constantemente pior que os demais, com resultados estatisticamente significativos. Em relação ao VAR, seu desempenho foi inferior para um e sete passos a frente, sendo preterido entre 52 e 54% das vezes, apresentando erro quadrático médio maior. Quando comparado aos modelos AIC-J e HQ-PIC, o modelo VP foi estatisticamente pior em praticamente todos os casos, obtendo maior erro quadrático médio entre 52 e 56% das vezes. A única exceção foi a previsão 6 passos a frente, situação em que o resultado por um pouco não foi significativo. Além desses, o único resultado significativo foi a superioridade do AIC-J em relação ao Var para seis passos a frente.

Analisando os resultados das tabelas (6.1) e (6.1) notamos que o comportamento do RRMSFE relativo nos dois subperíodos apresentou um padrão semelhante. Os modelos VECM (HQ-PIC) e VECM(AIC-J) apresentaram resultados ligeiramente superiores ao *benchmark* VAR em nível. Com exceção da previsão 1 passo a frente para o modelo VECM(HQ-PIC) na amostra com dados até 2012.3, ambos os modelos se mostraram mais eficazes que o VAR em nível em todas as demais previsões para os dois subperíodos. Os ganhos variaram entre 1 e 2%. Já para o modelo de valor presente (VP), a precisão das previsões realizadas foi consistentemente inferior à obtida pelo VAR, tanto para a amostra completa quanto para àquela com dados até 1994.

As estrelas nas tabelas (6.1) e (6.1)) indicam rejeição a 10% para o teste de Giacomini e White (2006). Como podemos perceber, apesar dos modelos VECM(HQ-PIC) e VECM(AIC-J) apresentarem valores de RRMSFE menores que 1 em quase todos os casos, apenas em poucos deles os resultados são estatisticamente significativos. Em relação ao modelo de VP este apresentou RRMSE maior do que 1 em todos os casos, entretanto, só obtivemos significância para um e sete passos a frente na amostra até 1994.

Modelos	Horizonte							
	1	2	3	4	5	6	7	8
VECM(HQ-PIC)	1.001	0.991	0.987	0.984	0.995	0.983	0.985	0.980
VECM(AIC-J)	0.998	0.991	0.988	0.990*	0.991	0.991	0.988	0.992
PV	1.034	1.008	1.027	1.011	1.033	1.070	1.058	1.031

RRMSFE relativo das previsões dos modelos em relação ao *benchmark* VAR em nível. O símbolo,*, indica significância a 10% no teste de Giacomini e White (2006).

Tabela 5: RRMSFE relativo ao VAR para a amostra até 2012

Modelos	Horizonte							
	1	2	3	4	5	6	7	8
VECM(HQ-PIC)	0.992	0.991	0.988	0.992	0.992	0.990	0.990	0.994
VECM(AIC-J)	0.996	0.993	0.990	0.991*	0.991	0.990*	0.989	0.992
PV	1.046*	1.010	1.009	1.013	1.033	1.015	1.016*	1.020

RMSFE relativo das previsões dos modelos em relação ao *benchmark* VAR em nível. O símbolo,*, indica significância a 10% no teste de Giacomini e White (2006).

Tabela 6: RMSFE relativo ao VAR para a amostra até 1994

No apêndice desse artigo, encontramos gráficos das raízes dos erros quadráticos de previsão dos modelos em questão para um passo à frente. Comparamos os modelos sempre em relação ao *benchmark* VAR em nível. Em uma primeira análise, constatamos uma discrepância muito grande entre os resultados anteriores a crise e aqueles que incluem as crises “com” e financeira. A média das raízes dos erros quadráticos de previsão subiu abruptamente, algo em torno de 3 vezes ao que era observado no período pré-crises. Nenhum dos modelos foi capaz de ter um bom desempenho durante os períodos de crise, de modo que eles se mostraram muito mais úteis em períodos com baixa volatilidade.

Outra informação relevante que pode ser extraída dos gráficos é o melhor desempenho do modelo com restrições de VP frente ao VAR em determinados momentos. Apesar do pior desempenho do VP como um todo nas previsões fora da amostra, esse resultado não foi absoluto, apresentando variações ao longo do tempo.

Podemos concluir que os modelos VECM(HQ-PIC) e VECM(AIC-J), impondo restrições parciais apresentaram melhor desempenho no geral, sobretudo este último que obteve alguns resultados estatisticamente superiores ao VAR. Por outro lado, o modelo de valor presente (VP) teve a pior performance, sobretudo no período até 1994, quando as diferenças foram significativas. Entretanto, os ganhos de forma geral foram pequenos e excluindo da comparação o VP, poucas vezes significativos. Outro fato relevante foi a diferença acentuada de precisão de todos os modelos durante os períodos de crise e os de estabilidade, sendo neste último drasticamente mais precisos.

7 Conclusão

Recentemente, tem se tornado um assunto muito debatido em econometria a robustez e as limitações de modelos estruturais e na forma reduzida. Apesar de muito ser debatido a respeito de hipóteses e teorias, relativamente pouco ainda é feito sobre aplicações empíricas desses modelos. Nesse artigo, buscamos dar uma contribuição a respeito da questão se restrições estruturais ajudam ao realizarmos previsões.

Com esse objetivo, exploramos as relações impostas por modelos de Valor Presente. Discutimos tanto as relações de cointegração impostas pelo modelo que são bem conhecidas desde o trabalho de Campbell e Shiller (1987), quanto restrições fracas de ciclo comum, expostas recentemente por Hecq e Issler (2012). Ambas as condições são necessárias para validar modelos de Valor Presente. Sendo assim, trabalhamos em um cenário que ambos os fenômenos foram incorporados.

Com o intuito de verificarmos se a introdução de restrições impostas pelo modelo de Valor Presente melhoram as previsões, comparamos esse modelo de VP restrito com modelos mais relaxados. Os demais modelos utilizados foram os seguintes: Var(p+1), VECM(AIC-J), VECM(HQ-PIC). Sendo o primeiro um modelo irrestrito e os demais modelos parcialmente restritos. Para a seleção das defasagens do modelo, do posto de cointegração e do posto das matrizes de curto prazo, utilizamos um procedimento ótimo de escolha conjunta desenvolvido em Athanasopoulos et al. (2011) e também critérios mais tradicionais AIC, BIC e o usual Johansen (1988, 1991).

Tendo em mãos os modelos a serem utilizados, realizamos previsões fora da amostra baseados em dados de preços e dividendos agregados, extraídos da base de dados de Robert Shiller, compreendidos entre 1871 e 2012. Para a comparação entre os resultados obtidos por esses modelos, utilizamos o teste de Giacomini e White (2006), que testa a igualdade de habilidade de previsão condicional.

Encontramos que os modelos VECM(HQ-PIC) e VECM(AIC-J), impondo restrições parciais apresentaram melhor desempenho de um modo geral, com destaque para o último que apresentou resultados mais significantes do ponto de vista estatístico. O modelo com todas as restrições de VP, se mostrou o menos preciso, apesar de momentos de superioridade em relação aos demais. Cabe ressaltar também o fato de que todos os modelos se comportaram extremamente melhores na amostra que considera dados até 1994 apenas, período anterior as mais recentes crises mundiais. Desse modo, podemos concluir que os modelos trabalham melhor em períodos de estabilidade do que em períodos de alta volatilidade.

Referências

- ATHANASOPOULOS, G., DE CARVALHO GUILLÉN, O. T., ISSLER, J. V. e VAHID, F. (2011): “Model selection, estimation and forecasting in VAR models with short-run and long-run restrictions,” *Journal of Econometrics*, 164(1), 116 – 129, [jce:titlejAnnals Issue on Forecastingj/ce:titlej](#).
- CAMPBELL, J., LO, A. e MACKINLAY, A. (1997): *The econometrics of financial markets*. Princeton Univ. Press, Princeton, NJ, 2. print. with corrections edn.
- CAMPBELL, J. Y. (1987): “Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis,” *Econometrica*, 55(6), 1249–73.
- CAMPBELL, J. Y. e SHILLER, R. J. (1987): “Cointegration and Tests of Present Value Models,” *Journal of Political Economy*, 95(5), pp. 1062–1088.
- COWLES, A. (1938): *Common-stock indexes, 1871-1937*, Monograph (Cowles Commission for Research in Economics). Principia press, inc.
- DIEBOLD, F. X. e MARIANO, R. S. (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13(3), 253–63.
- ENDERS, W. (2004): *Applied econometric time series*, Wiley series in probability and statistics. Wiley, Hoboken, NJ, 2. ed. edn.
- ENGLE, R. F. e KOZICKI, S. (1993): “Testing for Common Features,” *Journal of Business & Economic Statistics*, 11(4), pp. 369–380.
- GIACOMINI, R. e WHITE, H. (2006): “Tests of Conditional Predictive Ability,” *Econometrica*, 74(6), 1545–1578.
- GOYAL, A. e WELCH, I. (2004): “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction,” NBER Working Papers 10483, National Bureau of Economic Research, Inc.
- HAMILTON, J. (1994): *Time series analysis*. Princeton Univ. Press, Princeton, NJ.
- HECQ, A. e ISSLER, J. V. (2012): “A Common-Feature Approach for Testing Present-Value Restrictions with Financial Data,” Economics Working Papers (Ensaio Economicos da EPGE) 728, Graduate School of Economics, Getulio Vargas Foundation (Brazil).
- HECQ, A., PALM, F. C. e URBAIN, J.-P. (2006): “Common cyclical features analysis in VAR models with cointegration,” *Journal of Econometrics*, 132(1), 117 – 141, [jce:titlejCommon Featuresj/ce:titlej](#).
- ISSLER, J. V. e VAHID, F. (2001): “Common cycles and the importance of transitory shocks to macroeconomic aggregates,” *Journal of Monetary Economics*, 47(3), 449–475.
- JOHANSEN, S. (1988): “Statistical analysis of cointegration vectors,” *Journal of Economic Dynamics and Control*, 12, 231 – 254.
- (1991): “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models,” *Econometrica*, 59(6), pp. 1551–1580.

- (2000): “Modelling of cointegration in the vector autoregressive model,” *Economic Modelling*, 17(3), 359 – 373.
- JOHANSEN, S. e SWENSEN, A. R. (1999): “Testing exact rational expectations in cointegrated vector autoregressive models,” *Journal of Econometrics*, 93(1), 73–91.
- LETTAU, M. e LUDVIGSON, S. C. (2001): “Consumption, Aggregate Wealth, and Expected Stock Returns,” *Journal of Finance*, 56(3), 815–849.
- (2004): “Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption,” *American Economic Review*, 94(1), 276–299.
- PHILLIPS, P. C. B. (1996): “Econometric Model Determination,” *Econometrica*, 64(4), pp. 763–812.
- PHILLIPS, P. C. B. e PLOBERGER, W. (1996): “An Asymtotic Theory of Bayesian Inference for Time Series,” *Econometrica*, 64(2), pp. 381–412.
- VAHID, F. e ENGLE, R. F. (1993): “Common Trends and Common Cycles,” *Journal of Applied Econometrics*, 8(4), 341–60.

A Apêndice

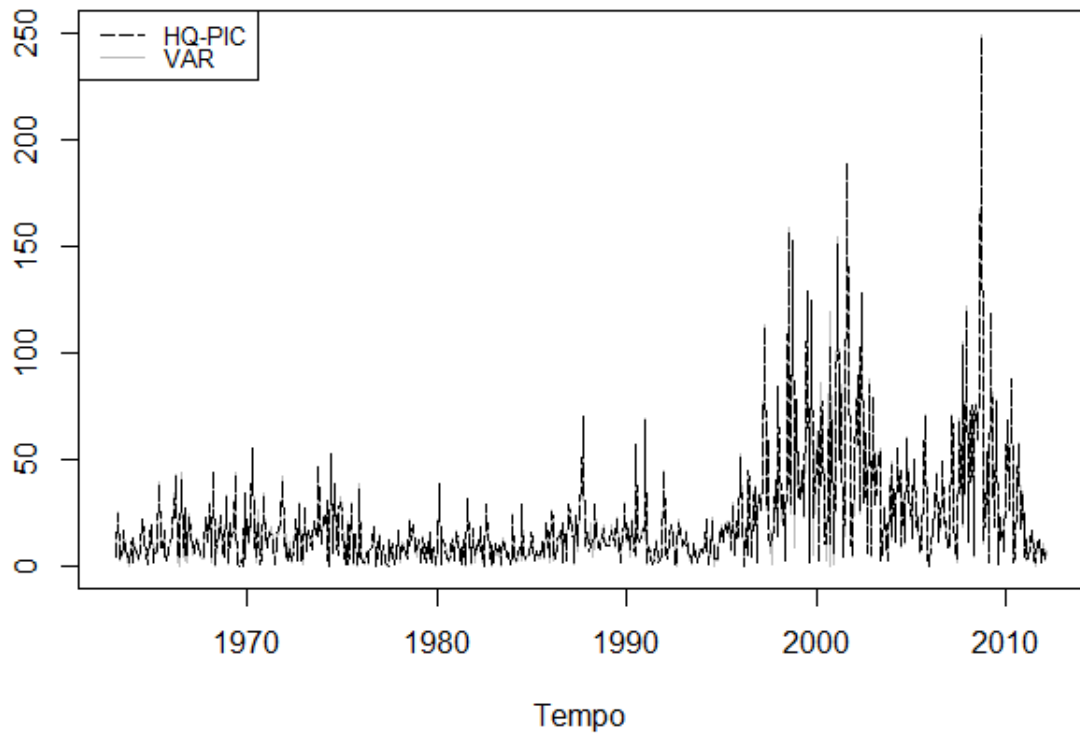


Figura 3: Raíz dos erros quadráticos do VAR e do HQ-PIC para 1 passo a frente até 2012

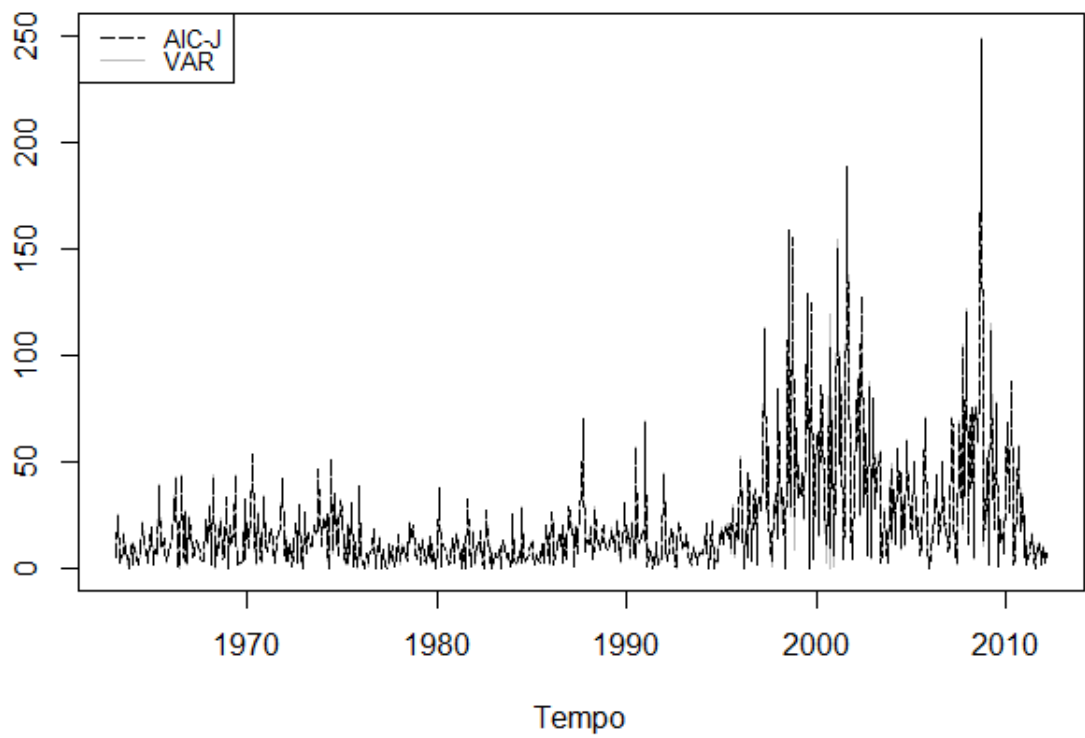


Figura 4: Raíz dos erros quadráticos do VAR e do AIC-J para 1 passo a frente até 2012

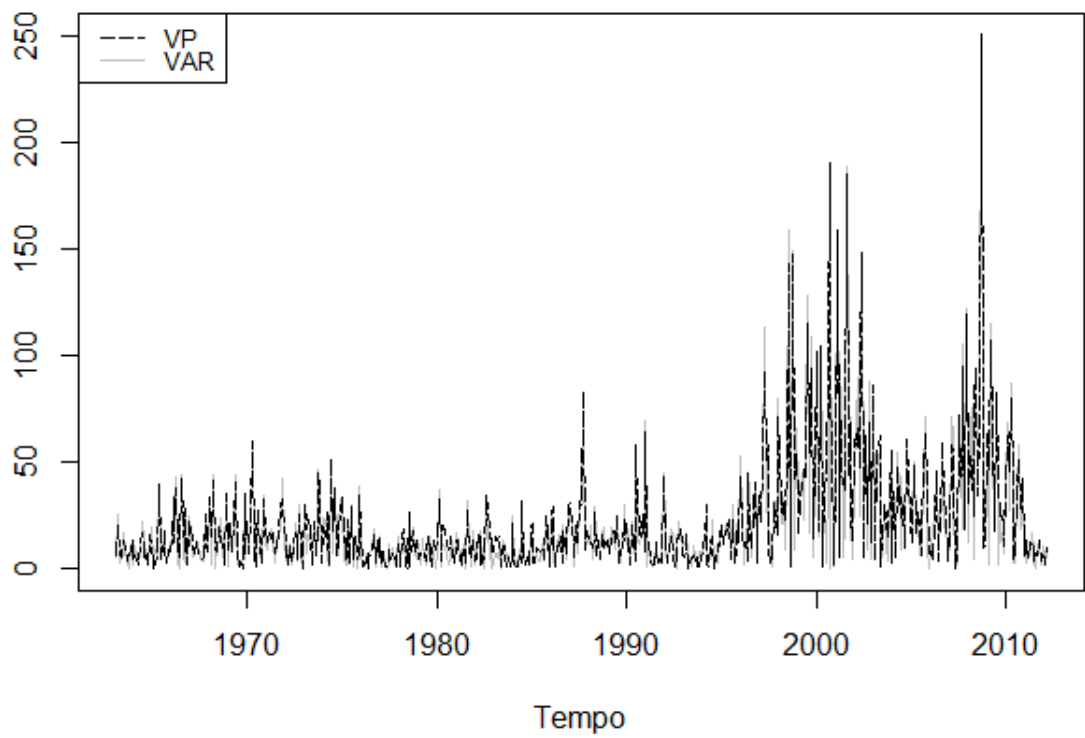


Figura 5: Raíz dos erros quadráticos do VAR e do VP para 1 passo a frente até 2012

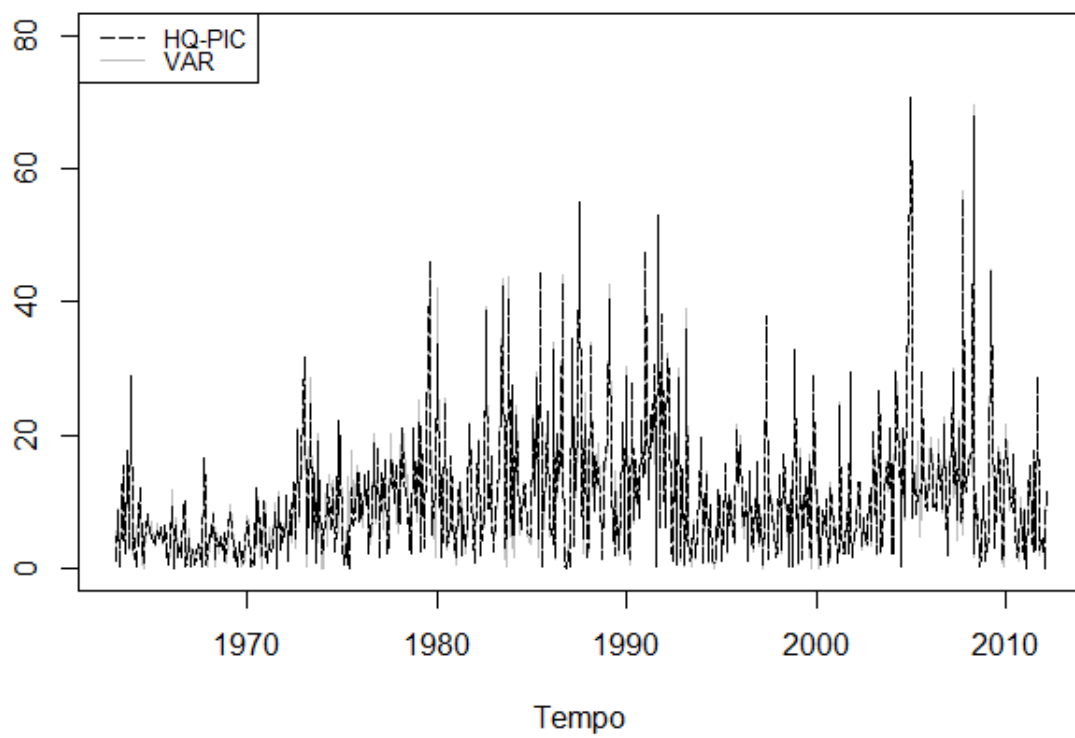


Figura 6: Raíz dos erros quadráticos do VAR e do HQ-PIC para 1 passo a frente até 1994

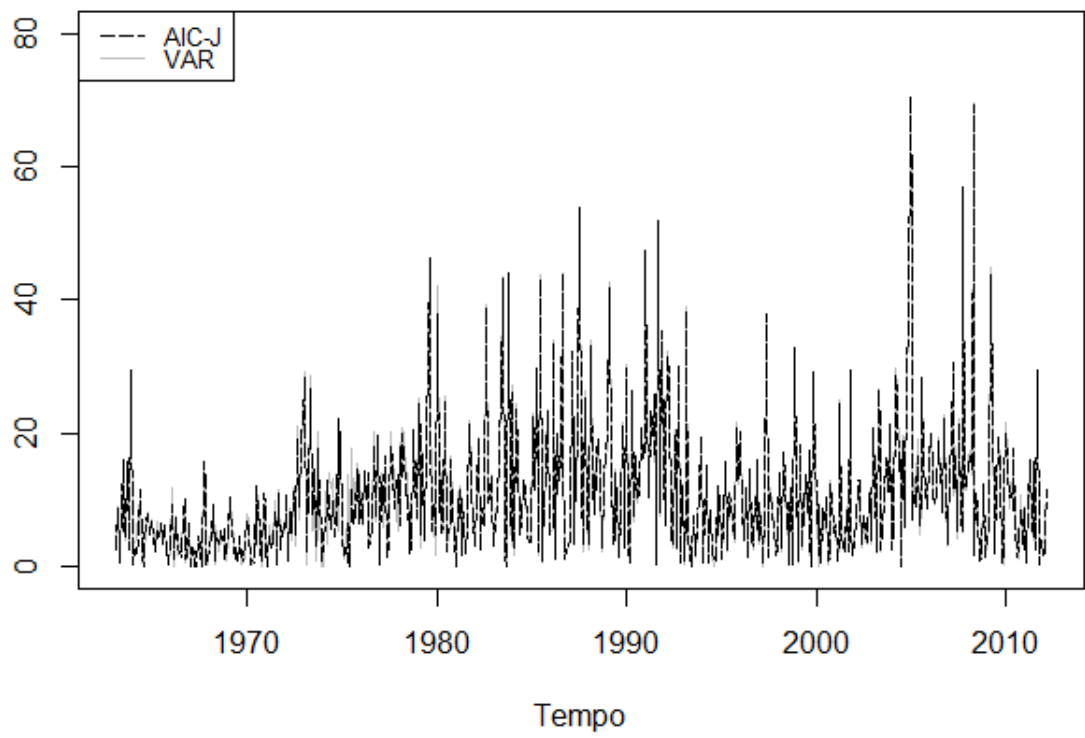


Figura 7: Raíz dos erros quadráticos do VAR e do AIC-J para 1 passo a frente até 1994

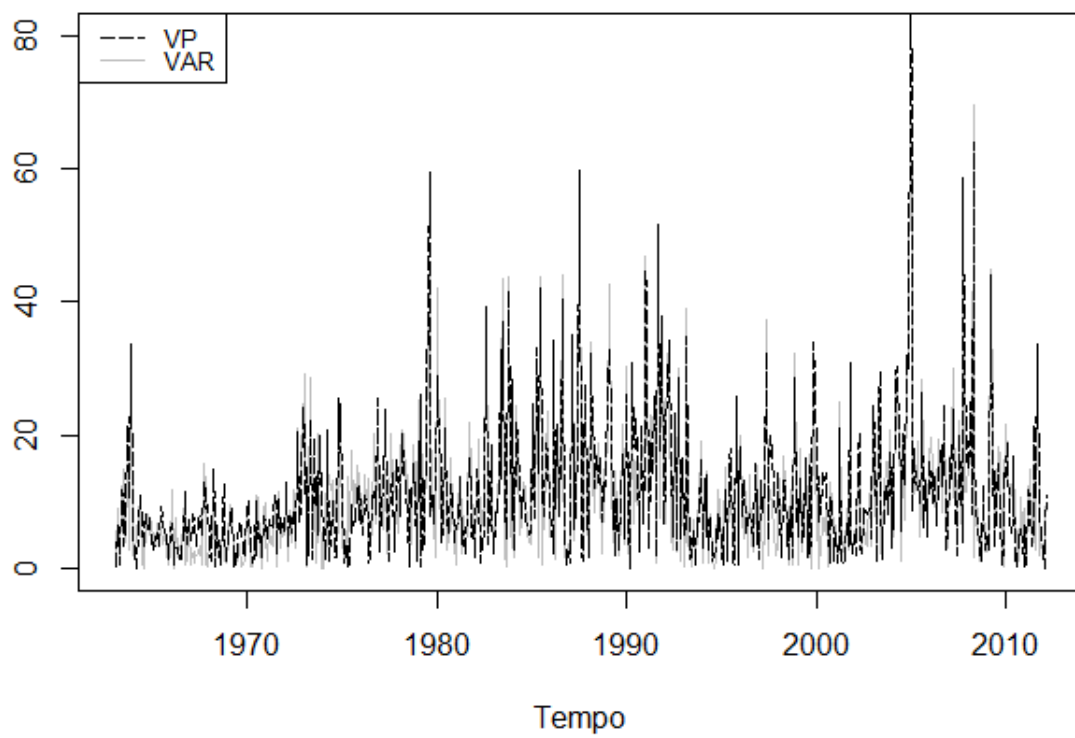


Figura 8: Raíz dos erros quadráticos do VAR e do VP para 1 passo a frente até 1994