

Econometria III

Professor Thiago Morello

Nota de aula 3

(referências: Bueno, capítulos 2 e 3; Enders, cap.2)

1 Previsão com modelos ARMA

1.1 Introdução e caso do processo estocástico AR(1)

Prever uma série de tempo é atribuir, com base nos dados disponíveis até o momento corrente, valores a instantes futuros de tempo. Trata-se, por exemplo, de atribuir um valor ao PIB brasileiro de 2027, tomando-se por base a série de valores do PIB brasileiro disponível até o ano corrente de 2026. Há diversas maneiras de converter a série disponível em uma previsão e os modelos ARMA fornecem fórmulas específicas para tanto. Tais fórmulas variam com a especificação ARMA.

As fórmulas de previsão referentes à família ARMA comumente utilizadas correspondem ao que se entende por previsor ótimo. Ou seja, tais fórmulas minimizam o erro de previsão. E, pois, assim como no caso da regressão linear em dados de corte transversal, as fórmulas do modelo ARMA são obtidas a partir da expectativa do valor futuro da série condicional à informação disponível até o instante corrente – pois tal expectativa condicional é o previsor que minimiza o erro quadrático médio (Hamilton, 1994, p.72). A informação disponível no instante t é igual aos valores da série de tempo observados até então, i.e., $y_t, y_{t-1}, y_{t-2}, \dots, y_0$. Será utilizada a notação sintética I_t para representar essa sequência de valores.

A previsão para o valor futuro um passo à frente da série é $E[y_{t+1}|I_t]$. Assim como em Bueno (2008), seja considerada inicialmente a especificação $ARMA(1,0) = AR(1) : y_t = c + \phi y_{t-1} + e_t$. Ou seja, suponha-se que a metodologia de Box-Jenkins apontou que a série de tempo empírica sob análise é mais bem reproduzida pela especificação $AR(1)$. Neste caso:

$$E[y_{t+1}|I_t] = E[(c + \phi y_t + e_{t+1})|I_t] = c + \phi E[y_t|I_t] + \phi E[e_{t+1}|I_t].$$

Neste ponto nota-se que (i) $E[y_t|I_t] = E[y_t|y_t, y_{t-1}, y_{t-2}, \dots, y_0]^1 = y_t$ (pois condiciona-se no próprio valor de y_t , o qual já foi observado, sendo pois conhecido no instante “ t ”) e (ii) $E[e_{t+1}|I_t] = E[e_{t+1}] = 0$ (pois a perturbação em t , por ser um ruído branco, não possui relação com as perturbações em instantes anteriores $t-1, t-2$, as quais compõem os

¹ Enders (2015, p.79) escreve a expectativa condicional à informação disponível até t como incluindo em tal informação também os valores das perturbações. Porém, eu acredito que tal notação é pouco esclarecedora uma vez que as perturbações nunca são observadas.

valores observados da série; e, na última passagem considera-se que um ruído branco tem expectativa nula).

Com isso:

$$E[y_{t+1}|I_t] = c + \phi y_t$$

Uma vez que os parâmetros c e ϕ são estimados como parte da metodologia de Box-Jenkins, a única informação que o previsor que se acaba de calcular requer, para prever o valor um passo à frente, é o valor corrente da série.

Cabe calcular os previsores (ou fórmulas de previsão) para valores dois passos e três passos adiante.

Dois passos adiante:

$$E[y_{t+2}|I_t] = E[c + \phi y_{t+1} + e_{t+2}|I_t] = c + \phi E[y_{t+1}|I_t] + E[e_{t+2}|I_t] = \{1\} = c + \phi(c + \phi y_t) + 0 = c(1 + \phi) + \phi^2 y_t.$$

{1} A expectativa condicional $E[y_{t+1}|I_t]$ é o previsor ótimo um passo adiante, e foi substituída pela fórmula calculada anteriormente.

Três passos adiante:

$$E[y_{t+3}|I_t] = E[c + \phi y_{t+2}|I_t] = c + \phi E[y_{t+2}|I_t] = \{1\} = c + \phi(c(1 + \phi) + \phi^2 y_t) = c(1 + \phi + \phi^2) + \phi^3 y_t.$$

{1} Novamente foi feita substituição da expectativa do valor futuro pela fórmula calculada no passo anterior.

Já é possível deduzir a fórmula geral, referente à j passos adiante:

$$E[y_{t+j}|I_t] = c \sum_{s=0}^{j-1} \phi^s + \phi^j y_t = c/(1 - \phi) + \phi^j y_t.$$

Esta é, pois, a fórmula do previsor ótimo para uma série mais bem reproduzida por um modelo AR(1).

1.2 Caso do processo estocástico MA(1)

Cabe agora considerar o caso de uma série mais bem reproduzida por um modelo MA(1). Iniciando pela previsão um passo adiante:

$$E[y_{t+1}|I_t] = E[c + \theta e_t + e_{t+1}|I_t] = E[c + \theta L e_t + e_{t+1}|I_t] = c + E[e_{t+1}(1 + \theta L)|I_t] \rightarrow$$

$$E[y_{t+1} \cdot (1 + \theta L)^{-1}|I_t] = c + E[e_{t+1}|I_t] \rightarrow E[(y_{t+1} - c)(1 + \theta L)^{-1}|I_t] = E[e_{t+1}|I_t] \rightarrow$$

$$E[e_{t+1}|I_t] = E[(y_{t+1} - c)(1 + \theta L)^{-1}|I_t] = \{1\} =$$

$$E[e_{t+1}|I_t] = E[(1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \dots)^{-1}(y_{t+1} - c) + \dots|I_t]$$

$$E[e_{t+1}|I_t] = E[(y_{t+1} - c) - \theta(y_t - c) + \theta^2(y_{t-1} - c) - \theta^3(y_{t-2} - c) + \dots|I_t]$$

$$0 = E[(y_{t+1}-c) - \theta (y_t-c) + \theta^2(y_{t-1}-c) - \theta^3(y_{t-2}-c) + \dots | I_t]$$

$$0 = E[y_{t+1}-c | I_t] - \theta (y_t-c) + \theta^2(y_{t-1}-c) - \theta^3(y_{t-2}-c) + \dots$$

$$E[y_{t+1}-c | I_t] = \theta (y_t - c) - \theta^2(y_{t-1}-c) + \theta^3(y_{t-2}-c) + \dots$$

$$E[y_{t+1} | I_t] = c + \theta (y_t - c) - \theta^2(y_{t-1}-c) + \theta^3(y_{t-2}-c) + \dots$$

Agora considerando a previsão dois passos adiante:

$$E[y_{t+2}|I_t] = E[c + \theta e_{t+1} + e_{t+2}|I_t] \rightarrow$$

$$E[y_{t+2}|I_t] = E[c + \theta L e_{t+2} + e_{t+2}|I_t] = E[c + (1+\theta L)e_{t+2}|I_t] \rightarrow$$

$$E[(y_{t+2}-c)(1+\theta L)^{-1}|I_t] = E[e_{t+2}|I_t] \rightarrow$$

$$E[(y_{t+2}-c)(1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \theta^4 L^4 - \dots) | I_t] = 0 \rightarrow$$

$$E[(y_{t+2}-c) | I_t] - \theta E[(y_{t+1}-c)|I_t] + \theta^2 E[(y_t-c) | I_t] - \theta^3 E[(y_{t-1}-c) | I_t] + \theta^4 E[(y_{t-2}-c) | I_t] - \dots = 0 \rightarrow$$

$$E[(y_{t+2}-c) | I_t] = \theta E[(y_{t+1}-c)|I_t] - \theta^2 E[(y_t-c) | I_t] + \theta^3 E[(y_{t-1}-c) | I_t] - \theta^4 E[(y_{t-2}-c) | I_t] + \dots = 0 \rightarrow$$

$$E[(y_{t+2}-c) | I_t] = \theta E[y_{t+1}|I_t] - \theta c - \theta^2(y_t-c) + \theta^3(y_{t-1}-c) - \theta^4(y_{t-2}-c) + \dots = 0 \{1\} \rightarrow$$

$$E[(y_{t+2}-c) | I_t] = \theta \{c + \theta (y_t - c) - \theta^2(y_{t-1}-c) + \theta^3(y_{t-2}-c) + \dots\} - \theta c - \theta^2(y_t-c) + \theta^3(y_{t-1}-c) - \theta^4(y_{t-2}-c) + \dots = 0 \rightarrow$$

$$E[(y_{t+2}-c) | I_t] = \theta^2(y_t - c) - \theta^3(y_{t-1}-c) + \theta^4(y_{t-2}-c) + \dots - \theta^2(y_t-c) + \theta^3(y_{t-1}-c) - \theta^4(y_{t-2}-c) + \dots = 0 \rightarrow$$

$$E[(y_{t+2}-c) | I_t] = 0 \rightarrow$$

$$E[y_{t+2}|I_t] = c \rightarrow$$

O que sugere, por indução, que a fórmula para a previsão j passos adiante é:

$$E[y_{t+j}|I_t] = c + \theta (y_t - c) - \theta^2(y_{t-1}-c) + \theta^3(y_{t-2}-c) + \dots, \text{ se } j = 1$$

$$E[y_{t+j}|I_t] = c, \text{ se } j > 1$$

Tem-se aqui um resultado fundamental: a previsão para um número de passos superior à ordem do modelo MA(1) é igual à média incondicional da série observada. De fato, se $y_t \sim \text{MA}(1)$, então $E[y_t] = E[c + \theta e_{t-1} + e_t] = c$. Há, pois, uma deterioração rápida das previsões geradas pelo MA(1), as quais se tornam constantes logo após um instante de tempo adiante.

1.3 Caso geral

Conforme apresentado, a fórmula para cálculo de previsões depende do processo estocástico utilizado para reproduzir a série de tempo empírica. Mas o conhecimento

disponível é tal que todas as fórmulas já foram derivadas e são, portanto, conhecidas (elas se encontram no capítulo 4 de Hamilton, 1994). Há duas características fundamentais a assinalar quanto às fórmulas associadas a todas as especificações ARMA possíveis, conforme estabelece Enders (2015):

1. Elas convergem para a média incondicional com j tendendo ao infinito;
2. A variância do erro de previsão, $u_t = y_{t+j} - E[y_{t+j}|I_t]$, aumenta com j , e, no limite, com $j \rightarrow \infty$, converge para a variância incondicional da série original (y_t);

Um exemplo em que a última propriedade é particularmente clara é o do processo AR(1) tratado duas subseções acima:

$$E[y_{t+j}|I_t] = c/(1-\phi) + \phi^j y_t.$$

No limite, com $j \rightarrow \infty$, $E[y_{t+j}|I_t] = c/(1-\phi)$, pois $|\phi| < 1$, uma vez que a condição de estacionaridade tem de ser atendida por um processo AR(1). E, conforme já visto na nota de aula anterior, se $y_t \sim \text{AR}(1)$, $E[y_t] = c/(1-\phi)$.

Para o caso do modelo MA(1) não há propriamente convergência, mas sim uma mudança brusca para uma previsão constante no tempo para um número de passos a frente superior à ordem do modelo.

O fundamental a notar é: as previsões geradas por processos ARMA deterioram-se progressivamente com o número de passos adiante para os quais se quer prever a série. Como corolário, previsões referentes a um ou poucos passos adiante são mais confiáveis.

1.4 Erro de previsão

O erro de previsão é simplesmente a diferença entre valor observado e valor previsto, i.e.:

$$u_{t+j} = y_{t+j} - \hat{y}_{t+j} = y_{t+j} - E[y_{t+j}|I_t]$$

A variância do erro de previsão, por sua vez, é igual a:

$$\begin{aligned} V(u_{t+j}|I_t) &= E[(u_{t+j} - E[u_{t+j}|I_{t+j}])^2] = \{1\} = E[u_{t+j}^2] = E[(y_{t+j} - \hat{y}_{t+j})^2] \\ &= \{2\} = E[(E[\hat{y}_{t+j}|I_{t+j}] - \hat{y}_{t+j})^2] = E[(\hat{y}_{t+j} - E[\hat{y}_{t+j}|I_{t+j}])^2] \\ &= V(\hat{y}_{t+j}|I_t) \end{aligned}$$

Em que foram utilizados os resultados: {1} de nulidade da expectativa condicional do erro de previsão (o previsor ótimo é não-viesado), {2} de ausência de viés do previsor ótimo, i.e., $E[\hat{y}_{t+j}|I_{t+j}] = y_{t+j}$.

Em suma:

$$V(u_{t+j}|I_t) = V(\hat{y}_{t+j}|I_t)$$

O resultado é: a variância do erro de previsão é igual à variância da previsão.

O resultado anterior pode ser combinado com um segundo resultado para gerar uma conclusão bastante importante sobre como a dispersão das previsões varia com o horizonte de previsão. Este segundo resultado será estabelecido também a partir do processo AR(1) e diz respeito à variância, ou volatilidade, do erro de previsão. Basta calculá-la para ver:

$V[u_{t+j}|I_t] = E[(y_{t+j} - E[y_{t+j}|I_t])^2|I_t] = E[(y_{t+j} - c/(1-\phi) - \phi^j y_t)^2|I_t] = g(j)$, em que $g(\cdot)$ é uma função genérica.

Com base na passagem anterior, é possível apresentar um argumento simples porém não rigoroso, para a relação entre tamanho do horizonte de previsão e volatilidade do erro de previsão. Assumindo que a série não aumenta ou diminui sistematicamente com j , mas oscila, exibindo momentos de aumento e de redução, a função $g(\cdot)$ é crescente em “ j ”. Isso pois $|\phi| < 1$, e, portanto, quanto maior j , menor é a parcela que explicitamente contém j e é subtraída das demais parcelas e, portanto, maior é o resultado. Ou seja, a volatilidade do erro de previsão, esta uma medida para a incerteza associada à previsão, aumenta com a extensão do erro de previsão. Esta conclusão é verdadeira e uma demonstração formal, para o caso do processo AR(1) é provida no apêndice ao final desta nota de aula.

Resumindo os parágrafos anteriores, tem-se:

- A variância da previsão é equivalente à variância do erro de previsão;
- A variância do erro de previsão é crescente com o horizonte de previsão.

Conclui-se, pois, que a variância da previsão é crescente com o horizonte de previsão. Ou seja, o grau de incerteza inerente às previsões é tão maior quanto maior o horizonte de previsão.

Para finalizar, considera-se o intervalo de confiança para a previsão. Este pode ser construído a partir da variância do erro de previsão (conforme já discutido). Para isso, são usados três ingredientes, (i) a fórmula do previsor, (ii) a fórmula da variância do erro de previsão, (iii) limites do intervalo de confiança de 95% iguais a 1,96, assumindo-se distribuição assintoticamente normal. Com o que se obtém:

$$\hat{y}_{t+j} \pm 1.96 \sqrt{V(u_{t+j}|I_t)}$$

Pode-se assim saber qual é o intervalo que contém, com 95% de probabilidade, o real valor da série j instantes de tempo adiante. O que é mais informativo do que a previsão pontual.

Cabe ressaltar uma característica do intervalo de confiança desnudada pela fórmula acima: quanto maior o horizonte de previsão, maior a variância do erro de previsão, e consequentemente, maior a amplitude do intervalo de confiança das previsões. Algo

coerente, pois é comum tomar a amplitude de um intervalo de confiança como medida do grau de incerteza enfrentado.

1.5 Previsão dentro e fora da amostra

Há dois usos para as previsões. A primeira é óbvia. Trata-se de conhecer valores futuros. A segunda é a comparação de especificações ARMA alternativas, em termos do desempenho preditivo. Ou seja, trata-se de saber quais de duas especificações possíveis cometem menor erro de previsão. Nesta sub-seção foca-se neste segundo uso. Ele pode ser feito a partir de dois tipos de previsão. A primeira e mais intuitiva é a previsão de níveis futuros da série. O que é chamado de previsão “fora da amostra” (out-of-sample prediction). Já a segunda é a previsão de níveis presentes ou passados, ou seja, já observados, e que por isso compõem a amostra. Trata-se de previsão “dentro da amostra” (in-sample prediction).

Para gerar previsões dentro da amostra, subdivide-se a última em duas subamostras, (i) a subamostra utilizada para estimar as especificações ARMA, ou “amostra de estimação”, com tamanho denotado por T_1 , e (ii) a subamostra com base na qual o desempenho de previsão é avaliado, ou “amostra para avaliação”, cujo tamanho é indicado por T_2 . Tem-se $T_1 + T_2 = T$.

1.6 Previsão dinâmica e estática

1.6.1 Base conceitual

As previsões podem ser geradas de duas maneiras que diferem em função da natureza da informação que incorporam. É esclarecedor tomar como exemplo a previsão dois passos à frente (dentro ou fora da amostra), $\hat{y}_{t+2}$. Esta geralmente é função da previsão um passo à frente, $\hat{y}_{t+2} = f(\hat{y}_{t+1})$ – veja a seção 1.1 para o caso do processo AR(1). Eis a diferença: pode-se utilizar, em tal função, (i) ou a previsão um passo a frente, e neste caso tem-se a assim chamada “previsão dinâmica” ($\hat{y}_{t+2} = f(\hat{y}_{t+1})$), ou (ii) o nível efetivamente observado da série, caracterizando a “previsão estática” ($\hat{y}_{t+2} = f(y_{t+1})$).

Colocando em outras palavras:

- A previsão dinâmica utiliza a previsão para um dado instante para prever o instante seguinte;
- Já a previsão estática utiliza o valor observado em um dado instante para prever o instante seguinte;
- A previsão dinâmica prevê múltiplos instantes à frente, usando em cada um deles previsões para instantes anteriores;
- Já a previsão estática prevê apenas um instante à frente, usando a observação do instante corrente.

A previsão estática, por ter por base um maior número de observações, tem maior confiabilidade, ou seja, menor probabilidade de errar. E também pois seu

horizonte de previsão nunca é maior do que um passo adiante - o que reduz a convergência para a média incondicional e o aumento da incerteza.

A queda da qualidade da previsão dinâmica é perceptível de uma perspectiva geral e embasada nas fórmulas dos processos ARMA. O caso AR(1) é bastante ilustrativo, conforme aponta Enders (2015, p.79):

A previsão para h instante adiante é: $E[y_{t+h}|I_t] = \phi E[y_{t+h-1} + e_{t+h}|I_t] = \phi E[y_{t+h-1}|I_t]$. Essa mesma fórmula pode ser usada para substituir a previsão para $t+h-1$, ou seja para $h-1$ instantes à frente, pela previsão para $h-2$ instantes à frente. O que se tem é:

- $E[y_{t+h}|I_t] = \phi E[y_{t+h-1}|I_t]$ (a previsão h instantes adiante incorpora a previsão $h-1$ instantes adiante)
- $E[y_{t+h-1}|I_t] = \phi E[y_{t+h-2}|I_t]$ (a previsão h instantes adiante incorpora a previsão $h-2$ instantes adiante)
- $E[y_{t+h-2}|I_t] = \phi E[y_{t+h-3}|I_t]$ (a previsão h instantes adiante incorpora a previsão $h-3$ instantes adiante)
- ...
- $E[y_{t+2}|I_t] = \phi E[y_{t+1}|I_t] + e_{t+2}$ (a previsão h instantes adiante incorpora a previsão $h-3$ instantes adiante)

Ou seja, quanto mais adiante se prevê, mais previsões têm de ser incorporadas, e como cada previsão é sujeita a erro, e há acúmulo de erros de previsão, maior é o erro total acumulado – ver o apêndice abaixo, em que é estabelecido que o erro de previsão j passos adiante para um processo AR(1) é um acúmulo de perturbações, e, pois, de erros de estimação.

Já a previsão estática reduz o acúmulo de erros de previsão ao substituir uma previsão, feita antes da realização de um novo valor para a série, pelo valor observado após a realização. Ou seja, quando $t+1$ passa a ser o dia de hoje, substitui-se $E[y_{t+1}|I_t]$ por $E[y_{t+1}|I_{t+1}] = y_{t+1}$, reduzindo o erro de previsão acumulado em magnitude equivalente ao erro de previsão referente a $t+1$.

1.6.2 Computação

As abordagens de previsão estática e dinâmica estão ambas disponíveis nos softwares econométricos². Os livros texto de Morettin e Tolo (2005, seção 9.3) e Hamilton (1994, seção 4.5) referem-se à previsão dinâmica como o acréscimo, à previsão referente ao instante $t+j$, de um múltiplo do erro de previsão calculado com base em uma nova realização da série, esta referente ao instante $t+1$. Contudo, como Morettin e Tolo (2005) deixam claro, os parâmetros do processo ARMA têm de ser recalculados. Tendo isso em vista e também a afirmação de Bueno (2008, p. 83), a maneira recomendada de fazer a previsão estática é reestimando o processo ARMA a cada nova realização da série e então fazendo a previsão um passo a frente. É isso que será, nesta disciplina,

² Por exemplo, no caso do software Stata®, o leitor interessado pode consultar as páginas do manual do software no seguinte link: <https://www.stata.com/manuals13/tsarimapostestimation.pdf>.

entendido como previsão estática. Já a previsão dinâmica é simplesmente a previsão para mais de um passo adiante sem reestimação do processo ARMA.

2 Algoritmo alternativo

Cabe recordar que a especificação ARIMA selecionada pela metodologia de Box-Jenkins é a que melhor reproduz a série de tempo empírica sob análise. Porém, há um fato fundamental: conforme afirmam Enders (2015, p.82) e Box e Jenkins (20XX, p.), não devemos nos enganar ao pensar que tal processo é sempre aquele que melhor prevê valores futuros da série. O fundamento da diferença entre (i) melhor reproduzir uma série e (ii) melhor prever níveis futuros da série está na natureza distintas dos dados em questão. A reprodução de uma série tem por base dados já observados, enquanto que a previsão se refere a dados futuros, ainda não observados. Se há uma diferença significativa entre os níveis da série já observados e os futuros, então a especificação selecionada pode prever relativamente mal a série.

Caso, sublinhe-se, o objetivo da análise seja a previsão, o esclarecimento do parágrafo anterior sugere expandir o primeiro algoritmo da metodologia Box-Jenkins. Ao invés de classificar especificações candidatas apenas com base em critérios de liberdade, também pode-se considerar o erro agregado de previsão. Ou seja, selecionaria-se a especificação com melhor desempenho preditivo e com graus de liberdade se não mínimos, ao menos não máximos.

Antes de apresentar a segunda versão do algoritmo, a qual prioriza a previsão, cabe apresentar algumas formas de agregar o erro de previsão. Todas elas evitam o problema de cancelamento de valores iguais em absoluto, mas com sinais diferentes. Seja ao elevar ao quadrado os resíduos ou ao tomar o valor absoluto dos mesmos. Uma delas toma a razão entre erro de previsão absoluto e a variável dependente, assim informando quanto à porcentagem que o erro de previsão representa. As três agregações para o erro de previsão consideradas nesta disciplina, em consonância com Bueno (2008, p.82), são:

Raiz do erro quadrático médio = root mean square error =

$$MSE = \frac{1}{T_2} \sqrt{\sum_{h=1}^{T_2} \hat{u}_{t+h}^2}$$

$$\text{Erro absoluto médio} = \text{mean absolute error} = MAE = \frac{1}{T_2} \sum_{h=1}^{T_2} |\hat{u}_{t+h}|$$

Erro absoluto percentual médio = mean absolute percentual error =

$$MAPE = \frac{1}{T_2} \sum_{h=1}^{T_2} \frac{|\hat{u}_{t+h}|}{\hat{y}_{t+h}}$$

Esclarece-se que as três medidas são calculadas a partir de previsões dentro da amostra, para uma fração da amostra de tamanho T_2 .

Apêndice Relação entre horizonte de previsão e volatilidade do erro de previsão: caso do processo estocástico AR(1)

(este apêndice é uma explicação mais detalhada da demonstração feita por Enders (2015, p.80, equações 2.49 e 2.50))

A variância do erro da previsão j passos adiante é dada por $V[u_{t+j}|I_t] = E[(y_{t+j} - E[y_{t+j}|I_t])^2|I_t]$. Seja considerado o caso em que se deseja reproduzir a série empírica y_t , $t = 1, 2, \dots$, com base em um processo AR(1). Neste caso, é adequado afirmar que a série empírica real é equivalente ao processo AR(1), aumentada, porém, por um erro de estimação. Ocorre que, por simplicidade, pode-se assumir que o termo de perturbação do processo AR(1) já contém este erro. Com isso pode-se escrever a série original como um processo AR(1): $y_{t+j} = c + \phi y_{t+j-1} + e_{t+j}$. Com isso tem-se:

Considerando que a série temporal empírica

$$V[u_{t+j}|I_t] = E[(\{c + \phi y_{t+j-1} + e_{t+j}\} - E[y_{t+j}|I_t])^2|I_t].$$

Utilizando raciocínio análogo, pode-se substituir y_{t+j} pela especificação AR(1) também dentro da expectativa $E[y_{t+j}|I_t]$:

$$V[u_{t+j}|I_t] = E[(\{c + \phi y_{t+j-1} + e_{t+j}\} - E[c + \phi y_{t+j-1} + e_{t+j}|I_t])^2|I_t] \leftrightarrow$$

$$V[u_{t+j}|I_t] = E[(\phi y_{t+j-1} + e_{t+j} - \phi E[y_{t+j-1}|I_t])^2|I_t] \leftrightarrow$$

$$V[u_{t+j}|I_t] = E[(e_{t+j} + \phi(y_{t+j-1} - E[y_{t+j-1}|I_t]))^2|I_t] \leftrightarrow$$

$$V[u_{t+j}|I_t] = E[(e_{t+j} + \phi u_{t+j-1})^2|I_t] \leftrightarrow$$

Voltando mais um passo:

$$V[u_{t+j}|I_t] = E[(e_{t+j} + \phi(y_{t+j-1} - E[y_{t+j-1}|I_t]))^2|I_t]$$

$$V[u_{t+j}|I_t] = E[(e_{t+j} + \phi(\phi y_{t+j-2} + e_{t+j-1} - c - E[y_{t+j-2}|I_t]))^2|I_t]$$

$$V[u_{t+j}|I_t] = E[(e_{t+j} + \phi(\phi(y_{t+j-2} - E[y_{t+j-2}|I_t]) + e_{t+j-1}))^2|I_t]$$

$$V[u_{t+j}|I_t] = E[(e_{t+j} + \phi^2(y_{t+j-2} - E[y_{t+j-2}|I_t]) + \phi e_{t+j-1})^2|I_t]$$

$$V[u_{t+j}|I_t] = E[(e_{t+j} + \phi e_{t+j-1} + \phi^2(u_{t+j-2}))^2|I_t]$$

Se continuarmos a iteração recursiva, chegaremos a:

$V[u_{t+j}|I_t] = E[(e_{t+j} + \phi e_{t+j-1} + \phi^2 e_{t+j-2} + \dots + \phi^j e_t)^2 | I_t]$. Convertendo o quadrado interno em um produto, chega-se a:

$$\begin{aligned}
 V[u_{t+j}|I_t] &= E \left[\sum_{k=0}^j \phi^{2k} e_{t+j-k}^2 + \sum_{k=0}^j \sum_{\substack{m=0 \\ m \neq k}}^j \phi^{k+m} e_{t+j-k} e_{t+j-m} \middle| I_t \right] \leftrightarrow \\
 V[u_{t+j}|I_t] &= \sum_{k=0}^j \phi^{2k} E[e_{t+j-k}^2 | I_t] + \sum_{k=0}^j \sum_{\substack{m=0 \\ m \neq k}}^j \phi^{k+m} E[e_{t+j-k} e_{t+j-m} | I_t] \leftrightarrow \{1\} \\
 V[u_{t+j}|I_t] &= \sum_{k=0}^j \phi^{2k} \sigma^2 \leftrightarrow V[u_{t+j}|I_t] = \sigma^2 \sum_{k=0}^j \phi^{2k} \leftrightarrow \{2\} \\
 V[u_{t+j}|I_t] &= \sigma^2 \frac{1 - \phi^{2kj}}{1 - \phi^{2k}} = g(j)
 \end{aligned}$$

Uma vez que $|\phi| < 1$, a função $g(\cdot)$ é crescente em j , pois quanto maior j , menor ϕ^{2kj} e, portanto, maior o numerador de $g(\cdot)$. Quanto às passagens $\{1\}$ e $\{2\}$, a primeira utilizou os fatos de que uma perturbação ruído branco tem variância constante σ^2 e autocorrelações nulas. A passagem $\{2\}$ utilizou a fórmula da soma de uma progressão geométrica finita $(a_1(1-q^N)/(1-q))$, $a_1 = 1$, $q = \phi^{2k}$ e $N = j$.