

PREVISÕES BAYESIANAS E MODELOS GRÁFICOS

Ralph dos Santos Silva

<http://www.ralphsilva.org/previsoesbayesianasemodelosgraficos.html>

Instituto de Matemática
Centro de Ciências Matemáticas e da Natureza
Universidade Federal do Rio de Janeiro

Séries temporais

Uma **Série Temporal** é um conjunto de observações y_t , cada uma observada no tempo específico t . Notação: $\{y_t\}$.

Uma série temporal é tempo-discreta se o conjunto de índices de tempo for um conjunto discreto, por exemplo $T = \{1, 2, 3, \dots\}$.

Uma série temporal é tempo-contínua se as observações ocorrem continuamente em um intervalo de tempo, por exemplo $T \in [0, 1]$.

Neste curso trataremos de séries temporais a tempo-discreto.

A seguir, veremos

- **Estacionariedade;**
- **Função de Autocorrelação;** e
- **Autoregressivos e Médias Móveis (ARMA).**

Estacionariedade

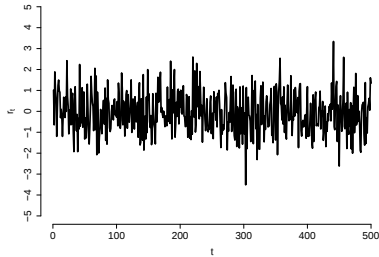
Uma série temporal $\{y_t\}$ é **(fracamente) estacionária** se a média e a variância de Y_t forem invariantes no tempo e a covariância de Y_t com Y_{t+h} depender somente de h .

Na estacionariedade fraca temos que

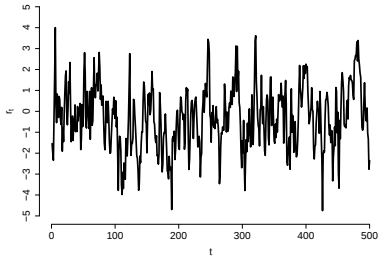
$$E(Y_t) = \mu, \quad \text{Var}(Y_t) = \sigma^2 \quad \text{e} \quad \text{Cov}(Y_t, Y_{t+h}) = \gamma(h), \quad \text{para } h \geq 0.$$

Sob a hipótese de estacionariedade fraca, esperamos que uma série temporal (observada) esteja em torno de um valor constante e com uma variação constante.

Exemplo



(a) Modelo Ruído Branco com $\sigma = 1$.



(b) Modelo AR(1) com $\phi = 0,7$ e $\sigma = 1$.

Figura 1: Exemplos de séries temporais fracamente estacionárias.

Exemplo

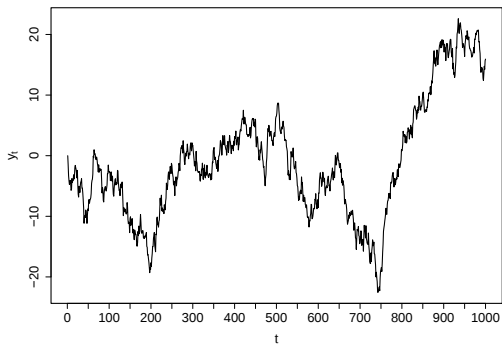


Figura 2: Exemplo de série temporal não estacionária (passeio aleatório).

Função de autocovariância

A função de autocovariância de defasagem h é definida por

$$\gamma(h) = \text{Cov}(Y_t, Y_{t+h}).$$

Temos que $\gamma(0) = \text{Var}(Y_t)$.

Além disso, sob estacionariedade fraca, temos que

$$\gamma(h) = \text{Cov}(Y_t, Y_{t+h}) = \text{Cov}(Y_t, Y_{t-h}) = \gamma(-h),$$

isto é,

$$\gamma(h) = \gamma(-h).$$

Função de autocorrelação

Considere uma série Y_t fracamente estacionária.

Estamos interessados na correlação linear entre Y_t e seus valores passados Y_{t-h} (ou futuros Y_{t+h}).

O coeficiente de correlação entre Y_t e Y_{t-h} é chamado de autocorrelação de defasagem (lag) h de Y_t . Definimos por

$$\rho(h) = \frac{\text{Cov}(Y_t, Y_{t-h})}{\sqrt{\text{Var}(Y_t)\text{Var}(Y_{t-h})}} = \frac{\text{Cov}(Y_t, Y_{t-h})}{\text{Var}(Y_t)} = \frac{\gamma(h)}{\gamma(0)},$$

pois $\text{Var}(Y_t) = \text{Var}(Y_{t-h})$ pela hipótese de estacionariedade fraca.

Temos que

- $\rho(0) = 1$;
- $\rho(h) = \rho(-h)$ para todo h ; e
- $-1 \leq \rho(h) \leq 1$ para todo h .

Uma série fracamente estacionária Y_t não é correlacionada serialmente se, e somente se, $\rho(h) = 0$ para todo $h > 0$.

Estimação da função de autocorrelação

Para uma dada amostra $\{y_t\}_{t=1}^n$, a autocorrelação (amostral) de defasagem 1 de r_t é

$$\hat{\rho}(1) = \frac{\sum_{t=2}^n (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad \text{em que} \quad \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t.$$

Sob algumas condições gerais, $\hat{\rho}(1)$ é um estimador consistente de $\rho(1)$.

Se $\{Y_t\}$ for uma sequência independente e identicamente distribuída (iid) e $E(Y_t^2) < \infty$, então $\hat{\rho}(1)$ é assintoticamente normal com média 0 e variância $1/n$:

$$\hat{\rho}(1)\sqrt{n} \approx \mathcal{N}(0, 1).$$

Podemos testar

$$\begin{cases} H_0 : & \rho(1) = 0 \\ H_1 : & \rho(1) \neq 0. \end{cases}$$

Estimação da função de autocorrelação: defasagem $h \geq 0$

A autocorrelação (amostral) de defasagem h de y_t é definida por

$$\hat{\rho}(h) = \frac{\sum_{t=h+1}^n (y_t - \bar{y})(y_{t-h} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad \text{para } 0 \leq h \leq n-1.$$

Se $\{Y_t\}$ for uma sequência iid com $E(Y_t^2) < \infty$, então $\hat{\rho}(h)$ é assintoticamente normal com média 0 e variância $1/n$ para todo inteiro positivo e fixo h :

$$\hat{\rho}(h)\sqrt{n} \approx \mathcal{N}(0, 1).$$

Para h fixo, podemos testar

$$\begin{cases} H_0 : & \rho(h) = 0 \\ H_1 : & \rho(h) \neq 0. \end{cases}$$

Ver arquivo exemplo_1.r

Modelo ruído branco

Uma série temporal Y_t é chamada de **ruído branco** se $\{Y_t\}$ for uma sequência de variáveis aleatórias não correlacionadas e identicamente distribuídas com média zero e variância constante e finita (σ^2).

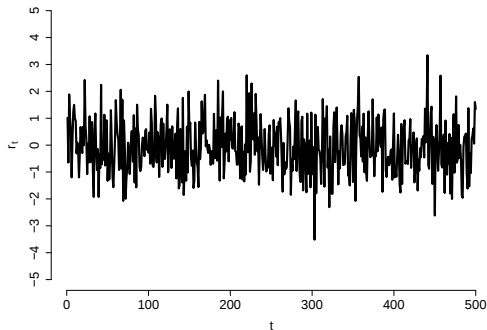


Figura 3: Exemplo de uma série temporal **ruído branco** com $\sigma = 1$.

Ver arquivo exemplo_2.r

Modelo passeio aleatório

Soma acumulada de variáveis aleatórias ε_t não correlacionadas com $E(\varepsilon_t) = 0$ e $\text{Var}(\varepsilon_t) = \sigma^2 < \infty$.

$$Y_0 = \mu \quad \text{e} \quad Y_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t = Y_{t-1} + \varepsilon_t, \quad \text{para } t = 1, 2, \dots$$

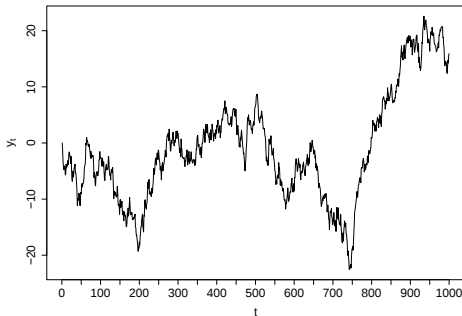


Figura 4: Exemplo de um passeio aleatório gaussiano com $\mu = 0$ e $\sigma = 1$.

Ver arquivo exemplo_2.r

Modelo autoregressivo de ordem 1

Um modelo autoregressivo de ordem 1, AR(1), é dado por

$$Y_t = \mu + \phi(Y_{t-1} - \mu) + \varepsilon_t = \alpha + \phi Y_{t-1} + \varepsilon_t, \quad \text{com } \alpha = \mu(1 - \phi),$$

em que ε_t é um ruído branco com média zero e variância σ_a^2 .

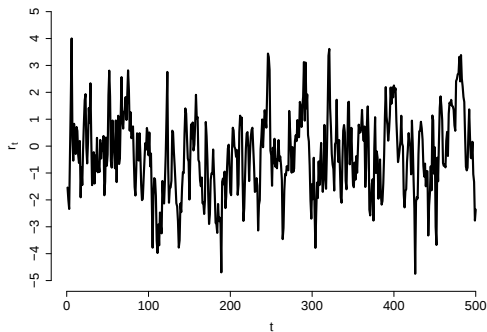
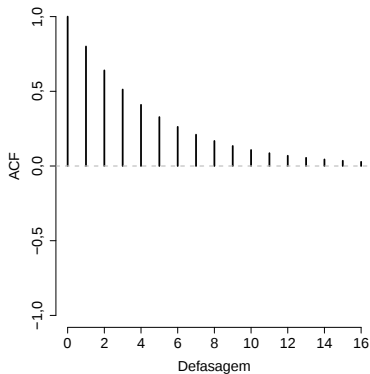
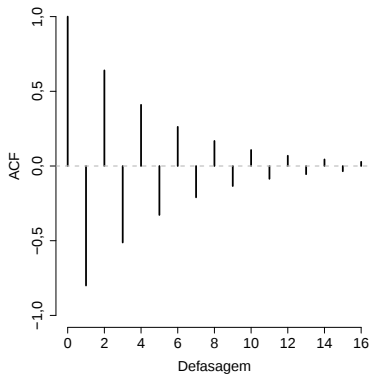


Figura 5: Exemplo de uma série temporal **AR(1)** com $\phi = 0,7$ e $\sigma = 1$.



(a) $\phi = 0,8$.



(b) $\phi = -0,8$

Figura 6: Função de autocorrelação do modelo AR(1).

Ver arquivo exemplo_2.r

Modelos de Markov de primeira ordem

Defina $\mathbf{Y}_{1:t} = (Y_1, Y_2, \dots, Y_t)^\top$ como uma coleção de variáveis aleatórias.



Figura 7: Um modelo gráfico direcionado de uma cadeia de Markov de primeira ordem.

Note no DAG acima que $Y_{t+1} \perp \mathbf{Y}_{1:t-1} | Y_t$. Além disso,

$$f(\mathbf{y}_{1:T} | \boldsymbol{\psi}) = f(y_1 | \boldsymbol{\psi}) \prod_{t=2}^T f(y_t | y_{t-1}, \boldsymbol{\psi}), \quad \text{com } \mathbf{y}_{1:t} = (y_1, y_2, \dots, y_t)^\top.$$

Exemplo

Considere o modelo dado por

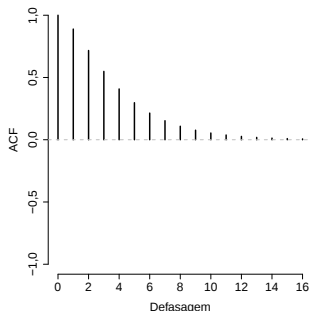
$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad \text{tal que } E(\varepsilon_t) = 0 \quad \text{e} \quad \text{Var}(\varepsilon_t) = \sigma^2.$$

- Para $|\phi| < 1$, temos o modelo **autoregressivo de ordem 1**; e
- Para $\phi = 1$, temos o modelo **passeio aleatório**.

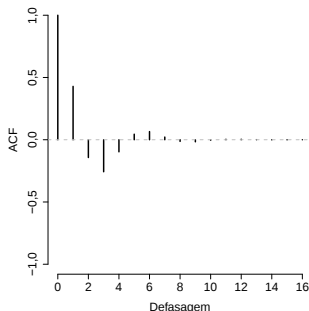
Modelo autoregressivo de ordem 2

Um modelo AR(2) tem a forma

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \varepsilon_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t.$$



(a) $\phi_1 = 1,2$ e $\phi_2 = -0,35$



(b) $\phi_1 = 0,6$ e $\phi_2 = -0,4$

Figura 8: Função de autocorrelação do modelo AR(2).

Ver arquivo exemplo_2.r

Modelos de Markov de segunda ordem

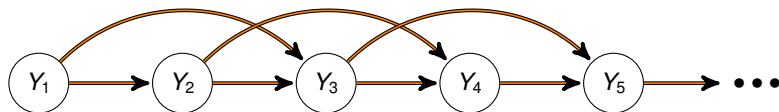


Figura 9: Um modelo gráfico de uma cadeia de Markov de segunda ordem.

Note no DAG acima que $Y_{t+1} \perp Y_{1:t-2} | (Y_t, Y_{t-1})$. Além disso,

$$f(\mathbf{y}_{1:T} | \boldsymbol{\psi}) = f(y_1, y_2 | \boldsymbol{\psi}) \prod_{t=3}^T f(y_t | y_{t-1}, y_{t-2}, \boldsymbol{\psi}).$$

Exemplo

Considere o modelo autoregressivo de ordem 2 dado por

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, \quad \text{tal que} \quad E(\varepsilon_t) = 0 \quad \text{e} \quad \text{Var}(\varepsilon_t) = \sigma^2,$$

com possíveis restrições impostas sobre ϕ_1 e ϕ_2 .

Modelos médias móveis

O modelo de **médias móveis de ordem 1**, ou simplesmente **MA(1)**, é dado por

$$Y_t = \mu + \theta \varepsilon_{t-1} + \varepsilon_t.$$

O modelo **MA(q)** é dado por

$$Y_t = \mu + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

Modelos de médias móveis são sempre estacionários porque são combinações lineares de uma sequência de ruídos brancos.

A autocorrelação na defasagem h de um modelo MA(q) é sempre igual a zero para $h > q$, isto é, $\rho(h) = 0$ se $h > q$.

Ver arquivo exemplo_2.r

Modelos ARMA

Um modelo **autoregressivo e médias móveis (ARMA)** pode ser melhor que um AR simples ou um MA simples devido ao número reduzido de parâmetros necessários para se captar a estrutura de dependência dos dados.

O modelo **ARMA(p, q)** é dado por

$$\begin{aligned} Y_t &= \mu + \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} + \varepsilon_t \\ &= \alpha + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} + \varepsilon_t. \end{aligned}$$

A estimação do modelo ARMA(p, q) é feita, em geral, pelo método da máxima verossimilhança. Disponível no R através da função `arima`.

Análise de séries temporais

A verificação do gráfico da série temporal nos auxilia no seguinte:

- Permite verificar tendências;
- Ajuda a verificar a hipótese de estacionariedade;
- Ajuda a verificar a hipótese de passeio aleatório; e
- Ajuda, em alguns caso, a identificar possível sazonalidade.

O gráfico da função de autocorrelação amostral (ACF) nos auxilia no seguinte:

- Analisar o grau de dependência da série temporal;
- Se houver tendência na série, então haverá muita dependência;
- Se houver sazonalidade, então haverá padrão ciclico;
- Se não estacionário ou parecido com passeio aleatório, então haverá muita dependência;
- Se a série for não correlacionada (ruído branco), então todas as autocorrelações são próximas de zero.

Modelos em Espaço de Estados

Outra abordagem para modelagem de séries temporais.

As séries temporais não precisam ser estacionárias!

Modelos em espaço de estados

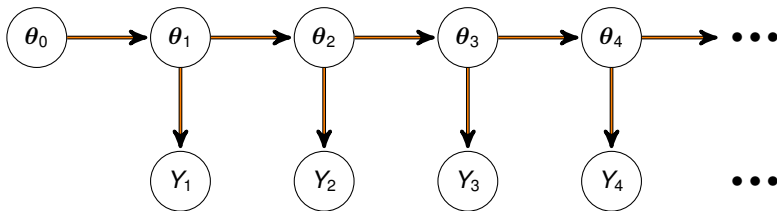


Figura 10: Um DAG de um modelo em espaço de estados.

Defina $\theta_{1:t} = (\theta_1, \theta_2, \dots, \theta_t)^\top$ como uma coleção de variáveis aleatórias.

Considera-se que

- $Y_{1:t}$ são variáveis observáveis; e
- $\theta_{1:t}$ são variáveis não observáveis (latentes).

Nestes modelos, temos as seguintes distribuições:

- $f(\theta_0 | \Psi)$, priori para o instante inicial;
- $f(\theta_t | \theta_{t-1}, \Psi)$, $t = 1, 2, \dots$, para a evolução dos estados; e
- $f(y_t | \theta_t, \Psi)$, $t = 1, 2, \dots$, para as observações.

A distribuição conjunta é dada por

$$f(\theta_0, \theta_{1:T}, \mathbf{y}_{1:T} | \Psi) = \left[\prod_{t=1}^T f(y_t | \theta_t, \Psi) \right] \left[\prod_{t=1}^T f(\theta_t | \theta_{t-1}, \Psi) \right] f(\theta_0 | \Psi)$$

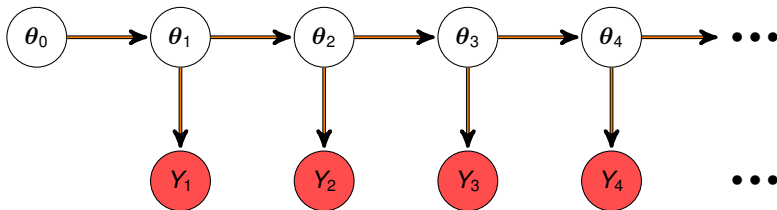


Figura 11: Um DAG de um modelo em espaço de estados.

Quando os vértices $\mathbf{Y}_{1:T}$ são observados, temos que

$$f(\theta_0, \theta_{1:T}, \mathbf{y}_{1:T} | \Psi) = \left[\prod_{t=1}^T f(y_t | \theta_t, \Psi) \right] \left[\prod_{t=1}^T f(\theta_t | \theta_{t-1}, \Psi) \right] f(\theta_0 | \Psi)$$

Modelos estáticos

Os **modelos estáticos** têm uma descrição fixa (através de parâmetros fixos) ao longo das unidades de observação.

- Análise de regressão,

$$Y_t = \beta_0 + \beta_1 z_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

- Modelos ARMA,

$$Y_t = \alpha + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

Em séries temporais, essa hipótese muitas vezes é violada:

as estruturas mudam com a passagem do tempo.

Nesta aula, faz-se uma introdução aos **modelos em espaço de estados**.

- **A descrição (os parâmetros) muda(m) com a passagem do tempo.**
- Os modelos em espaço de estados nos permitem a modelagem da tendência (média) e variância (volatilidade).
- Os modelos em espaço de estados incluem como casos particulares os modelos estáticos.

A passagem do tempo traz observações e aumenta o nosso conhecimento.

Em modelos em espaço de estados temos também perda de informação devido à passagem do tempo.

Por exemplo, o nível de vendas mês passado é mais relevante hoje que o nível de vendas em janeiro passado.

O modelo de regressão (linear múltiplo)

Uma variável resposta Y é função de um conjunto de variáveis explicativas $z_1, \dots, z_d$ através da relação

$$Y = \theta_0 + \theta_1 z_1 + \dots + \theta_d z_d + \varepsilon.$$

Em muitos casos, assume-se que $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

A equação acima pode ser compactamente escrita como:

$$Y = \mathbf{z}^\top \boldsymbol{\theta} + \varepsilon,$$

em que

$$\mathbf{z} = \begin{pmatrix} 1 \\ z_1 \\ \vdots \\ z_d \end{pmatrix} \quad \text{e} \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{pmatrix}.$$

A natureza das variáveis explicativas ou regressoras é bastante ampla.

Pode-se utilizar qualquer variável quantificável.

Os coeficientes de regressão $\theta_1, \dots, \theta_d$ informam sobre a influência que os regressores têm sobre a resposta Y .

Na prática, seus valores são desconhecidos e estimados a partir de uma coleção de observações feitas sobre o modelo.

Observamos respostas $y_1, \dots, y_T$ com seus respectivos regressores $\mathbf{z}_1, \dots, \mathbf{z}_T$. Simbolicamente, temos que

$$y_t = \mathbf{z}_t^\top \boldsymbol{\theta} + \varepsilon_t, \quad t = 1, \dots, T,$$

em que $\mathbf{z}_t = (z_{t1}, z_{t2}, \dots, z_{td})^\top$.

Modelo em espaço de estados

Em modelos em espaço de estados os parâmetros mudam com o passar do tempo.

O modelo de regressão é estendido para

$$Y_t = \mathbf{z}_t^\top \boldsymbol{\theta}_t + \varepsilon_t, \quad t = 1, \dots, T.$$

Note a **indexação de $\boldsymbol{\theta}$** .

Esta formulação cria uma profusão de parâmetros a serem estimados.

O modelo necessita de mais informação. Essa informação vem do fato que os parâmetros sucessivos estão intimamente relacionados.

Por exemplo, um parâmetro é igual ao seu antecessor mais uma pequena perturbação causada pelas mudanças às quais o sistema está sujeito.

Se o sistema é estático, como no modelo de regressão, temos que

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}.$$

Em modelos em espaço de estados, admitimos a forma mais geral dada por

$$\theta_t = \mathbf{G}_t \theta_{t-1} + \mathbf{w}_t,$$

em que

- $\mathbf{G}_t$ contém valores conhecidos (ou parâmetros fixos); e
- $\mathbf{w}_t$ é uma perturbação aleatória.

A equação acima é conhecida como **equação do sistema**.

A **matriz de evolução** $\mathbf{G}_t$ controla a parte determinística da evolução do sistema e estabelece a propagação do sistema ao longo do tempo.

A **perturbação** $\mathbf{w}_t$ é responsável pela introdução de incertezas devidas à passagem do tempo e consequente perda de informação.

Note que se $\mathbf{G}_t = \mathbf{I}$ e $\mathbf{w}_t = \mathbf{0}$, o modelo se reduz ao caso estático.

Exemplo

Suponha que uma série de **vendas** possa ser explicada pela respectiva série de **preços** através da seguinte relação:

$$vendas_t = \mu_t + \beta_t \times preço_t + v_t$$

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

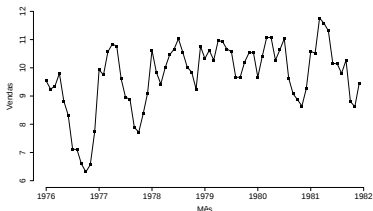
$$\beta_t = \beta_{t-1} + w_{2t}.$$

As **vendas** são explicadas pelo **preço** em uma regressão dinâmica.

A maior (ou menor) estabilidade dessa relação será controlada pela magnitude dos incrementos (erros) w_{1t} e w_{2t} .

Exemplo

Analisaremos a série de vendas mensais de um certo tipo doce no Reino Unido de janeiro de 1976 a dezembro de 1981. Também utilizaremos a série de preços, na escala do logaritmo, deste doce como variável explicativa.



(a) Vendas



(b) Preços

Figura 12: Vendas e logaritmo dos preços mensais de um certo tipo doce no Reino Unido de janeiro de 1976 a dezembro de 1981.

Ver arquivo exemplo_1.r