

Modelos para séries temporais sazonais  
aplicados a dados de venda do mercado de  
*travel retail*

Mariana Filgueiras Barrozo



Universidade Federal do Rio de Janeiro  
Instituto de Matemática  
Departamento de Métodos Estatísticos

2018

# Modelos para séries temporais sazonais aplicados a dados de venda do mercado de *travel retail*

Mariana Filgueiras Barrozo

Projeto Final de Curso submetido ao Corpo Docente do Departamento de Métodos Estatísticos do Instituto de Matemática da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para a obtenção dos títulos de Bacharel em Ciências Atuariais e Bacharel em Estatística.

Orientador: João Batista de Moraes Pereira

Rio de Janeiro, RJ - Brasil

2018

ii

# Modelos para séries temporais sazonais aplicados a dados de venda do mercado de *travel retail*

Mariana Filgueiras Barrozo

Projeto Final de Curso submetido ao Corpo Docente do Departamento de Métodos Estatísticos do Instituto de Matemática da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para a obtenção dos títulos de Bacharel em Ciências Atuariais e Bacharel em Estatística.

Aprovada por:

---

Professor João B. M. Pereira  
IM-UFRJ

---

Professora Kelly C. M. Gonçalves  
IM-UFRJ

---

Professor William L. Leão  
IM-UFRJ

Rio de Janeiro, RJ - Brasil

2018

# Agradecimentos

Agradeço ao meu orientador João Batista por toda dedicação, paciência e motivação. Aos professores que tive durante a graduação por todo o ensinamento. À minha família e amigos que sempre me incentivaram a continuar. À Iara Gil por toda paciência e apoio durante este processo. Ao pessoal da Dufry por todo o apoio. E à Fernanda Oliveira pelo incentivo da escolha do tema do projeto.

# Resumo

O comércio nesses últimos anos vem sendo muito afetado por causa da crise financeira no Brasil. Com isso, a necessidade de previsões de vendas faz-se cada vez mais importante, pois possibilita a tomada de decisões estratégicas. Decisões essas que podem estar relacionadas a compra de produtos para loja, criação de promoções, elaboração do orçamento para o ano seguinte, entre outras. Neste trabalho, modelamos de duas formas distintas as séries temporais da receita colocada e da receita resgatada de dados do mercado de *travel retail*. Especificamente, foram utilizados o modelo SARIMA e o modelo de regressão harmônica uma vez que os dados apresentam comportamento sazonal. A partir disso, analisamos e comparamos seus ajustes para as nossas séries de dados, assim como fazemos previsões das receitas para tempos futuros.

*Palavras-Chave:* Séries temporais, Modelo SARIMA, Modelo de regressão harmônica, Previsão, Travel retail.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Série temporal</b>	<b>4</b>
2.1	Função Média, Variância e Autocovariância . . . . .	5
2.2	Estacionariedade . . . . .	5
2.2.1	Ruído Branco . . . . .	6
2.3	Tendência . . . . .	6
2.4	Sazonalidade . . . . .	7
<b>3</b>	<b>Modelagem</b>	<b>9</b>
3.1	Modelos para séries estacionárias . . . . .	9
3.1.1	Modelo AR . . . . .	9
3.1.2	Modelo MA . . . . .	10
3.1.3	Modelo ARMA . . . . .	10
3.1.4	Modelo ARIMA . . . . .	11
3.2	Identificação do modelo . . . . .	11
3.2.1	ACF e PACF . . . . .	12
3.2.2	Estimação por Máxima Verossimilhança . . . . .	12
3.2.3	AIC e BIC . . . . .	13
3.2.4	Teste Ljung-Box . . . . .	14
3.2.5	Teste Shappiro-Wilk . . . . .	15
3.3	Modelos para séries temporais com comportamento sazonal . . . . .	15
3.3.1	Modelo de regressão harmônica . . . . .	15

3.3.2	Modelo SARIMA . . . . .	16
<b>4</b>	<b>Análise da receita colocada</b>	<b>17</b>
4.1	Modelo SARIMA aplicado à receita colocada . . . . .	20
4.1.1	Identificação do modelo . . . . .	21
4.1.2	Análise dos resíduos . . . . .	24
4.1.3	Previsão . . . . .	27
4.2	Modelo de regressão harmônica aplicado à receita colocada . . . . .	29
4.2.1	Identificação do modelo . . . . .	29
4.2.2	Análise dos resíduos . . . . .	29
4.2.3	Previsão . . . . .	36
<b>5</b>	<b>Análise da receita resgatada</b>	<b>37</b>
5.1	Modelo SARIMA aplicado à receita resgatada . . . . .	40
5.1.1	Identificação do modelo . . . . .	41
5.1.2	Análise dos resíduos . . . . .	44
5.1.3	Previsão . . . . .	47
5.2	Modelo de regressão harmônica aplicado à receita resgatada . . . . .	49
5.2.1	Identificação do modelo . . . . .	49
5.2.2	Análise dos resíduos . . . . .	49
5.2.3	Previsão . . . . .	55
<b>6</b>	<b>Conclusão</b>	<b>57</b>

# Capítulo 1

## Introdução

Entender o comportamento do consumidor e do mercado, assim como fazer previsões de variáveis relacionadas a este meio sempre foi essencial. Neste contexto, modelos e técnicas estatísticas são amplamente empregados e vêm ajudando cada vez mais as empresas a tomarem decisões estratégicas. Principalmente hoje em dia, em que o avanço tecnológico permite análises rápidas de grandes bancos de dados.

Muitos conjuntos de dados são observados no tempo. A estes chamamos de séries temporais. Dados deste tipo podem apresentar diversos tipos de comportamento no tempo: tendências, comportamentos sazonais ou cíclicos ou até mesmo comportamentos puramente aleatórios. Entender este comportamento possibilita a previsão. Este é um dos papéis da análise de séries temporais.

Neste trabalho, analisamos dados de uma empresa do ramo de *travel retail*. Entende-se como *travel retail* a empresa que tem lojas em aeroportos, onde as pessoas estão indo viajar. Lojas estas tanto na área nacional onde vende produtos com o preço compatível ao mercado local, como lojas na área internacional onde vende produtos sem a incidência dos impostos dos países locais. Neste caso iremos analisar as lojas que vendem os produtos sem o preço dos impostos no valor final do mesmo, o que é uma oportunidade bem atrativa para o passageiro.

Porém, assim como as empresas do ramo de comércio, as vendas foram prejudicadas pela crise financeira que o Brasil está passando. Uma vez que os produtos são vendidos em dólar, o preço dos mesmos, transformado em real, acaba sendo um pouco mais elevado do que seria em um país estável economicamente. Com isso, várias medidas estão sendo adotadas em empresas desse ramo como, por exemplo, o serviço de reserva dos produtos para que eles já estejam separados de antemão, o que possibilita ao comprador fazer suas compras com mais facilidade e rapidez no ato da sua viagem.

A partir disso, a previsão de reserva dos pedidos e resgate dos mesmos se tornou algo essencial para tomada de decisão. Especificamente, neste trabalho, analisamos através do modelo SARIMA e do modelo de regressão harmônica duas séries temporais que são descritas a seguir.

1) **Receita Colocada:** Quando o passageiro vai viajar para algum destino internacional, ele pode reservar os produtos que quer comprar assim que passar pela loja no aeroporto, tanto nas lojas de embarque quanto nas lojas de desembarque. Produtos como bebidas, perfumes, eletrônicos podem ser reservados. E, para isso, não é preciso fazer nenhum tipo de pagamento ou pré-pagamento com cartão ou dinheiro. Sendo assim, chamamos de **Receita Colocada** todo o montante em dólar que foi reservado para os passageiros, seja ele feito pelo *call center*, *site* ou *loja*.

2) **Receita Resgatada:** Quando o passageiro passar pela loja, ele pode solicitar ao atendente a sua reserva. A partir disso, ele pode levar toda a sua reserva ou levar somente alguns produtos desta reserva. Além disso, o cliente pode adicionar algum produto que veja na hora que queira comprar. Assim, todo este montante de venda a partir de uma reserva chamamos de **Receita Resgatada**.

Os dados a serem analisados neste trabalho são dados diários de janeiro/2015 até fevereiro/2018. Em uma análise exploratória inicial da receita colocada, temos que o seu valor mínimo é de \$39,90 e o seu valor máximo é de \$15.497,24. O valor médio

dessa nossa variável é de \$5.202,52. Já ao analisarmos a receita resgatada, temos um valor mínimo de \$164,00 e um valor máximo de \$12.446,00. O valor médio dessa nossa variável é de \$3.618,80.

A seguir, introduziremos formalmente a definição de série temporal e seus principais conceitos e comportamentos, tais como estacionariedade e sazonalidade. Em seguida, serão apresentados os modelos que serão utilizados para a análise das séries temporais apresentadas anteriormente: o modelo SARIMA e o modelo de regressão harmônica. Além disso, vamos mostrar como identificar o melhor ajuste para os nossos dados. Por fim, utilizaremos os modelos ajustados para previsão das séries temporais de interesse.

# Capítulo 2

## Série temporal

Uma série temporal pode ser definida como um conjunto de observações ordenadas no tempo. Segundo [Morettin e Tolo \(2006\)](#), os principais pontos de interesse da análise de séries temporais são:

- a) Investigar como as séries foram geradas;
- b) Fazer previsão para tempos futuros;
- c) Descrever o comportamento da série;
- d) Analisar periodicidades relevantes aos dados.

Assim, para descrevermos seu comportamento, temos que levar vários fatores em consideração como, por exemplo, presença de estacionariedade, tendência, sazonalidade, entre outros. Entender o comportamento da série permite o ajuste de um modelo adequado, que, por sua vez, permite a previsão para tempos futuros. Os principais conceitos sobre série temporais serão apresentados a seguir.

## 2.1 Função Média, Variância e Autocovariância

Considere um processo estocástico  $\{Z_t, t \in T\}$ , em que  $T$  é um conjunto cujos elementos correspondem a instantes de tempo. Comumente,  $T = \mathbb{Z}$ . Este processo serve de modelo para uma série temporal observada. Sua função média é definida por:

$$\mu_t = E(Z_t) \quad (2.1)$$

Ou seja, é a média em um determinado tempo  $t$ .

A função de autocovariância é definida por:

$$\gamma_{t,s} = Cov(Z_t, Z_s) = E[(Z_t - \mu_t)(Z_s - \mu_s)] = E(Z_t Z_s - \mu_t \mu_s). \quad (2.2)$$

Já a função de autocorrelação é dada por:

$$\rho_{t,s} = Corr(Z_t, Z_s) \quad (2.3)$$

$$Corr(Z_t, Z_s) = \frac{Cov(Z_t, Z_s)}{\sqrt{Var(Z_t)Var(Z_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}. \quad (2.4)$$

## 2.2 Estacionariedade

De acordo com [Cryer e Chan \(2008\)](#) podemos classificar a estacionariedade em dois tipos:

a) Estacionariedade estrita: quando  $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n})$  tem a mesma distribuição de  $(Z_{t_1-h}, Z_{t_2-h}, \dots, Z_{t_n-h})$  para todo  $t_1, t_2, \dots, t_n$  e toda escolha de tempo de lag  $h$ ;

b) Estacionariedade de segunda ordem (ou fraca): quando a média e variância da série são constantes ao longo do tempo e sua função de covariância depende apenas do lag  $h$ , isto é,  $\gamma_{t,t+h} = \gamma_{0,h}$  para todo tempo  $t$  e lag  $h$ .

### 2.2.1 Ruído Branco

Um dos principais processos estacionários é o ruído branco, uma sequência de variáveis aleatórias independentes e identicamente distribuídas. Assim, segue as características de um processo estacionário: média e variância constantes ao longo do tempo e função de covariância igual a zero para  $h \neq 0$ .

## 2.3 Tendência

Muitos modelos (processos) para séries temporais se baseiam na suposição de que a série temporal apresenta comportamento estacionário. Entretanto, um dos fatores que fazem com que a série não seja estacionária é a tendência. Uma maneira de lidar com a tendência é decompor a série temporal da forma  $Z_t = T_t + \epsilon_t$ , em que  $T_t$  corresponde à tendência e  $\epsilon_t$  é um processo estacionário, comumente assumido ser um ruído branco. Uma abordagem comum é estimar  $T_t$  por meio de modelos de regressão e  $\epsilon_t$  por meio de modelos para séries temporais estacionárias. Um exemplo de série com tendência é o valor do dólar mensalmente, que geralmente apresenta um crescimento (ou decrescimento).

Consideremos o seguinte modelo com tendência:

$$Z_t = T_t + \epsilon_t \tag{2.5}$$

onde  $\epsilon_t$  é um ruído branco com média zero e variância  $\sigma_{\epsilon_t}^2$  e  $T_t$  é a tendência.

Outra forma de tratar a tendência é estacionarizando a mesma através de diferenças. A primeira diferença é definida como

$$\Delta Z_t = Z_t - Z_{t-1}. \quad (2.6)$$

Tomando a segunda diferença:

$$\begin{aligned} \Delta^2 Z_t &= \Delta[\Delta Z_t] = \Delta[Z_t - Z_{t-1}] \\ \Delta^2 Z_t &= Z_t - 2Z_{t-1} + Z_{t-2}. \end{aligned} \quad (2.7)$$

De modo geral,

$$\Delta^n Z_t = \Delta[\Delta^{n-1} Z_t]. \quad (2.8)$$

Aplicar uma ou mais diferenças em uma série temporal muitas vezes remove a tendência, produzindo uma série estacionária, para a qual pode-se aplicar modelos para séries temporais estacionárias.

## 2.4 Sazonalidade

Dizemos que uma série temporal é sazonal quando apresenta um comportamento que se repete de tempo em tempo. A sazonalidade pode ser determinística, quando temos um padrão sazonal regular e estável no tempo assim a prevendo a partir de dados anteriores, ou estocásticas, quando o padrão sazonal varia de acordo com o tempo. Para citar um exemplo, séries mensais de temperatura são geralmente sazonais.

Analogamente ao caso anterior, uma maneira de lidar com a sazonalidade de uma série é decompô-la como:

$$Z_t = T_t + S_t + \epsilon_t. \quad (2.9)$$

em que  $T_t$  corresponde a tendência,  $S_t$  corresponde à sazonalidade e  $\epsilon_t$  a um processo estacionário. Este é um exemplo de um modelo de sazonalidade aditiva. Uma das abordagens para estimação da sazonalidade neste tipo de modelo é por meio de modelos de regressão harmônica, que serão discutidos no capítulo seguinte.

Outra forma de lidar com a sazonalidade de uma série é considerar diferenças sazonais com o intuito de produzir uma série estacionária. A diferença sazonal é definida como

$$\Delta_s(Z_t) = Z_t - Z_{t-s}, \quad (2.10)$$

em que  $s$  corresponde ao período sazonal.

# Capítulo 3

## Modelagem

### 3.1 Modelos para séries estacionárias

Como dito anteriormente, a modelagem estatística é uma importante ferramenta que empresas podem utilizar para tomada de decisões. Neste capítulo serão apresentados alguns modelos estatísticos que podem ser usados para modelar uma série temporal e fazer uma previsão para um tempo futuro da mesma. Os populares modelos ARIMA, também chamados de modelos de Box e Jenkins ([Box e Jenkins \(1970\)](#)).

#### 3.1.1 Modelo AR

O modelo autorregressivo  $AR(p)$  é um modelo para séries estacionárias e, como o próprio nome sugere, a variável no tempo  $t$  é regressada em função dos seus valores anteriores, ou seja,

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + \epsilon_t \quad (3.1)$$

onde  $p$  é a ordem do modelo AR,  $Z_t$  é a série no tempo  $t$ ,  $\epsilon_t$  é um ruído branco com média zero e variância  $\sigma_{\epsilon_t}^2$ , e  $\phi_1, \dots, \phi_p$  são os parâmetros a serem estimados. Os valores de  $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$  são independentes de  $\epsilon_t$ .

A equação pode ser reescrita utilizando o operador de retardo  $B$ , que opera de forma que  $BZ_t = Z_{t-1}$ . Logo, a equação ficará:

$$\phi_p(B)Z_t = \epsilon_t \quad (3.2)$$

onde  $\phi_p(B) = 1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p$ .

### 3.1.2 Modelo MA

Outro modelo para séries temporais estacionárias é o modelo de médias móveis  $MA(q)$ . Neste modelo, a variável no tempo  $t$  é definida em função dos erros dos tempos anteriores, ou seja,

$$Z_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \dots - \theta_q\epsilon_{t-q} \quad (3.3)$$

onde  $q$  é a ordem do modelo MA e  $\theta_1, \dots, \theta_q$  são os parâmetros a serem estimados,  $\epsilon_t$  é um ruído branco com média zero e variância  $\sigma_{\epsilon_t}^2$ .

Podemos reescrever a equação anterior com o operador de retardo  $B$ . Logo, a equação ficará:

$$Z_t = \theta_q(B)\epsilon_t \quad (3.4)$$

onde  $\theta_q(B) = 1 - \theta_1B - \theta_2B^2 - \dots - \theta_qB^q$ .

### 3.1.3 Modelo ARMA

Um modelo mais geral pode ser obtido quando junta-se as partes AR e MA, o que produz o modelo autorregressivo de médias móveis de ordens  $p$  e  $q$   $ARMA(p,q)$ . Supondo

que  $Z_t$  é estacionária, temos seguinte equação para descrever este modelo:

$$\phi_p(B)Z_t = \theta_q(B)\epsilon_t. \quad (3.5)$$

### 3.1.4 Modelo ARIMA

Foi visto no capítulo anterior que uma maneira de lidar com séries que não apresentam comportamento estacionário é tomar diferenças. Se colocarmos uma parte integrada (I) no modelo ARMA, indicando que a série temporal foi diferenciada, temos o modelo autorregressivo integrado de médias móveis ARIMA( $p,d,q$ ), em que  $p$  e  $q$  são as ordens do modelo ARMA e  $d$  corresponde a quantas diferenças foram tomadas da série original.

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)\epsilon_t. \quad (3.6)$$

Mais detalhes sobre modelos ARIMA podem ser vistos em [Morettin e Tolo \(2006\)](#), [Cryer e Chan \(2008\)](#), [Brockwell e Davis \(2010\)](#).

## 3.2 Identificação do modelo

Ao escolhermos um modelo, de acordo com [Cryer e Chan \(2008\)](#), temos que nos atentar ao princípio de parcimônia, onde o modelo utilizado deve ter o menor número de parâmetros que representa adequadamente a nossa série. Com isso, para identificarmos qual o melhor modelo para uma série temporal, primeiro temos que analisá-la de forma mais exploratória: verificar se ela é estacionária ou se necessário fazer alguma transformação nos dados. Se opta-se pela modelagem da série temporal por modelos ARIMA, uma grande ferramenta é a observação do comportamento das funções de autocorrelação e função de autocorrelação parciais amostrais. Detalhes serão visto a seguir.

### 3.2.1 ACF e PACF

Modelos ARMA apresentam funções de autocorrelação (ACF) e função da autocorrelação parcial (PACF) teóricas que se comportam de forma particular. Identificar esses comportamentos nas ACF e PACF amostrais auxilia no processo de identificação do modelo mais adequado para a série temporal sob análise. Abaixo temos a tabela 3.1 com um resumo de como podemos identificar o modelo a partir do ACF e PACF da nossa série de dados.

Tabela 3.1: ACF e PACF

Modelos	ACF	PACF
AR	É infinita em extensão. Decai de acordo com exponenciais e/ou senoides amortecidas.	É finita em extensão. Ela apresenta um corte após a defasagem $p$ .
MA	É finita em extensão. Ela apresenta um corte após a defasagem $q$ .	É infinita em extensão. Decai de acordo com exponenciais e/ou senoides amortecidas.
ARMA	É infinita em extensão. Decai de acordo com exponenciais e/ou senoides amortecidas após a defasagem $(q - p)$ .	Se comporta como uma FACP de MA puro.

### 3.2.2 Estimação por Máxima Verossimilhança

Uma vez identificado o modelo, parte-se para a estimação dos parâmetros. Esta estimação pode ser feita por máxima verossimilhança. Seja  $Z_1, \dots, Z_n$  uma amostra da nossa série e  $\theta$  o parâmetro ou vetor paramétrico a ser estimado. Com isso, temos a função de verossimilhança, que é dada por:

$$L(\boldsymbol{\theta}) = f(z_1, \dots, z_n | \boldsymbol{\theta}), \quad (3.7)$$

em que  $f(z_1, \dots, z_n | \theta)$  corresponde à função de probabilidade ou função de densidade conjunta de  $Z_1, \dots, Z_n$ , que, enquanto função de verossimilhança, é vista com função de  $\boldsymbol{\theta}$ .

O estimador de máxima verossimilhança é definido como o valor  $\hat{\boldsymbol{\theta}}$  que maximiza a função de verossimilhança, isto é,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{arg\,max}} L(\boldsymbol{\theta}; z_1, \dots, z_n). \quad (3.8)$$

Para mais detalhes sobre estimadores de máxima verossimilhança, ver [DeGroot e Schervish \(2003\)](#). Para estimadores de máxima verossimilhança no contexto de séries temporais, mais detalhes podem ser vistos em [Morettin e Tolo \(2006\)](#), [Cryer e Chan \(2008\)](#), [Brockwell e Davis \(2010\)](#).

### 3.2.3 AIC e BIC

Identificar o modelo e suas ordens a partir das ACF e PACF pode ser, em muitos casos, desafiador. Assim, uma vez que isto é feito, é aconselhável que também modelos de diferentes ordens (que sejam próximas das do modelo identificado) sejam ajustados e comparados. Para isso podemos utilizar o critério de informação de Akaike (AIC), que é definido por:

$$AIC = -2\log(L(\boldsymbol{\theta}; z_1, \dots, z_n)) + 2m \quad (3.9)$$

onde  $L$  é a função de verossimilhança do modelo ajustado, e  $m$  o número de parâmetros

do modelo. Este critério tende a escolher modelos com ordens mais altas.

Com isso, podemos utilizar também o critério de informação bayesiana (BIC), que é definido por:

$$BIC = -2\log(L(\theta; z_1, \dots, z_n)) + m\log(n) \quad (3.10)$$

onde  $n$  é o número de observações e  $m$  o número estimado de parâmetros. Quanto menor o valor desses dois critérios, melhor é o modelo ajustado.

### 3.2.4 Teste Ljung-Box

Uma suposição básica dos modelos ARIMA, conforme apresentado anteriormente, é a de que os erros se comportam como um ruído branco. Assim, após o ajuste de um modelo, é necessário verificar se os resíduos não apresentam evidências de dependência no tempo. Para este fim, podemos utilizar o teste de Ljung-Box para verificar a nossa hipótese nula ( $H_0$ ), onde os resíduos são independentes e são identicamente distribuídos.

Sob a hipótese nula ( $H_0$ ), a estatística de teste tem distribuição Qui-quadrada com  $m$  graus de liberdade. Com isso, a estatística do teste é definida por:

$$L = n(n+2) \sum_{k=1}^m \frac{\rho_k^2}{n-k} \quad (3.11)$$

onde  $\rho_k$  é a autocorrelação amostral,  $m$  é o número de parâmetros e  $n$  é o tamanho da amostra.

Assim, para que a hipótese nula seja aceita, precisamos que a autocorrelação tenda a zero, e, conseqüentemente, a nossa estatística de teste seja mínima. A hipótese nula ser rejeitada implica o modelo não estar explicando adequadamente a estrutura temporal da

série. Em seguida podemos fazer o teste Shappiro-Wilk, que será explicado na próxima subseção, para verificar a normalidade dos resíduos.

### 3.2.5 Teste Shappiro-Wilk

A normalidade dos erros é uma das premissas que possibilitam a estimação por máxima verossimilhança e o cálculo dos intervalos de confiança das previsões. A partir deste teste, podemos verificar a normalidade dos resíduos, nossa hipótese nula. Com isso, temos a seguinte estatística de teste:

$$W = \frac{(\sum_{i=1}^n a_i Z_{(i)})^2}{\sum_{i=1}^n (Z_i - \bar{Z})^2} \quad (3.12)$$

onde  $Z_{(i)}$  são os valores amostrais ordenados,  $a_i$  são constantes.

Assim, podemos concluir que valores altos da nossa estatística de teste ( $W$ ) fazem com que rejeitemos a hipótese nula.

## 3.3 Modelos para séries temporais com comportamento sazonal

### 3.3.1 Modelo de regressão harmônica

Em diversos contextos, séries temporais podem apresentar um comportamento sazonal ou cíclico. Para que a análise dessas séries seja feita de forma adequada, estas características devem ser levadas em conta na modelagem. Há diferentes abordagens na literatura para lidar com a sazonalidade de uma série temporal. Uma delas é decompor a série temporal e explicar a sazonalidade de forma determinística por meio de modelos de regressão. Em particular, pode-se utilizar os modelos de regressão harmônica. A regressão harmônica faz uma combinação linear de harmônicos (componentes senoidais). Sendo assim, a variável aleatória  $Z_t$  sendo expressa por:

$$Z_t = \mu + \sum_{n=1}^H (A_n \cos(w_n t) + B_n \sin(w_n t)) + \epsilon_t \quad (3.13)$$

onde  $\mu$  é a média de  $Z_t$ ,  $n = 1, \dots, H$  é o número de harmônicos no modelo,  $A_n$  é o coeficiente do cosseno,  $B_n$  é o coeficiente do seno,  $w_n = (2\pi)/s$  e  $s$  é o período sazonal.

Uma vez que a sazonalidade da série temporal seja explicada pela regressão (parte determinística do modelo), supõe-se que  $\epsilon_t$  seja um processo estacionário que pode ou não preservar dependência temporal. No primeiro caso, pode-se modelar esses erros por meio dos modelos ARMA descritos anteriormente. No segundo caso, pode-se assumir que os erros se comportam como ruído branco.

### 3.3.2 Modelo SARIMA

Diferentemente do modelo de regressão harmônica descrito anteriormente, a sazonalidade da série temporal no modelo SARIMA é explicada de forma puramente estocástica. Com isso, [Box e Jenkins \(1970\)](#) definiram o modelo ARIMA sazonal multiplicativo, ou seja, SARIMA( $p, d, q$ )( $P, D, Q$ ) $_s$ . Sendo ele representado por:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t \quad (3.14)$$

onde  $s$  é a ordem da parte sazonal,  $\Phi$  é o o parâmetro autorregressivo da parte sazonal de ordem  $P$ ,  $\Theta$  é o parâmetro de médias móveis de ordem  $Q$ ,  $\epsilon_t$  é o ruído branco com média zero e variância  $\sigma_{\epsilon_t}^2$ ,  $D$  é a diferença sazonal.

Uma das limitações do modelo SARIMA é que só conseguimos trabalhar com um tipo de sazonalidade de período não muito grande. Por exemplo, se temos uma série temporal mensal durante 4 anos, e nessa série temos uma sazonalidade mensal e uma anual, só conseguimos utilizar uma das duas.

# Capítulo 4

## Análise da receita colocada

Neste capítulo vamos analisar nossa série temporal da receita colocada. Os dados correspondem à receita diária de janeiro/2015 até fevereiro/2018. Em uma análise exploratória inicial da receita colocada, temos que o seu valor mínimo é de \$39,90 e o seu valor máximo é de \$15.497,24. O valor médio dessa nossa variável é de \$5.202,52.

A série pode ser observada na figura [4.1](#). Através da primeira análise deste gráfico, vimos a necessidade de transformar nossos dados para diminuir a escala dos mesmos e tentar induzir uma normalidade aos dados. Porém, vamos confirmar o comportamento dos mesmos, ao analisarmos o ACF e o PACF da nossa série temporal.

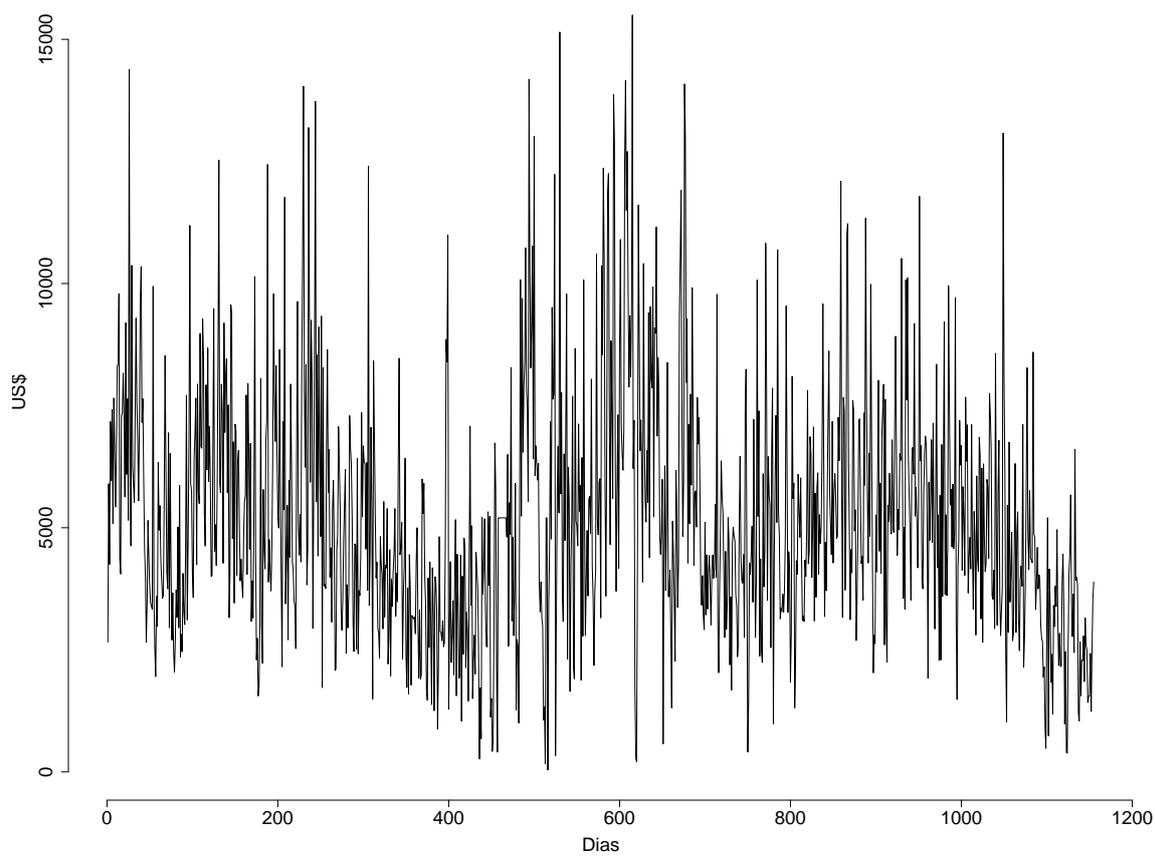
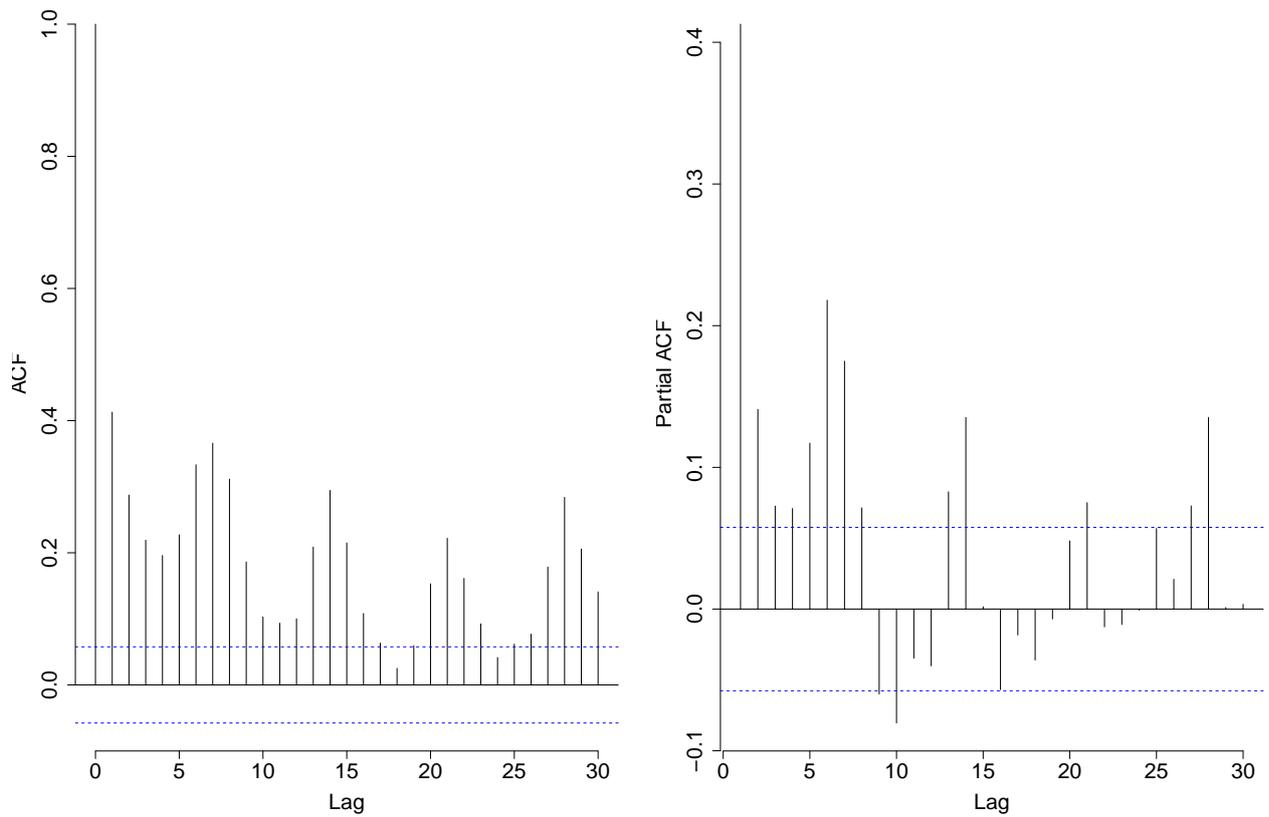


Figura 4.1: Gráfico da receita colocada

Nota-se, a partir do gráfico da série temporal e dos gráficos das ACF e PACF na figura 4.2, que os dados apresentam comportamento sazonal de período 7, semanal. Este resultado é condizente uma vez que os dados são diários.



(a) ACF

(b) PACF

Figura 4.2: ACF e PACF da receita colocada

Transformamos a nossa série temporal da receita colocada aplicando a raiz quadrada. Na figura 4.3 podemos ver o comportamento dos nossos dados transformados.

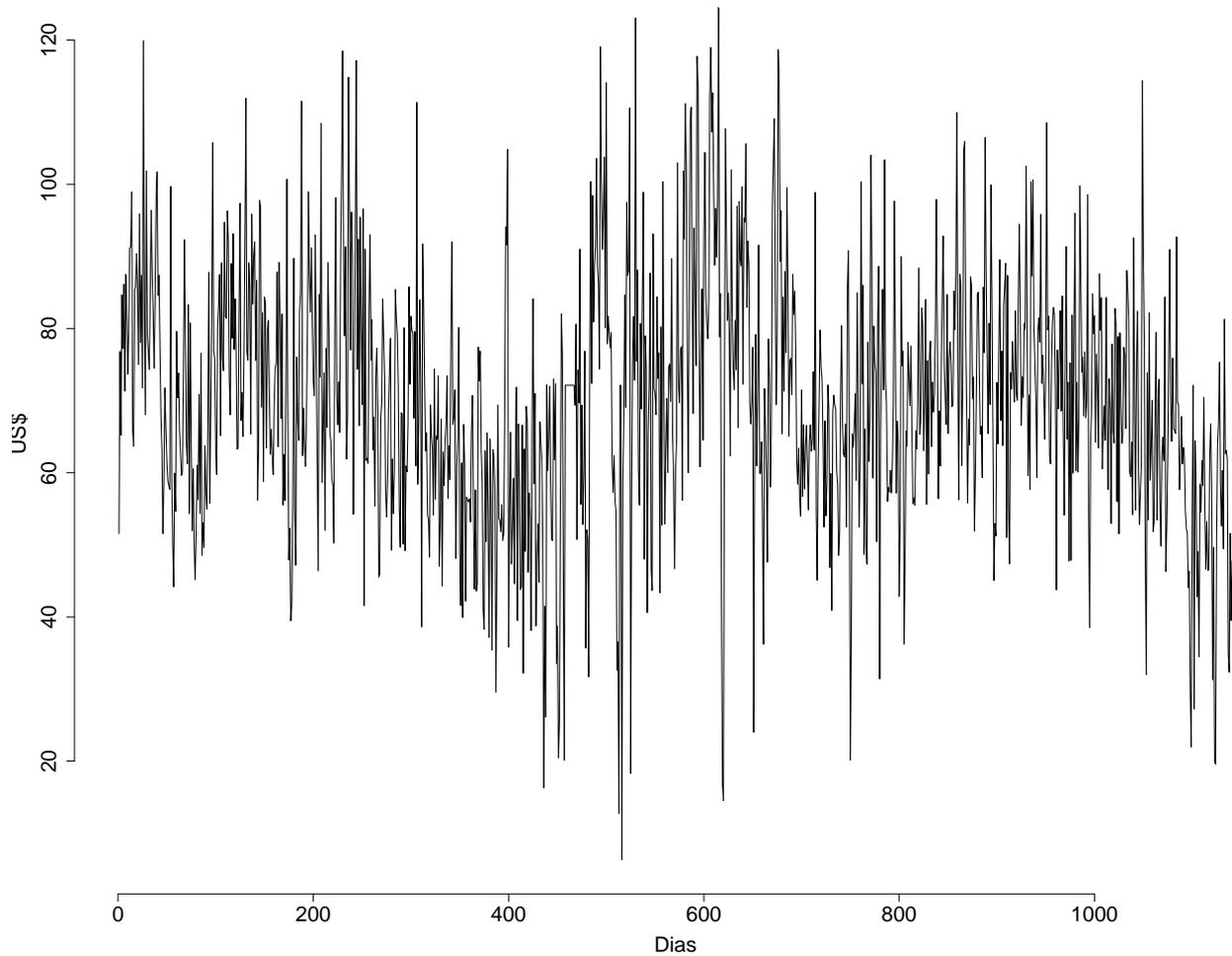


Figura 4.3: Gráfico da receita colocada transformada

## 4.1 Modelo SARIMA aplicado à receita colocada

Uma vez que a série não é estacionária, o próximo passo foi tomar as diferenças para estacionarizar a nossa série temporal. Assim, tiramos uma diferença sazonal de 7 dias e uma diferença comum para tirar qualquer tipo de tendência existente nos nossos dados. Na figura 4.4 temos o gráfico da nossa série de dados diferenciada. É possível observar que agora ela apresenta comportamento estacionário.

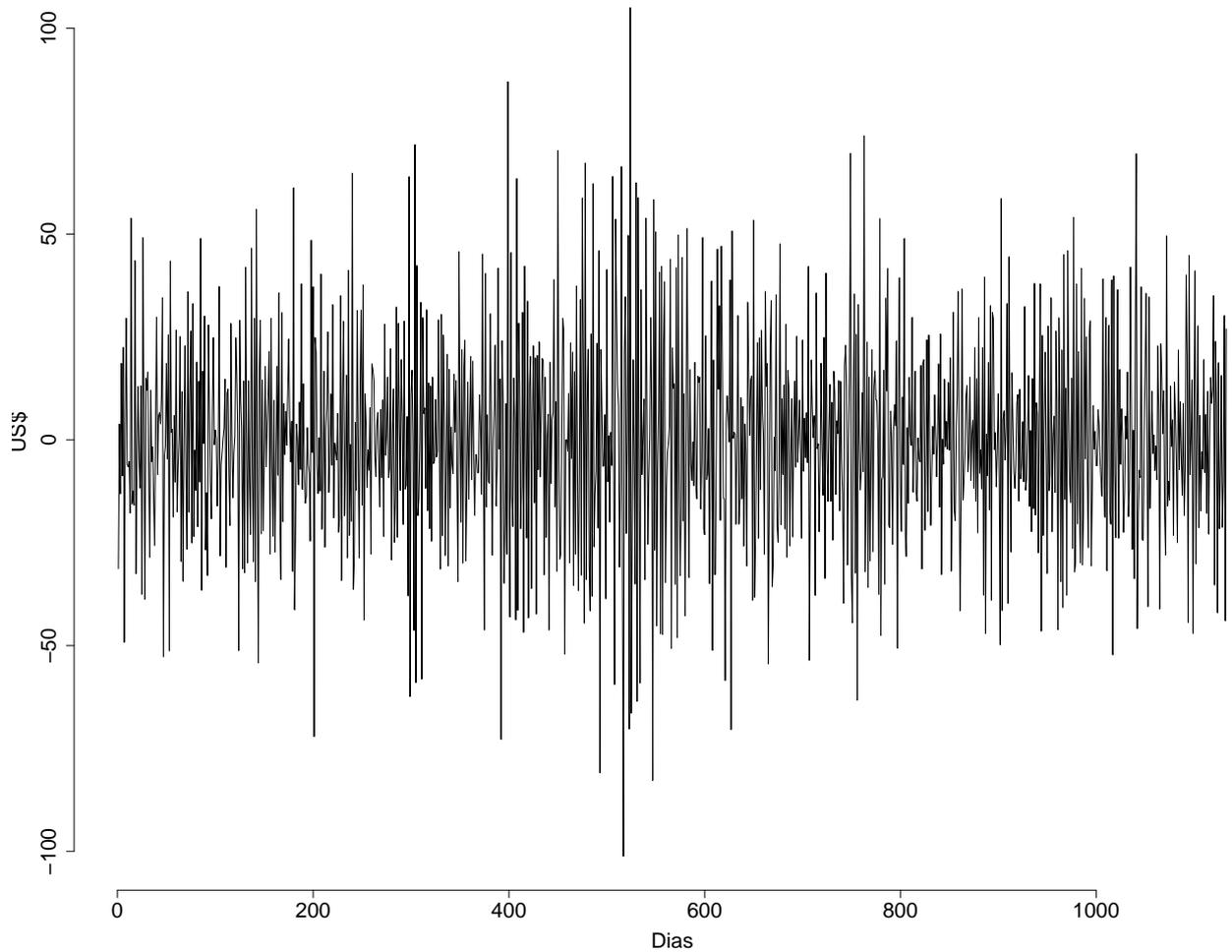


Figura 4.4: Gráfico da receita colocada diferenciada

### 4.1.1 Identificação do modelo

O primeiro passo para analisar qual modelo melhor se aplica aos nossos dados é olhar para a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF) da série diferenciada, ambos explicados anteriormente neste trabalho. Com isso, na [figura 4.5](#) temos os respectivos gráficos.

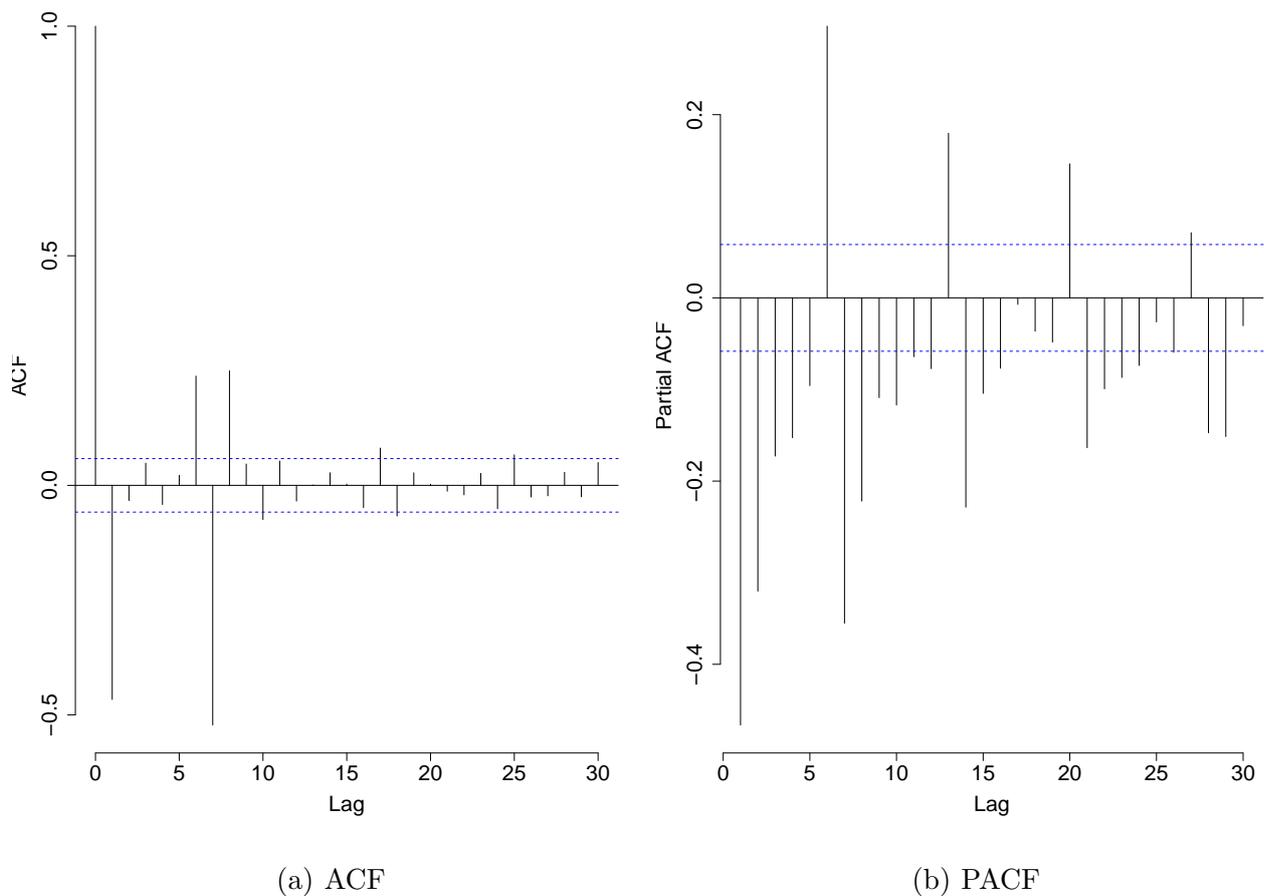


Figura 4.5: ACF e PACF da receita colocada diferenciada

Observando o gráfico da ACF, vemos que o lag 1 é significativo e observando o gráfico da PACF, percebemos um decrescimento exponencial a partir do lag 1. Este comportamento sugere um modelo um modelo de média móvel de ordem 1 para a série diferenciada. Por outro lado, quando observamos a ACF percebemos um decrescimento exponencial quando olhamos somente os lags 7, 14, 21 e assim por diante. Este comportamento das funções de autocorrelação e autocorrelação parcial sugerem um modelo de médias móveis de ordem 1 para a parte sazonal. Assim, o modelo proposto inicialmente é um SARIMA(0,1,1)(0,1,1)<sub>7</sub>. Entretanto, iremos testar alguns modelos com diferentes ordens para identificar qual melhor se adequa a nossa série temporal. Utilizaremos sempre a sazonalidade de 7 dias. Assim, iremos escolher o modelo que apresenta o menor AIC e BIC, ou seja, o que tem menos perda de informação.

Para o ajuste de todos os modelos e todas as análises posteriores, foi utilizado o programa **R Core Team (2013)**. Especificamente, para o ajuste, lançamos mão do comando **arima** do **R Core Team (2013)**, que estima os parâmetros do modelo pelo método da máxima verossimilhança, que no contexto de modelos normais, equivale à estimação por mínimos quadrados.

Portanto, testamos os seguintes modelos:

- a) SARIMA(0,1,1)(0,1,1)<sub>7</sub> com  $AIC = 9322$   $BIC = 9337$ ;
- b) SARIMA(0,1,1)(1,1,1)<sub>7</sub> com  $AIC = 9324$   $BIC = 9344$ ;
- c) SARIMA(1,1,1)(0,1,1)<sub>7</sub> com  $AIC = 9313$   $BIC = 9333$ ;
- d) SARIMA(1,1,1)(1,1,1)<sub>7</sub> com  $AIC = 9315$   $BIC = 9340$ ;
- e) SARIMA(1,1,2)(1,1,1)<sub>7</sub> com  $AIC = 9301$   $BIC = 9331$ ;
- f) SARIMA(1,1,2)(1,1,2)<sub>7</sub> com  $AIC = 9303$   $BIC = 9339$ ;
- g) SARIMA(2,1,2)(1,1,2)<sub>7</sub> com  $AIC = 9304$   $BIC = 9345$ ;

Entre os modelos ajustados, pode-se observar que o SARIMA(1,1,2)(1,1,1)<sub>7</sub> apresenta os menores valores tanto do AIC quanto do BIC. Logo, é o modelo que melhor se ajusta à nossa série temporal, segundo esses critérios; melhor até que o modelo identificado a princípio.

Assim, nossos parâmetros do modelo foram estimados por máxima verossimilhança e, com isso, a nossa equação se apresenta da seguinte forma:

$$(1-0,0222B^7)(1-0,8213B)(1-B)(1-B^7)Z_t = (1+B^7)(1+1,5898B-0,6019B^2)\epsilon_t \quad (4.1)$$

## 4.1.2 Análise dos resíduos

Uma vez o modelo ajustado, vamos agora analisar os resíduos para ver se as suposições do modelo são satisfeitas. Na figura 4.6 podemos observar o gráfico dos mesmos.

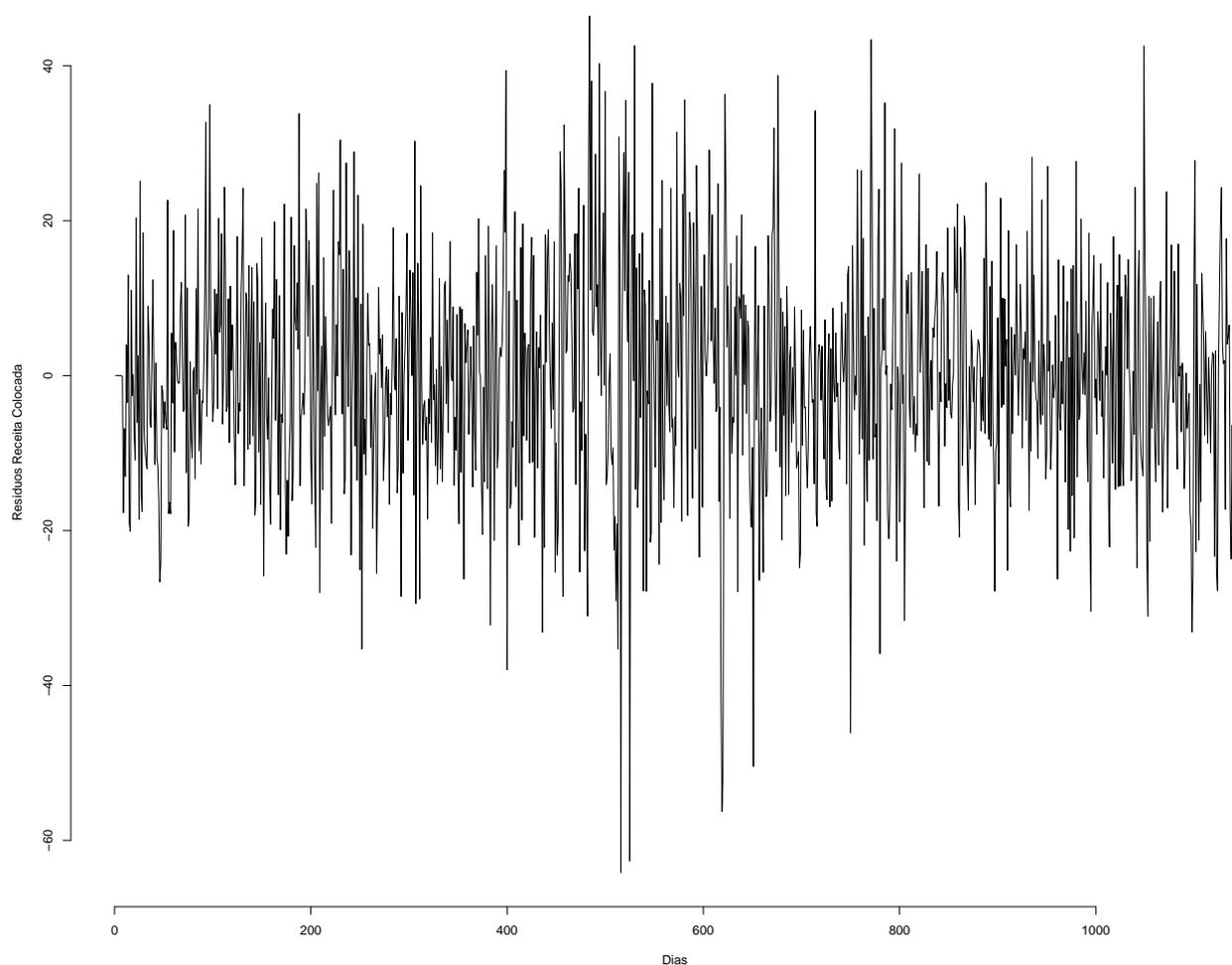


Figura 4.6: Gráfico dos resíduos da receita colocada

Em sequência, na figura 4.7 temos a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF) dos resíduos. É possível observar que praticamente todos os lags estão dentro dos limites, ou seja, não são significativos. Por isso, podemos assumir que nossos resíduos são independentes, o que será confirmado a seguir quando aplicarmos o teste de Ljung-Box nos mesmos. Este teste foi explicado em seções anteriores e mostramos que tem-se a hipótese nula de que os resíduos são independentes.

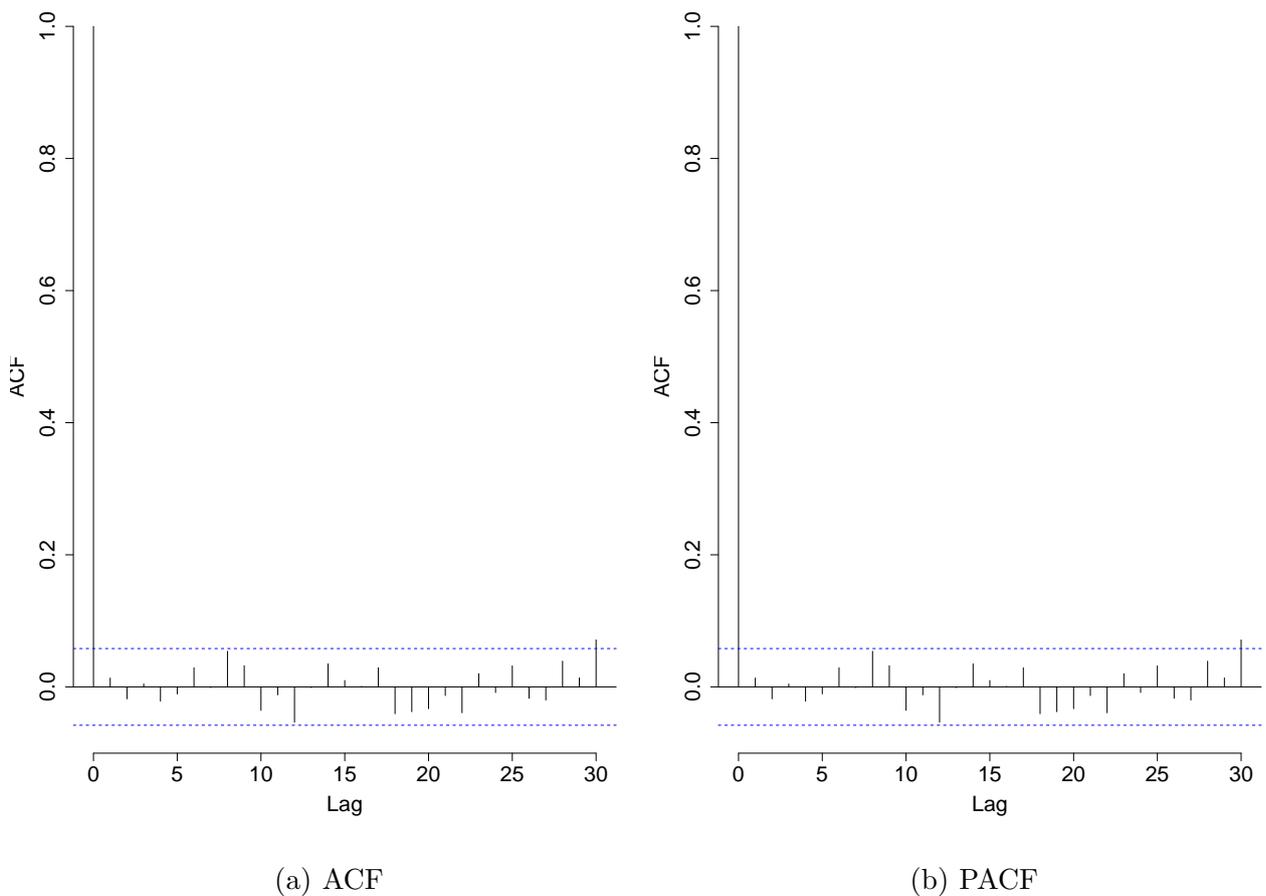


Figura 4.7: ACF e PACF dos resíduos

Ao aplicarmos o teste de Ljung-Box nos resíduos, obtemos o  $p - valor = 0,65$  e o  $\chi^2 = 0,2059$ . Esses valores nos mostram que não temos evidências para rejeitar a hipótese nula, ou seja, não temos evidências de que há dependência entre os resíduos no tempo. Este resultado corrobora com a hipótese do modelo de que os erros se comportam como um ruído branco.

A suposição de normalidade dos erros é importante para o cálculo dos intervalos de confiança da previsão, que será feita posteriormente. Para testar esta hipótese, aplicamos o teste de Shapiro-Wilk, além da análise visual por meio do gráfico *Q-Q plot*, que pode ser visto na figura 4.8.

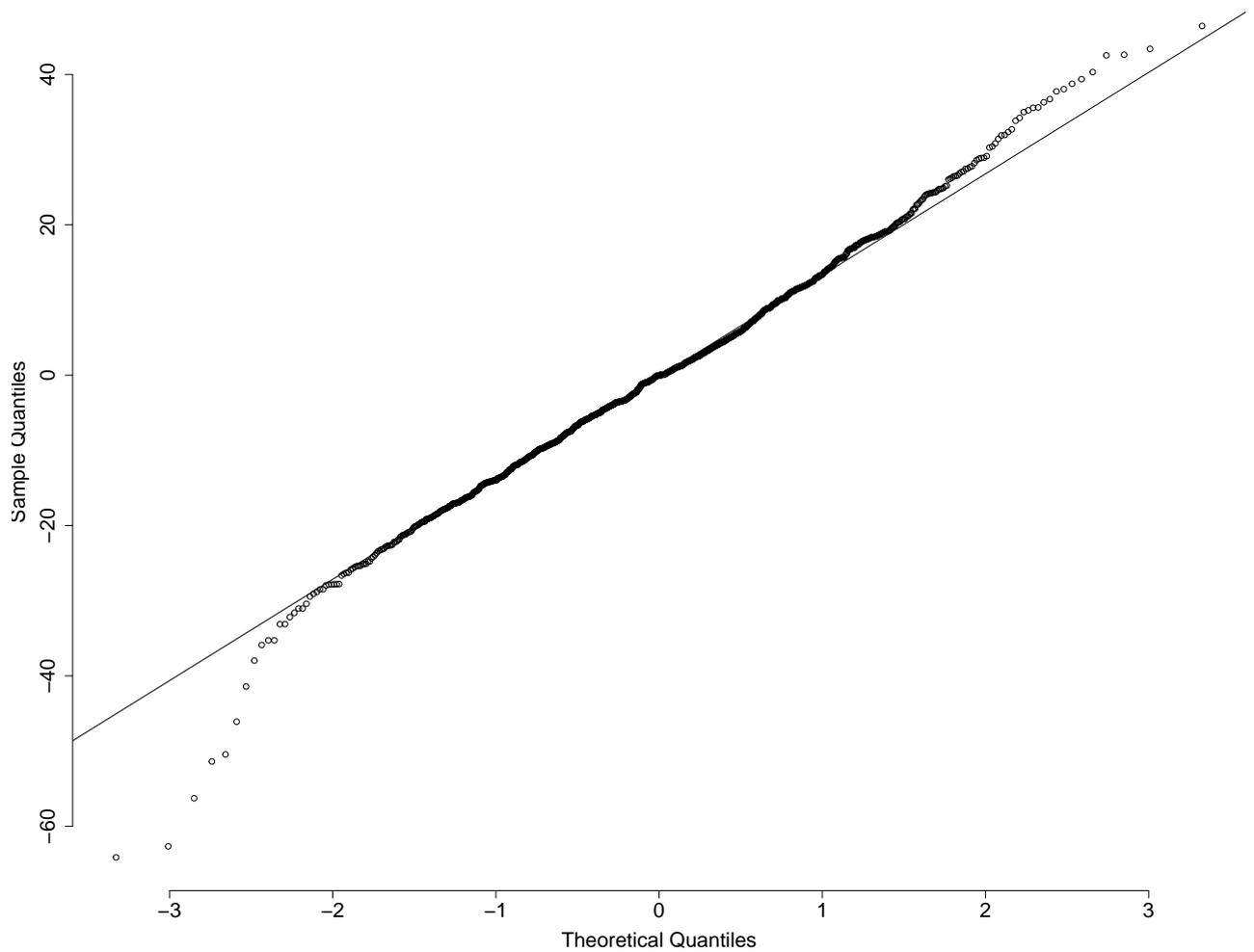


Figura 4.8: Normalidade dos resíduos

Observando o  $Q-Q$  plot, é possível analisar que a cauda dos mesmos não estão perto da linha normal. E, além disso, obtivemos uma estatística de teste  $W = 0,9918$  e um  $p$ -valor =  $5,804e^{-6}$ . Estes valores nos mostram que nossa hipótese nula não é satisfeita, ou seja, há evidências de que os resíduos não têm distribuição normal. Porém, isto pode ser explicado por causa de alguns dias em que o site de reserva não ficou em perfeito funcionamento e, com isso, a sua receita colocada não foi exatamente o que seria em um dia normal.

### 4.1.3 Previsão

Depois de todas as análises feitas, vamos agora fazer a previsão da nossa série temporal. Para testar a qualidade preditiva do modelo escolhido, retiramos as últimas 14 observações da série temporal e calculamos as previsões e os respectivos intervalos de 95% de confiança para os tempos correspondentes. Na figura 4.9, é possível observar os últimos 30 dias da série com a previsão nos últimos 14 dias e seus respectivos limites.

É possível ver que a previsão segue a tendência dos nossos dados observados. Portanto, podemos dizer que ela, junto com todo o nosso ajuste para a série temporal, foi adequado. Assim, seguiremos para o segundo ajuste, utilizando o modelo de regressão harmônica para explicar a sazonalidade.

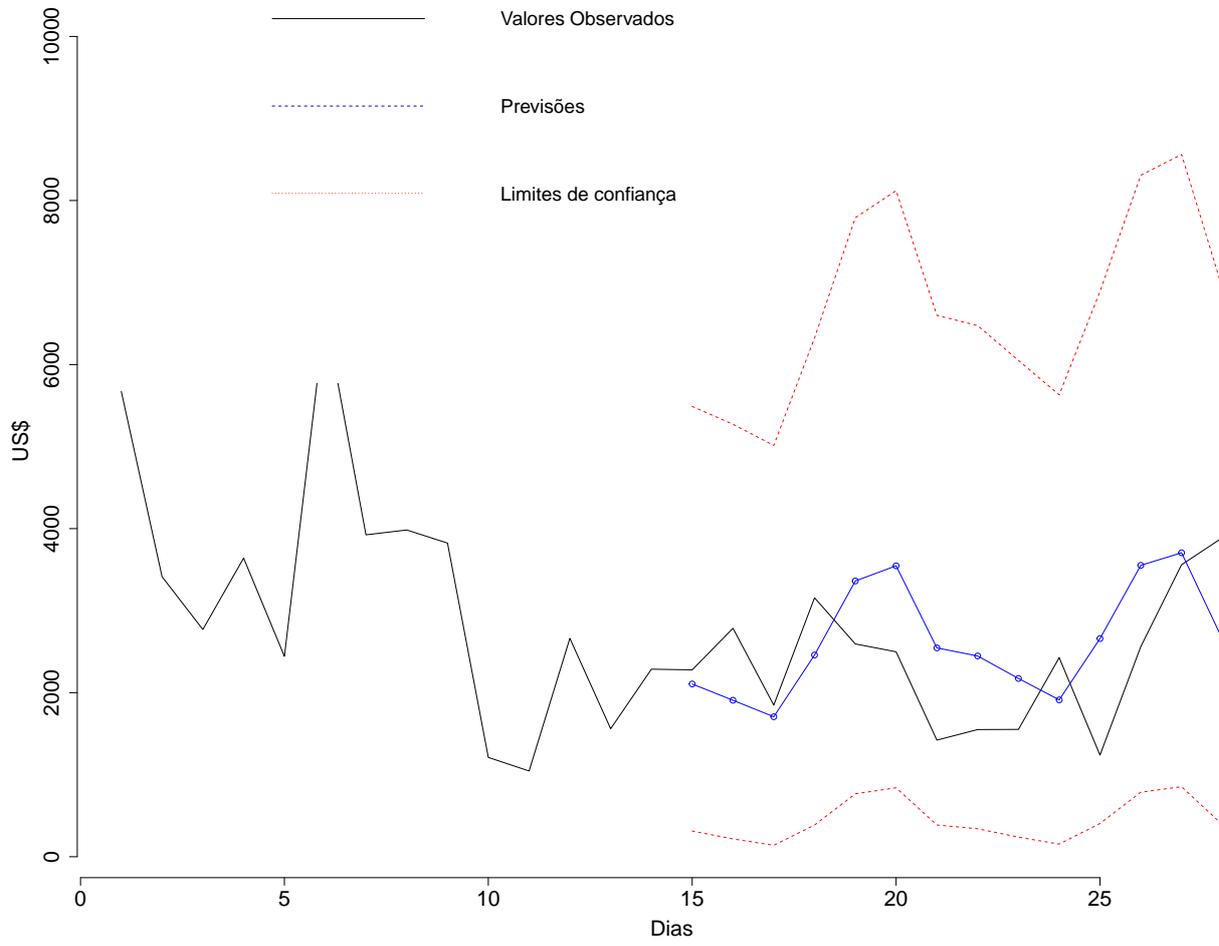


Figura 4.9: Previsão da receita colocada

## 4.2 Modelo de regressão harmônica aplicado à receita colocada

Nesta seção vamos explicar nossos dados através do modelo de regressão harmônica. Ao aplicar este modelo é possível incorporar a sazonalidade para os dois ciclos sazonais (semanal e anual) observados na série temporal. Propomos então o seguinte modelo:

$$Z_t = \beta_0 + \beta_1 \cos(2\pi w_1 t) + \beta_2 \sin(2\pi w_1 t) + \beta_3 \cos(2\pi w_2 t) + \beta_4 \sin(2\pi w_2 t) + \epsilon_t \quad (4.2)$$

em que  $w_1 = 1/7$ ,  $w_2 = 1/365$  e  $\beta_i$ , para  $i = 0, \dots, 4$ , são os coeficientes de regressão a serem estimados.

### 4.2.1 Identificação do modelo

O primeiro passo é ajustar o modelo acima para a série temporal considerando os erros independentes no tempo.

$$\hat{Z}_t = 70,51 + 3\cos(2\pi w_1 t) - 6,79\sin(2\pi w_1 t) - 6,46\cos(2\pi w_2 t) - 3,16\sin(2\pi w_2 t). \quad (4.3)$$

Com isso, o próximo passo é analisar a série temporal dos resíduos do ajuste para verificar se, depois de explicada a sazonalidade, sobrou estrutura temporal.

### 4.2.2 Análise dos resíduos

Primeiramente, vamos plotar os resíduos para vermos como eles se comportam, se aparentam ter um comportamento estacionário. Com isso, na figura 4.10, pode ser vista a série dos resíduos.

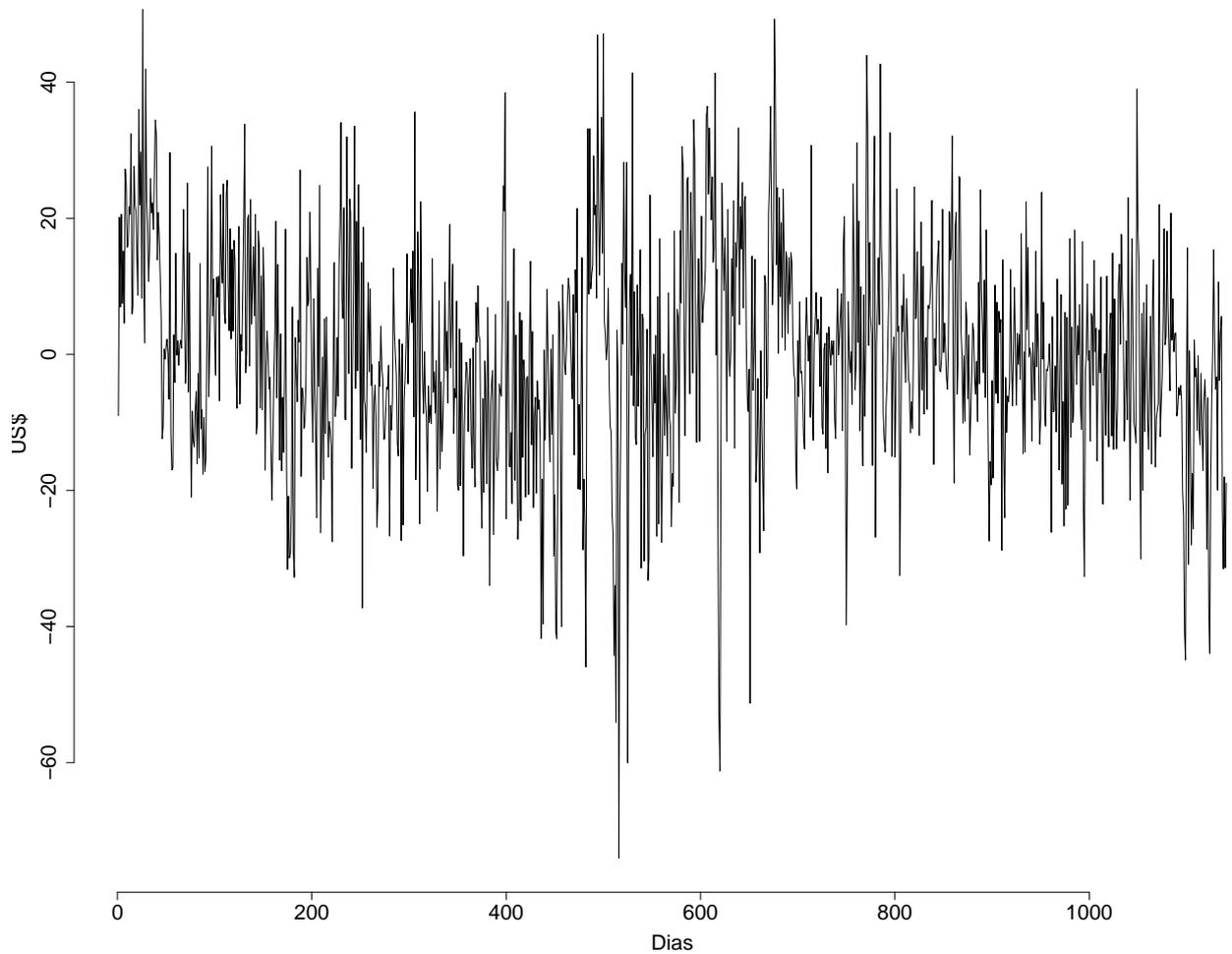


Figura 4.10: Plot dos Resíduos da Receita Colocada

Em sequência, abaixo temos a função de autocorrelação e a função de autocorrelação parcial dos resíduos. É possível observar que nossos resíduos não tem um comportamento de ruído branco. Vemos isso tanto no ACF quanto no PACF pois eles não estão dentro dos limites, assim não sendo independentes. Isso pode ser comprovado ao aplicarmos o teste de Ljung-Box nos resíduos, onde obtemos o  $p\text{-valor} < 2,2e^{-16}$  e o  $\chi^2 = 127$ . Apesar do nosso  $p\text{-valor}$  ser pequeno, a nossa  $\chi^2$  é muito grande. Assim, sendo necessário aplicarmos um outro modelo aos nossos dados.

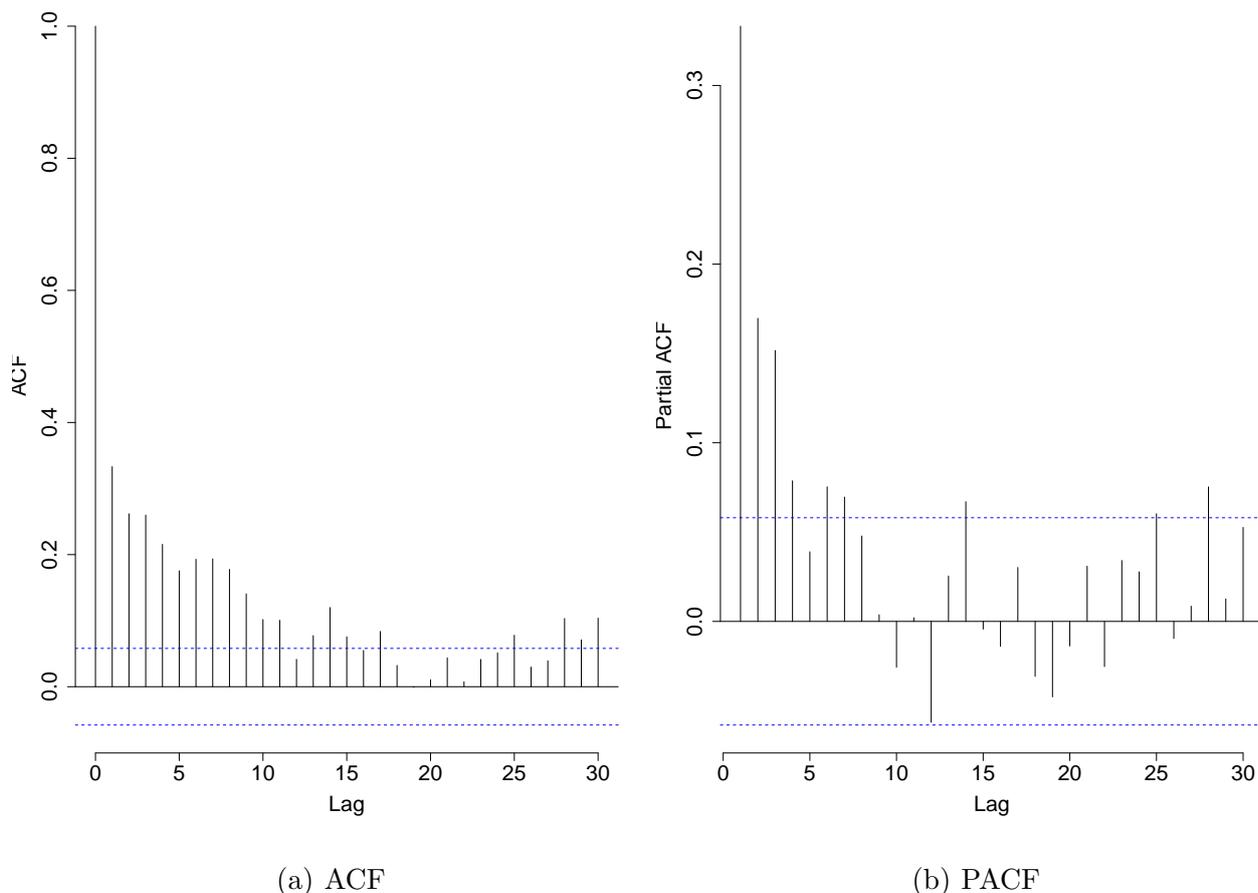


Figura 4.11: ACF e PACF dos resíduos

Observando os gráficos da ACF e PACF, sugerimos um modelo ARMA(1,1) para a série temporal dos resíduos. Uma abordagem possível é ajustar o modelo para os resíduos separadamente. O problema nesta abordagem é que o erro padrão das estimativas é subestimado pelo fato de não estarmos ajustando todos os parâmetros do modelo conjuntamente (os coeficientes de regressão e os parâmetros do modelo ARMA). Assim, aplicamos um modelo ARMA(1,1) juntamente com o modelo harmônico obtido acima. A partir desse modelos obtivemos um  $AIC = 9344,11$  e um  $BIC = 9384,42$ , que são maiores que os valores dos ajustes dos modelos SARIMA, indicando um "pior ajuste", segundo estes critérios. Segue abaixo a sua equação:

$$\hat{Z}_t = 0,90\hat{Z}_{t-1} - 0,69\epsilon_{t-1} + 70,45 + 3,02\cos(2\pi w_1 t) - 6,82\sin(2\pi w_1 t) - 6,49\cos(2\pi w_2 t) - 3,35\sin(2\pi w_2 t). \quad (4.4)$$

Após explicar a sazonalidade através do modelo de regressão e a estrutura temporal do erro por meio do modelo ARMA, espera-se que os resíduos desse ajuste se comportem como ruído branco. Com isso, na figura 4.12 podemos observar seu gráfico.

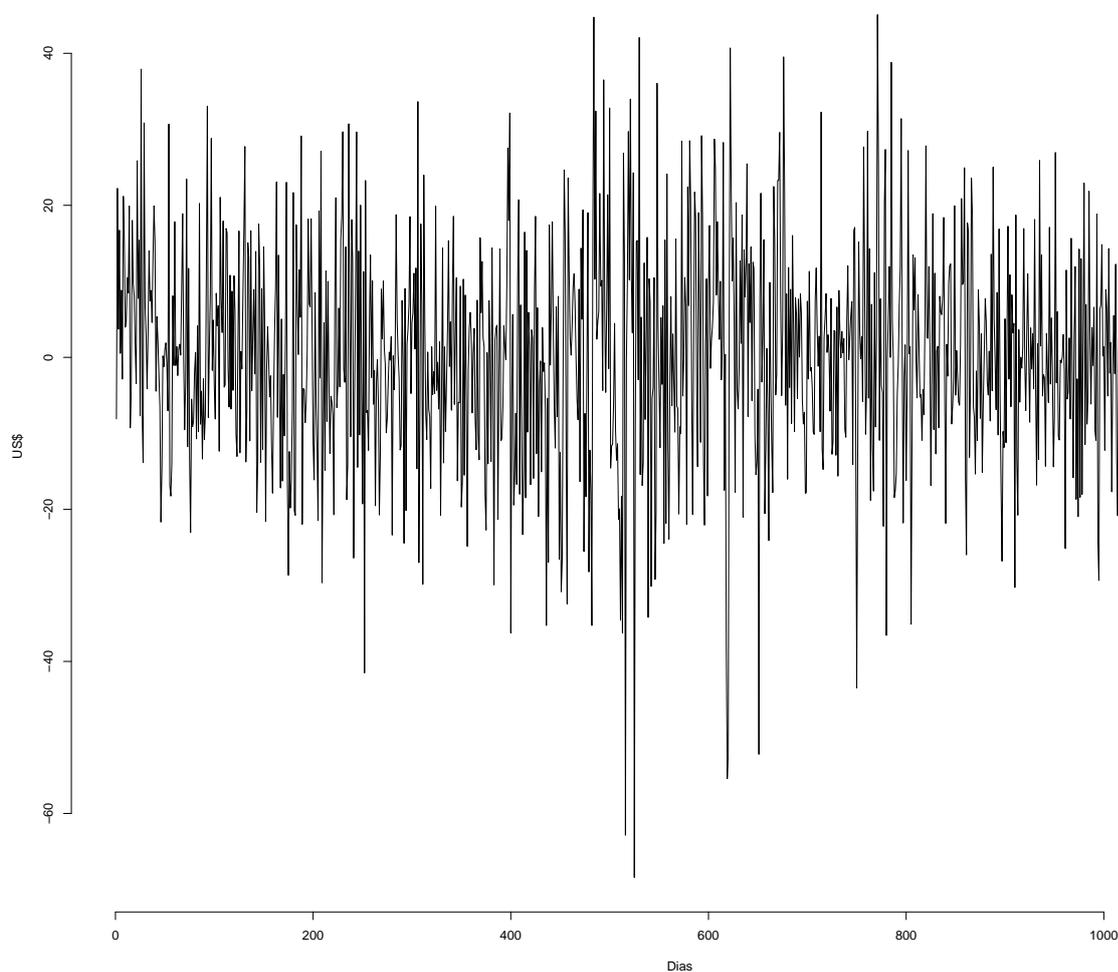


Figura 4.12: Gráfico dos resíduos da receita colocada

Podemos observar no gráfico que os nossos novos resíduos parecem estacionários. Com isso, iremos a seguir mostrar a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF) dos mesmos e depois verificar sua independência e normalidade. Na figura 4.13 temos suas respectivas ACF e PACF.

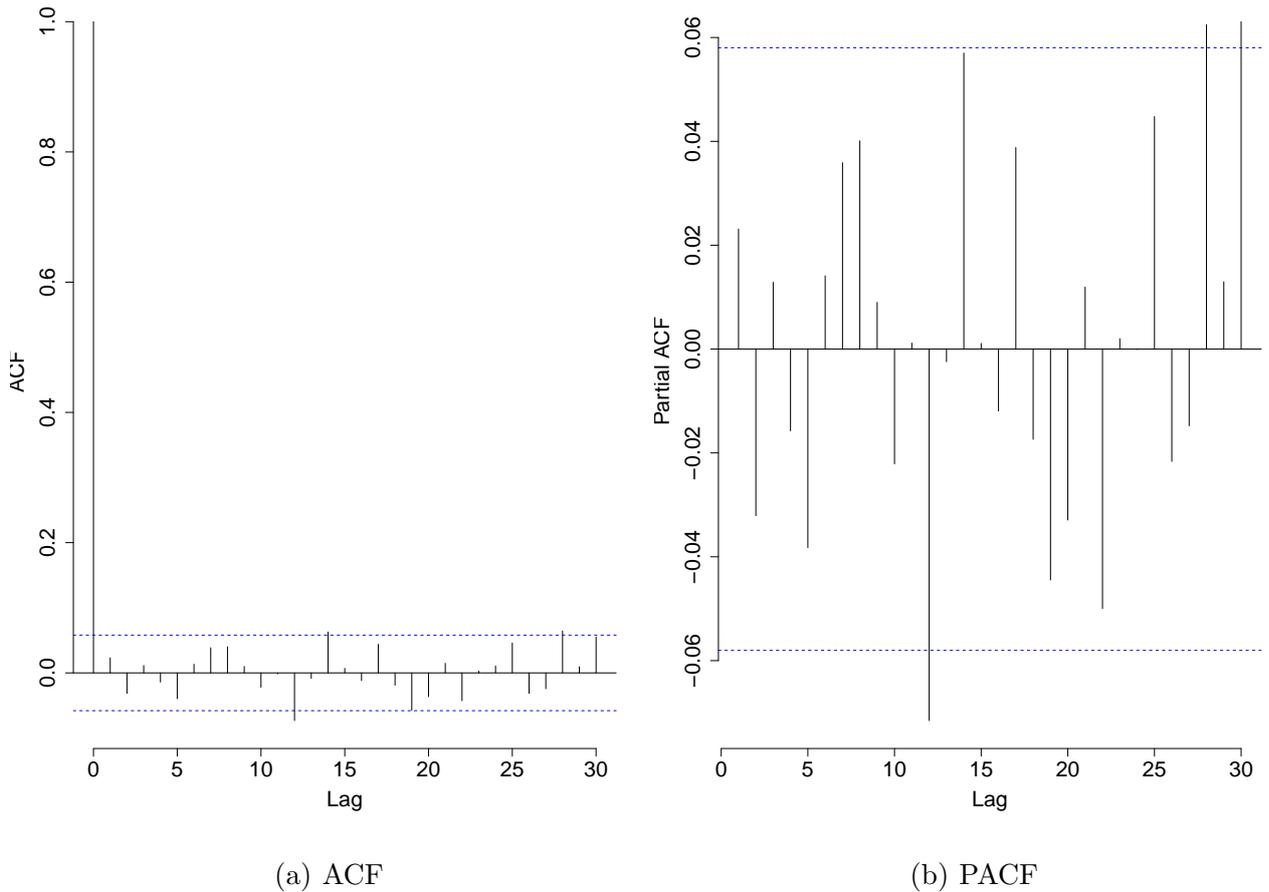


Figura 4.13: ACF e PACF dos resíduos

Em seguida, aplicamos o teste de Ljung-Box nos resíduos e obtivemos o  $p - valor = 0,4343$  e o  $\chi^2 = 0,6113$ . Esses valores nos mostram que aceitamos nossa hipótese nula, ou seja, os resíduos são independentes.

Assim, o nosso próximo passo é aplicar o teste de Shappiro-Wilk. E, com isso, verificar se a hipótese nula dele é aceita. Ou seja, se os nossos resíduos seguem uma distribuição normal. Na figura 4.14 temos o gráfico do nosso teste, nele é possível observar que a cauda dos mesmos não estão perto da linha da distribuição normal, o que poderia nos mostrar uma não normalidade dos resíduos.

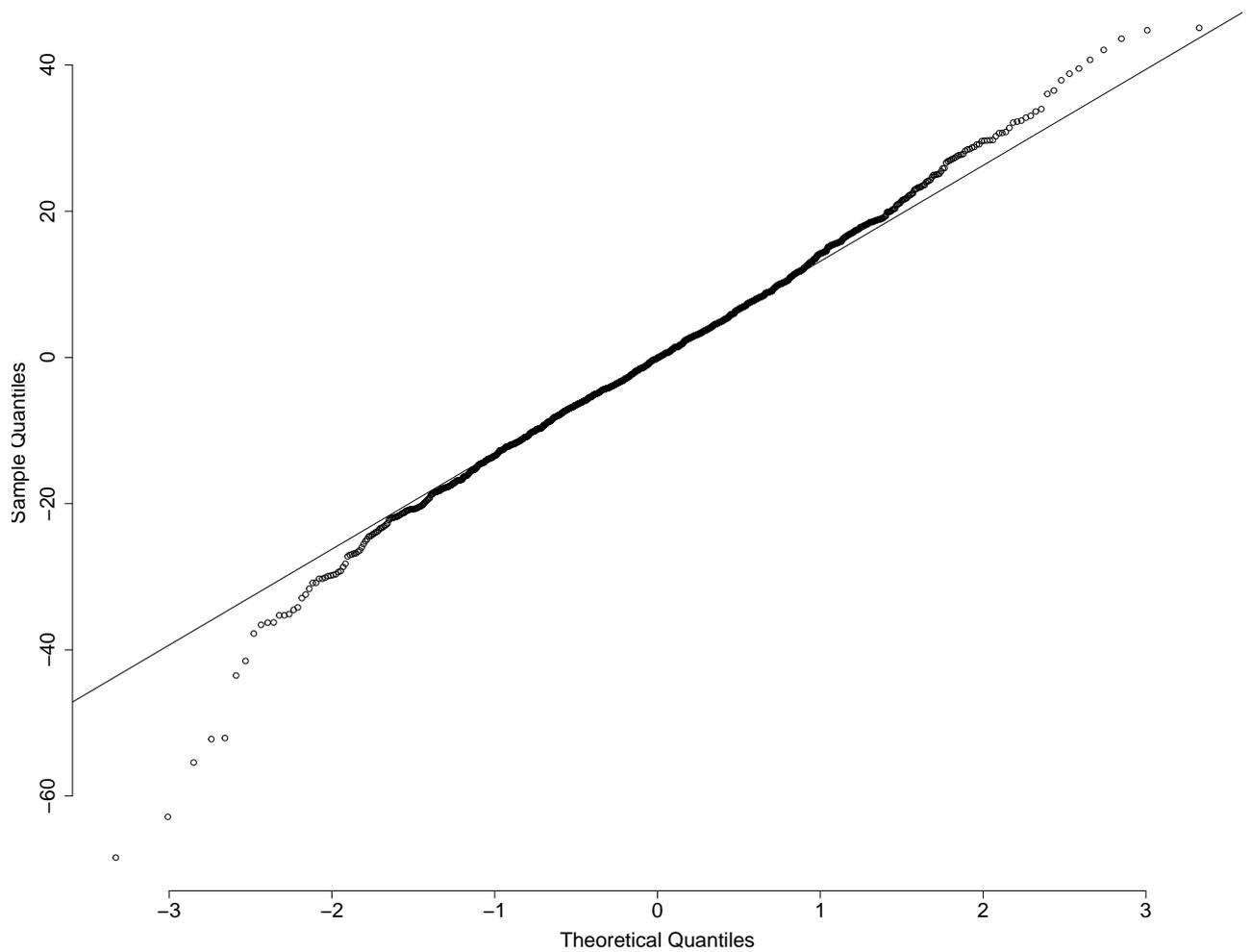


Figura 4.14: Normalidade dos resíduos

Além disso, obtivemos uma estatística de teste  $W = 0,9918$  e um  $p - valor = 5,847e^{-6}$ . Estes valores nos mostram que nossa hipótese nula não é aceita, ou seja, há evidências de que os resíduos não seguem uma distribuição normal. Porém, isto pode ser explicado pelo mesmo motivo da análise dos resíduos quando aplicamos o modelo SARIMA.

Uma das características do modelo de regressão harmônica, como dito anteriormente, é que podemos estimar de forma determinística a sazonalidade. A figura 4.15 apresenta a sazonalidade estimada.

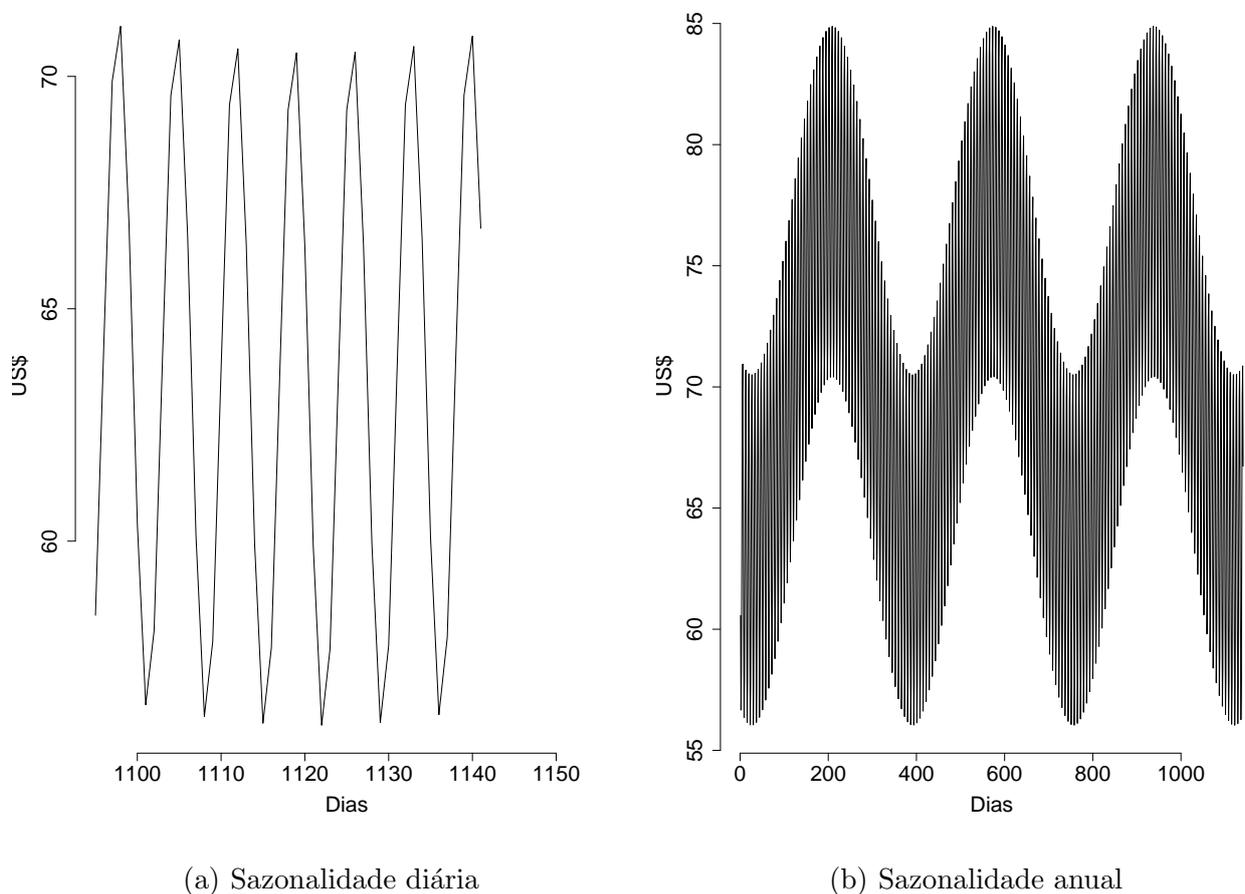


Figura 4.15: Sazonalidade diária e anual estimada

### 4.2.3 Previsão

Depois do ajuste do melhor modelo e da análise dos resíduos, vamos agora fazer a previsão da nossa série temporal de forma análoga à previsão do modelo SARIMA.

Na figura 4.16 é possível observar que a previsão segue a tendência dos nossos dados, também podemos ver que a partir do quinto valor previsto, a previsão se afasta um pouco dos nossos dados. Mesmo assim, podemos dizer que o ajuste e a previsão foram adequados.

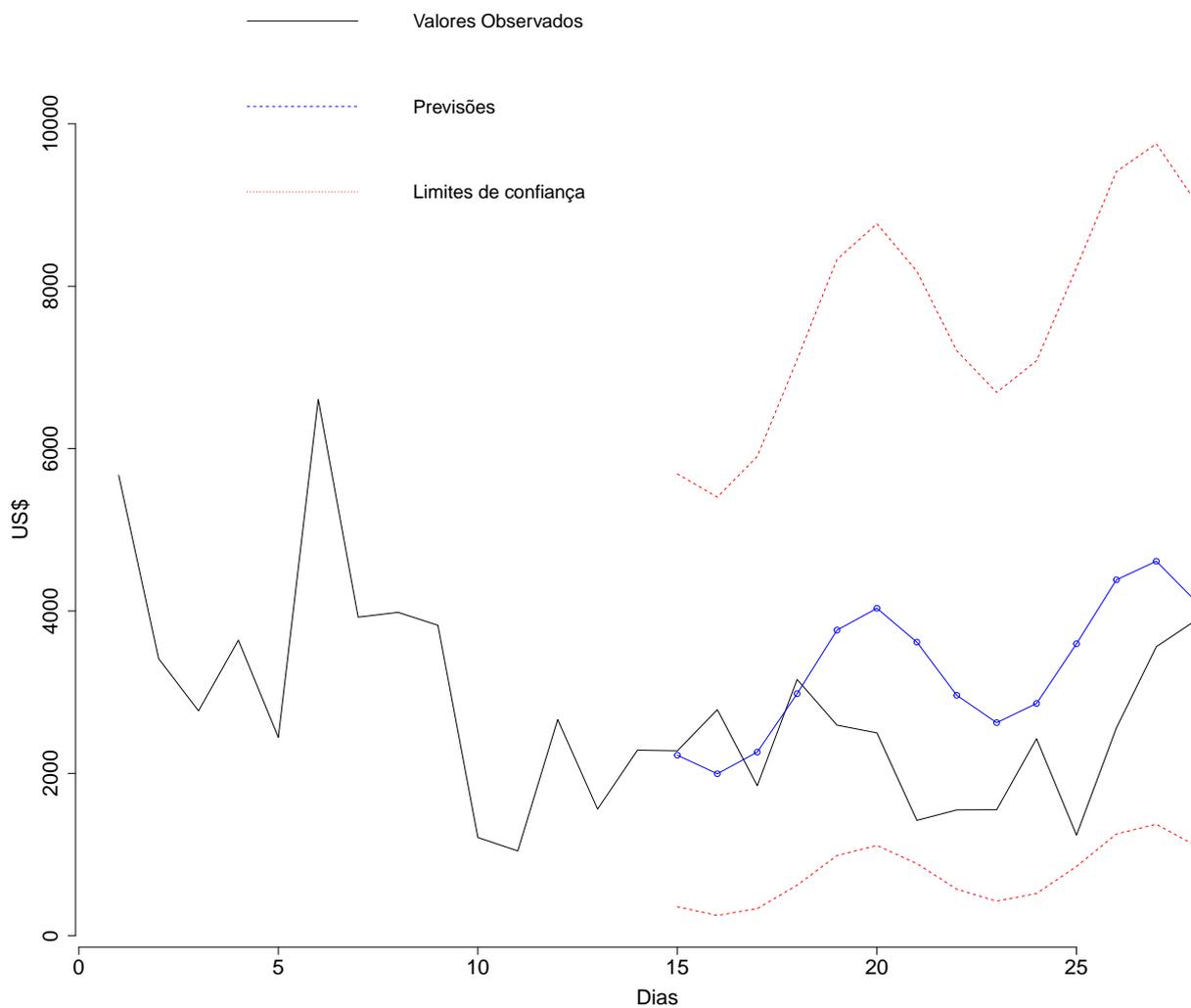


Figura 4.16: Previsão da receita colocada

# Capítulo 5

## Análise da receita resgatada

Agora vamos analisar os nossa série temporal da receita resgatada. Vamos fazer o mesmo processo que fizemos em relação a série da receita colocada. Analisando qual modelo melhor se adequa a nossa série temporal, utilizando tanto o modelo SARIMA quanto o modelo harmônico através de duas sazonalidades.

Ao analisarmos a série temporal da receita resgatada, temos um valor mínimo de \$164,00 e um valor máximo de \$12.446,00. O valor médio dessa nossa variável é de \$3.618,80.

Na figura [5.1](#) podemos ver o gráfico dos dados. Eles estão distribuídos diariamente e em uma escala em dólar. É possível observar que eles apresentam uma grande variação. Por isso, vimos a necessidade de transformar nossos dados para diminuir a escala dos mesmos e tentar induzir uma normalidade aos dados. Vamos confirmar este comportamento ao analisarmos a ACF e a PACF da nossa série temporal.

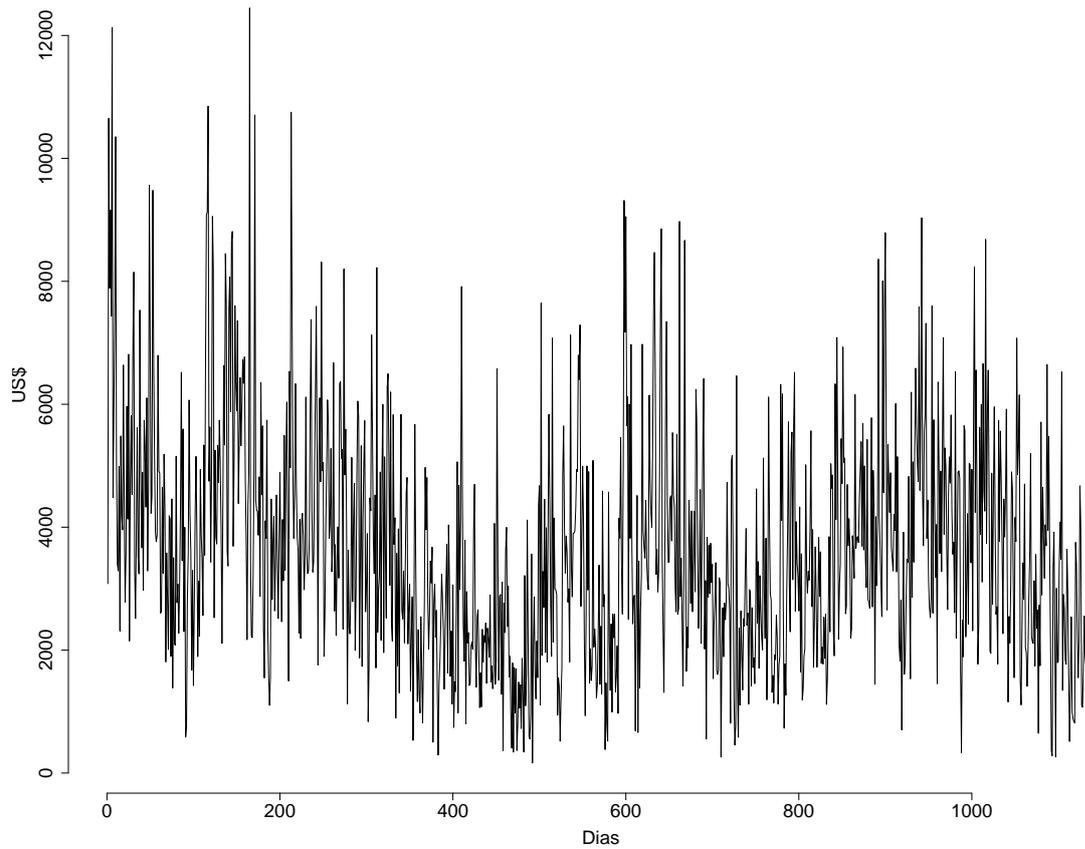
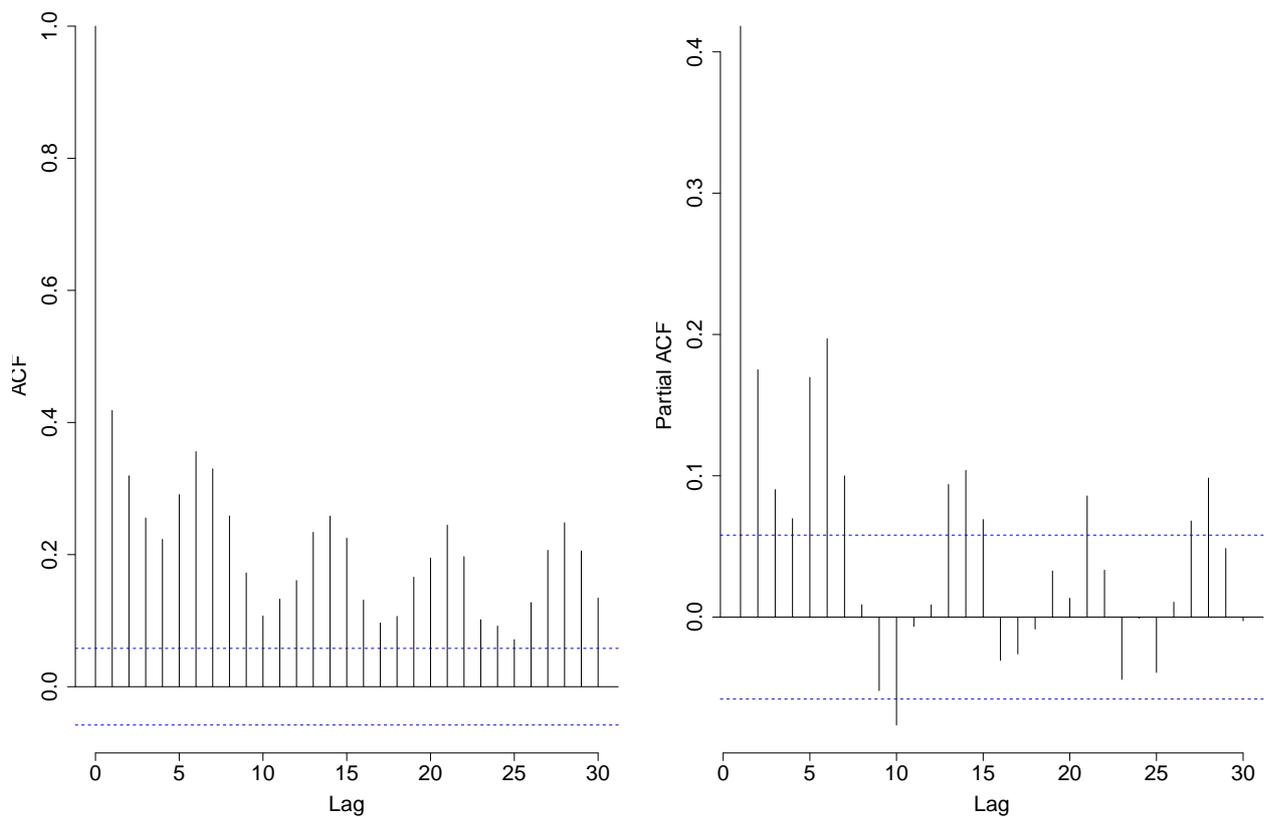


Figura 5.1: Gráfico da receita resgatada

Pode-se ver a partir do gráfico da série temporal e dos gráficos das ACF e PACF na figura 5.2, que os dados apresentam comportamento sazonal de período 7, semanal. Este resultado é compreensível, uma vez que os dados são diários.



(a) ACF

(b) PACF

Figura 5.2: ACF e PACF da receita resgatada

Assim, utilizamos a mesma transformação que fizemos na receita colocada, aplicamos a raiz quadrada na nossa série temporal da receita resgatada. Podemos ver na figura 5.3 o comportamento dos nossos dados transformados.

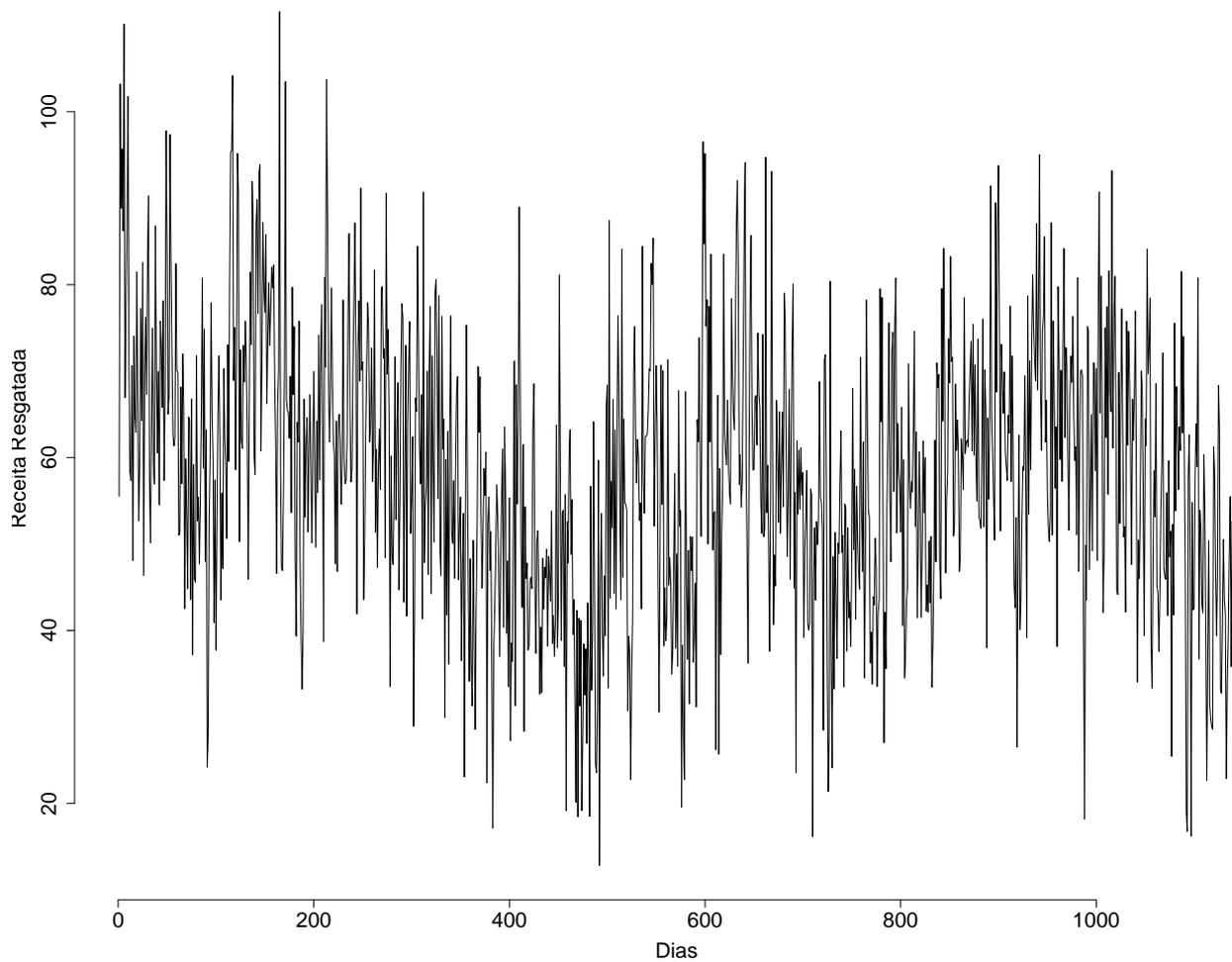


Figura 5.3: Gráfico da receita resgatada transformada

## 5.1 Modelo SARIMA aplicado à receita resgatada

Como a nossa série não é estacionária, o próximo passo foi tomar as diferenças para estacionarizar a nossa série temporal. Portanto, tiramos uma diferença sazonal de 7 dias e uma diferença comum para tirar qualquer tipo de tendência existente nos nossos dados. Podemos ver na figura 5.4 o gráfico da nossa série de dados diferenciada. É possível observar que agora ela apresenta um comportamento estacionário.

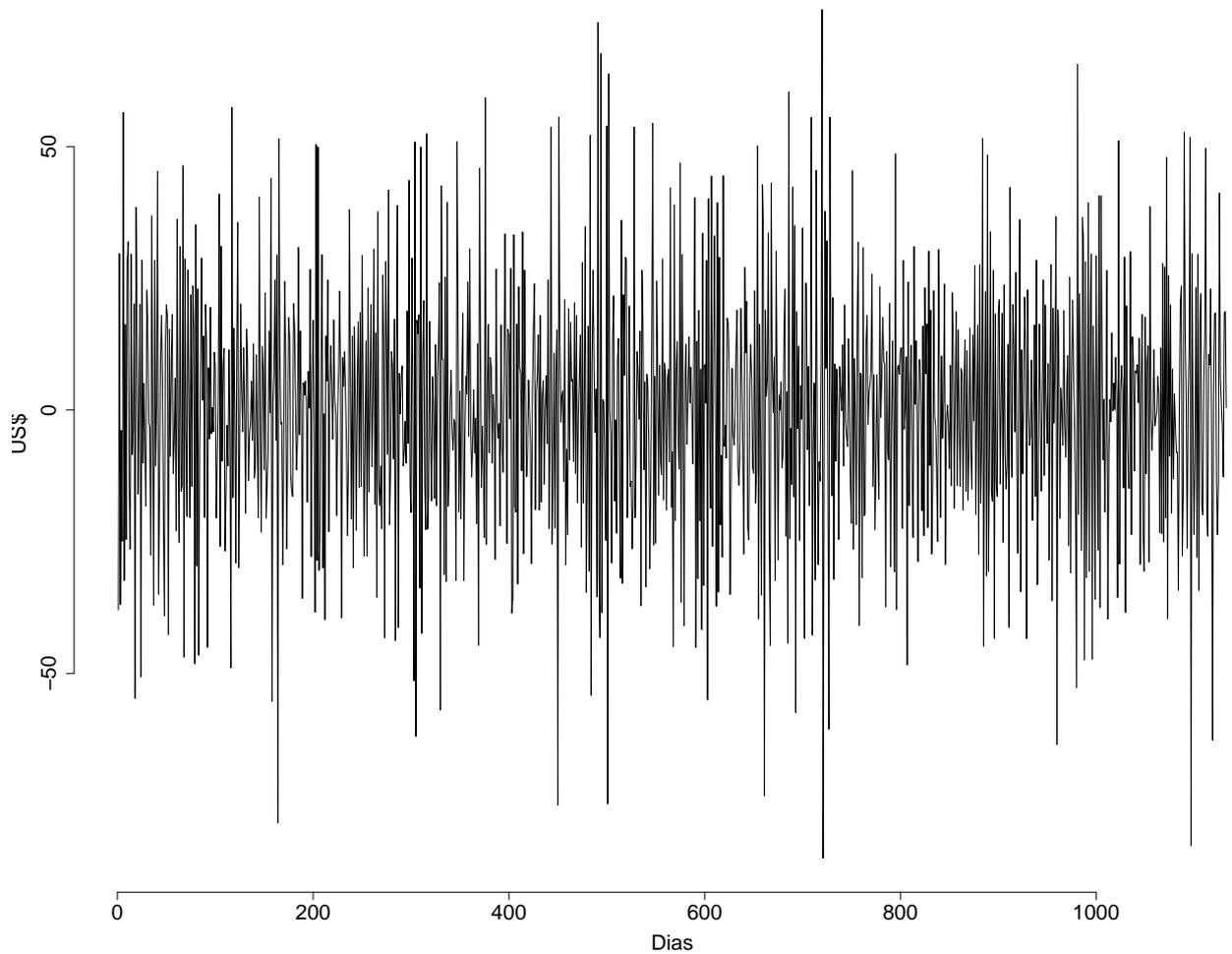


Figura 5.4: Gráfico da receita resgatada diferenciada

### 5.1.1 Identificação do modelo

Primeiramente, vamos analisar a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF) da série diferenciada, ambos explicados anteriormente neste trabalho. Assim, na figura 5.5 temos os respectivos gráficos para analisarmos qual modelo podemos ajustar aos nossos dados.

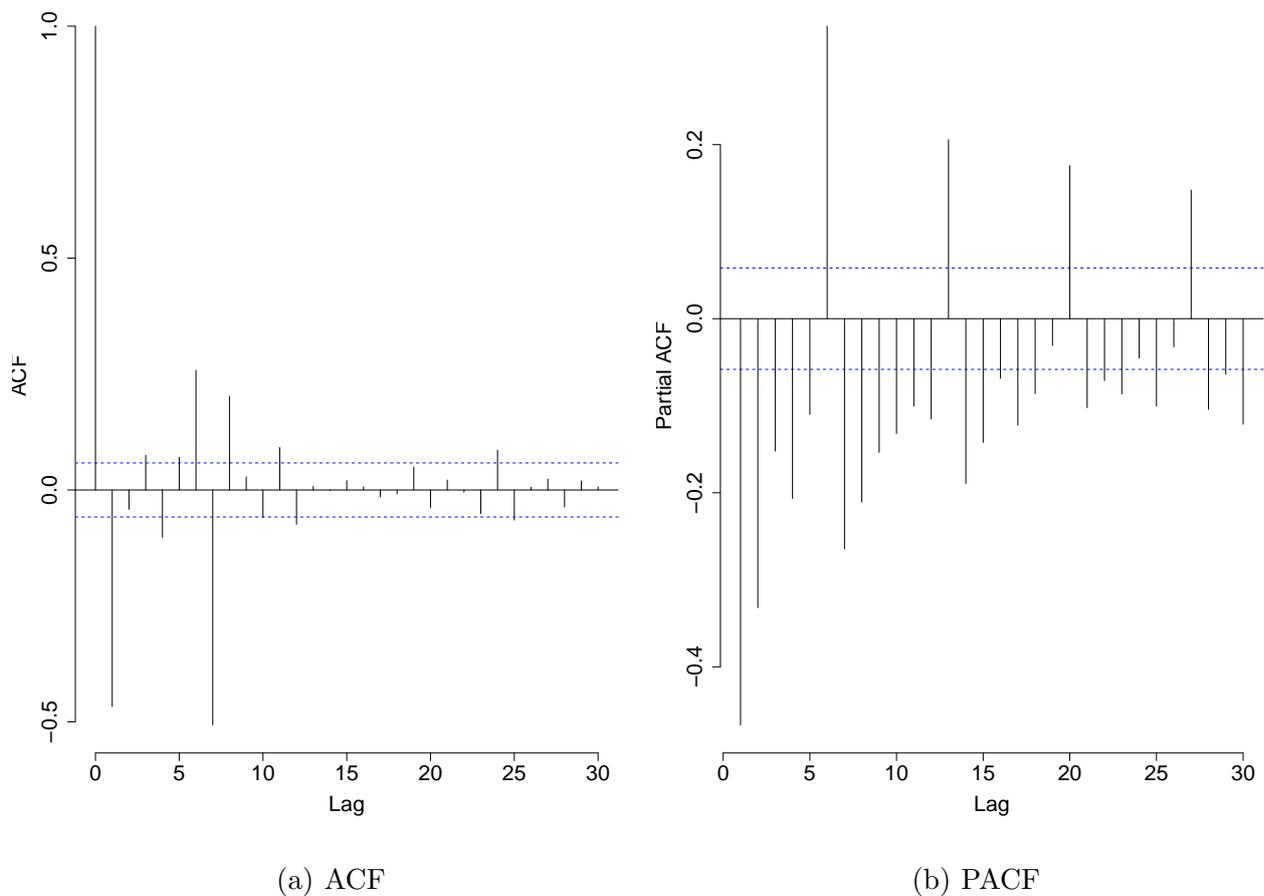


Figura 5.5: ACF e PACF da receita resgatada diferenciada

Os gráficos da figura 5.5 são muito parecidos com os da série temporal da receita colocada, podemos observar o mesmo comportamento. Por isso, o modelo sugerido também é um SARIMA(0,1,1)(0,1,1)<sub>7</sub>. A seguir iremos testar o modelo sugerido e outros modelos para identificar qual melhor se adequa a nossa série temporal. E iremos escolher o modelo que apresenta o menor AIC e BIC, ou seja, o que tem menos perda de informação.

Conforme falado anteriormente, foi utilizado o programa **R Core Team (2013)** para todas as análises neste projeto.

Com isso, testamos os seguintes modelos:

a) SARIMA(0,1,1)(0,1,1)<sub>7</sub> com  $AIC = 9087$   $BIC = 9102$ ;

b) SARIMA(1,1,1)(0,1,1)<sub>7</sub> com  $AIC = 9083$   $BIC = 9103$ ;

c) SARIMA(0,1,1)(1,1,1)<sub>7</sub> com  $AIC = 9087$   $BIC = 9107$ ;

d) SARIMA(1,1,1)(1,1,1)<sub>7</sub> com  $AIC = 9084$   $BIC = 9109$ ;

e) SARIMA(1,1,2)(1,1,1)<sub>7</sub> com  $AIC = 9070$   $BIC = 9100$ ;

f) SARIMA(1,1,2)(1,1,2)<sub>7</sub> com  $AIC = 9085$   $BIC = 9121$ ;

g) SARIMA(2,1,2)(1,1,1)<sub>7</sub> com  $AIC = 9082$   $BIC = 9117$ ;

A partir dos resultados obtidos a cima, é possível observar que o primeiro modelo SARIMA(1,1,2)(1,1,1)<sub>7</sub> apresenta os menores valores de AIC e BIC, assim ele sendo o melhor ajuste para a nossa série tempoal da receita resgatada.

A seguir temos a equação do nosso modelo. Os nossos parâmetros do ajuste foram estimados por máxima verossimilhança.

$$(1+0,0205B^7)(1-0,8657B)(1-B)(1-B^7)Z_t = (1+0,9783B^7)(1+1,6533B-0,6620B^2)\epsilon_t \quad (5.1)$$

### 5.1.2 Análise dos resíduos

Com isso, vamos agora analisar os resíduos desse nosso ajuste para ver se as suposições do modelo são satisfeitas. Na figura 5.6 temos o gráfico deles que estão distribuídos por dia.

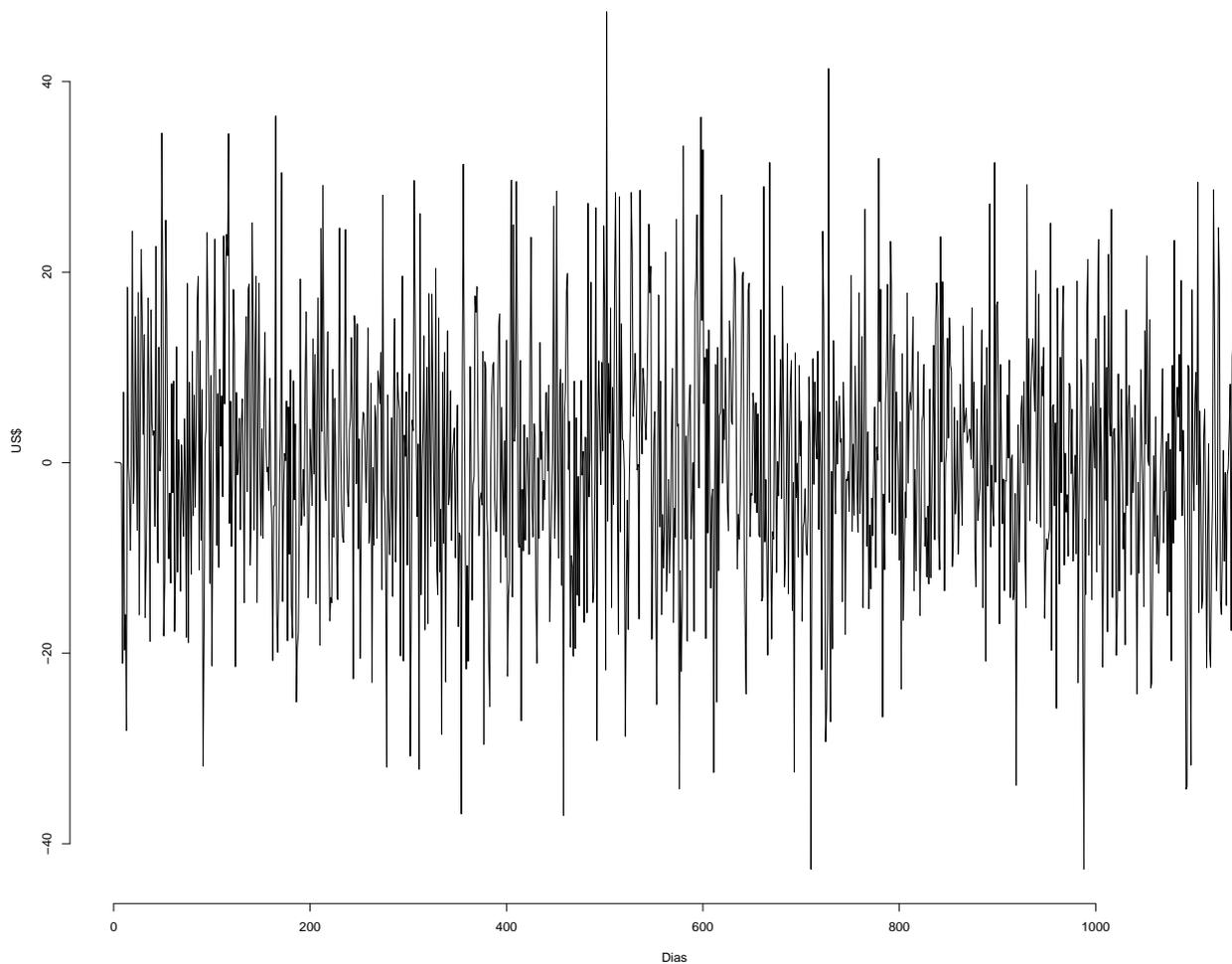


Figura 5.6: Gráfico dos resíduos da receita resgatada

Em seguida geramos a função de autocorrelação e a função de autocorrelação parcial dos resíduos. Na figura 5.7 podemos observar que os lags estão dentro dos limites. Por isso, podemos assumir que nossos resíduos são independentes. Para que isso seja confirmado, aplicaremos o teste de Ljung-Box nos mesmos. Este teste foi explicado em seções anteriores e mostramos que temos a hipótese nula de que os resíduos são independentes.

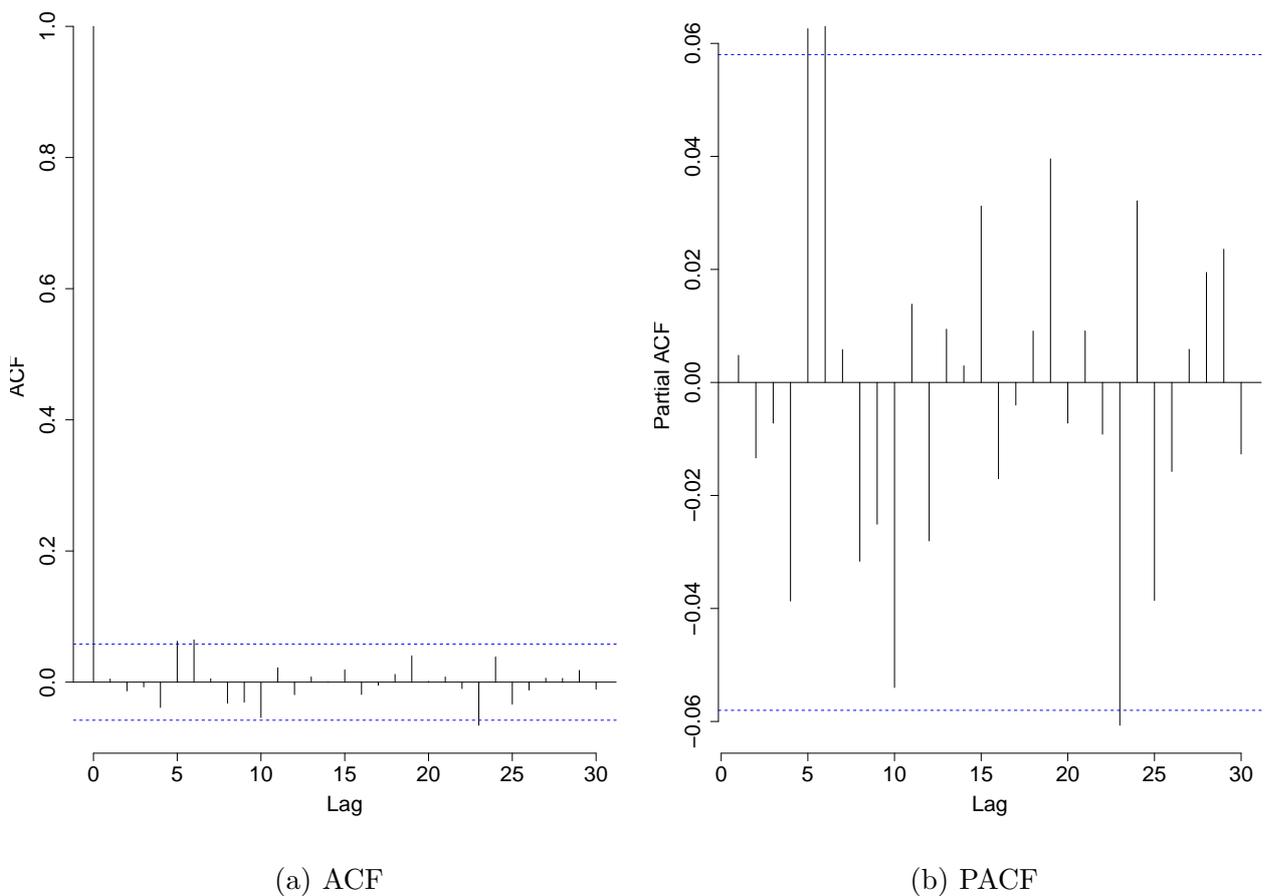


Figura 5.7: ACF e PACF dos resíduos

Ao aplicarmos o teste de Ljung-Box nos resíduos, obtemos o  $p - valor = 0,8711$  e o  $\chi^2 = 0,0263$ . A partir destes resultados, temos evidências de que nossos resíduos são independentes. A partir desse resultado, podemos observar que os mesmos se comportam como ruído branco.

Portanto, o próximo passo é aplicarmos o teste de Shappiro-Wilk. A partir deste teste vamos verificar se os nossos resíduos seguem uma distribuição normal. Na figura 5.8 temos o gráfico *Q-Q plot* do nosso teste.

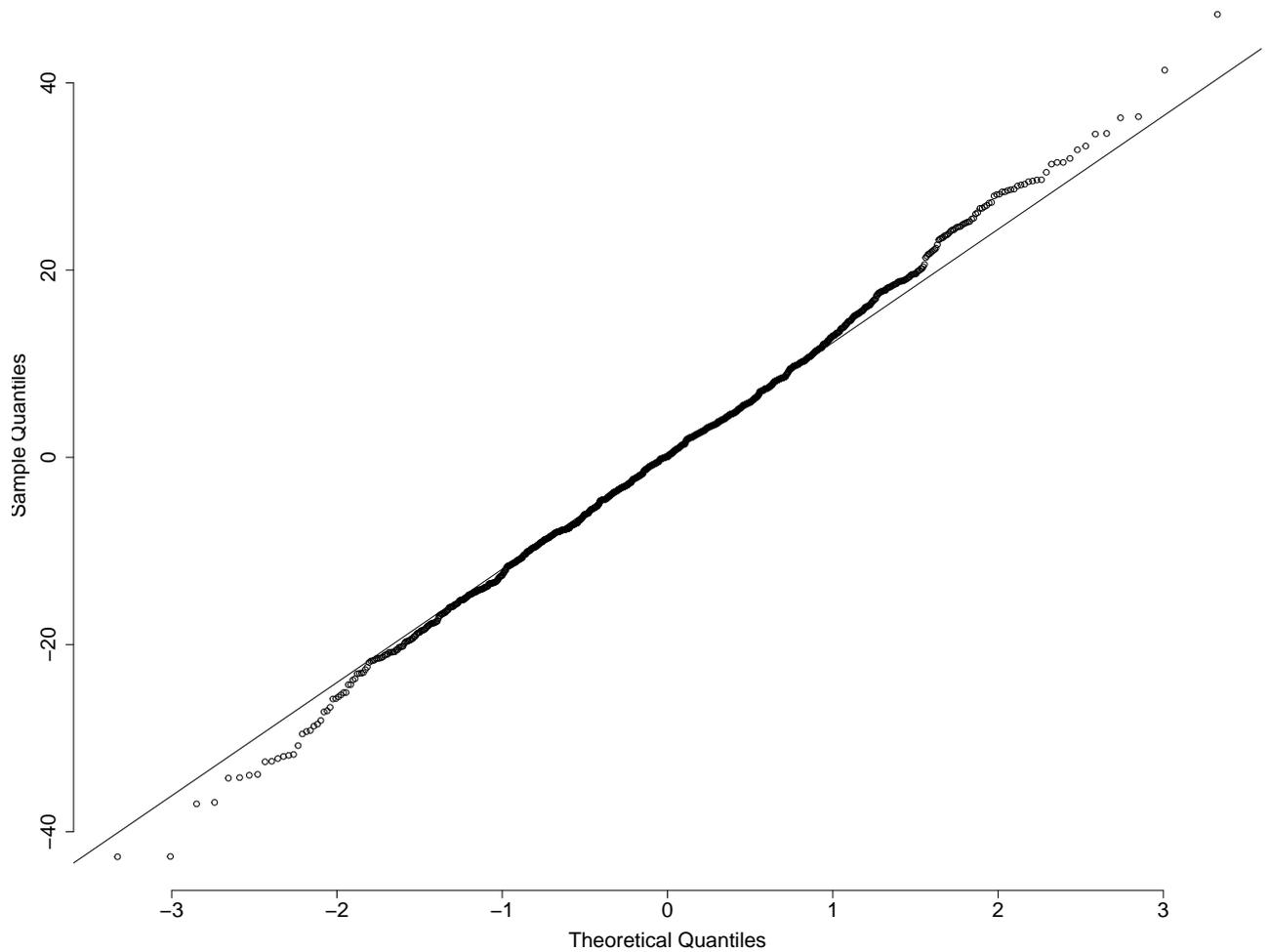


Figura 5.8: Normalidade dos resíduos

A partir do *Q-Q plot*, é possível observar que os valores da cauda não estão perto da linha da distribuição normal. E, além disso, obtivemos uma estatística de teste  $W = 0,9979$  e um  $p - \text{valor} = 0,1678$ . Estes valores obtidos a partir do teste de Shappiro-Wilk nos mostram que nossa hipótese nula é aceita, ou seja, nossos resíduos seguem uma distribuição normal.

### 5.1.3 Previsão

Depois ajustarmos nosso modelo e verificarmos nossos resíduos, agora faremos a previsão da nossa série temporal. Como anteriormente, retiramos as últimas 14 observações da série temporal e calculamos as previsões utilizando um intervalo de 95% de confiança. Na figura 5.9, é possível observar o último mês da série com a previsão para os últimos 14 dias e seus respectivos limites.

Podemos observar que a previsão acompanha o valor verdadeiro das observações. Logo, podemos dizer que o nosso ajuste utilizado foi adequado à nossa série. Portanto, seguiremos para a segunda modelagem a partir do modelo de regressão harmônica.

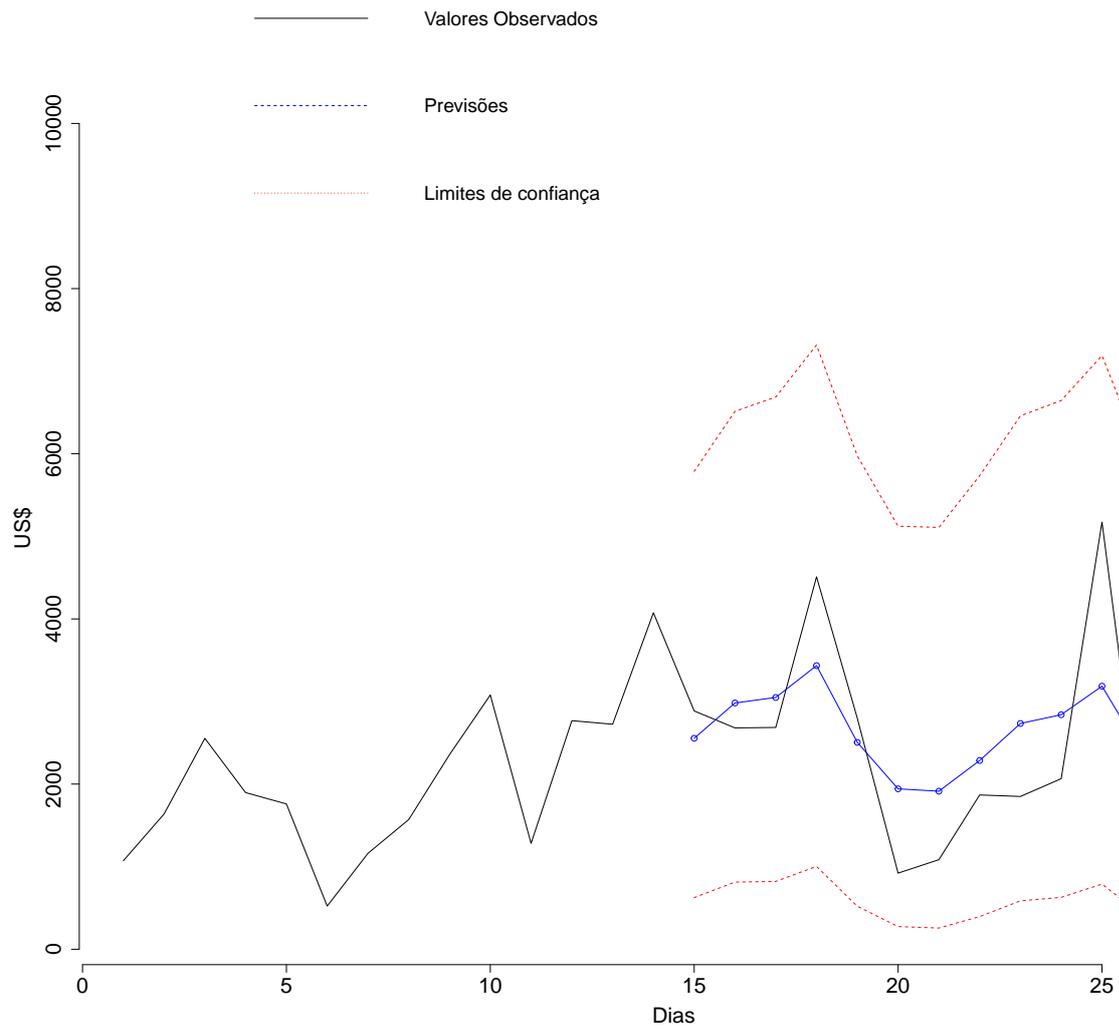


Figura 5.9: Previsão da receita resgatada

## 5.2 Modelo de regressão harmônica aplicado à receita resgatada

Agora vamos explicar nossa série temporal através do modelo de regressão harmônica. Como dito anteriormente, ao aplicarmos este modelo é possível utilizar dois ciclos sazonais (semanal e anual). A partir disso, estamos utilizando a equação 4.2, que foi proposta no capítulo anterior.

### 5.2.1 Identificação do modelo

A partir do modelo proposto acima, mostraremos o mesmo aplicado aos nossos dados:

$$\hat{Z}_t = 58,33 - 5,76\cos(2\pi w_1 t) + 1,17\sin(2\pi w_1 t) - 3,88\cos(2\pi w_2 t) - 3,46\sin(2\pi w_2 t). \quad (5.2)$$

Portanto, o próximo passo é analisar os resíduos deste ajuste e verificar se os mesmos são independentes e identicamente distribuídos. E analisar se eles seguem uma distribuição normal para que assim possamos seguir com a previsão dos dados.

### 5.2.2 Análise dos resíduos

Primeiramente, vamos analisar os resíduos para vermos como eles se comportam, se aparentam ter um comportamento estacionário. Com isso, na figura 5.10 dos resíduos.

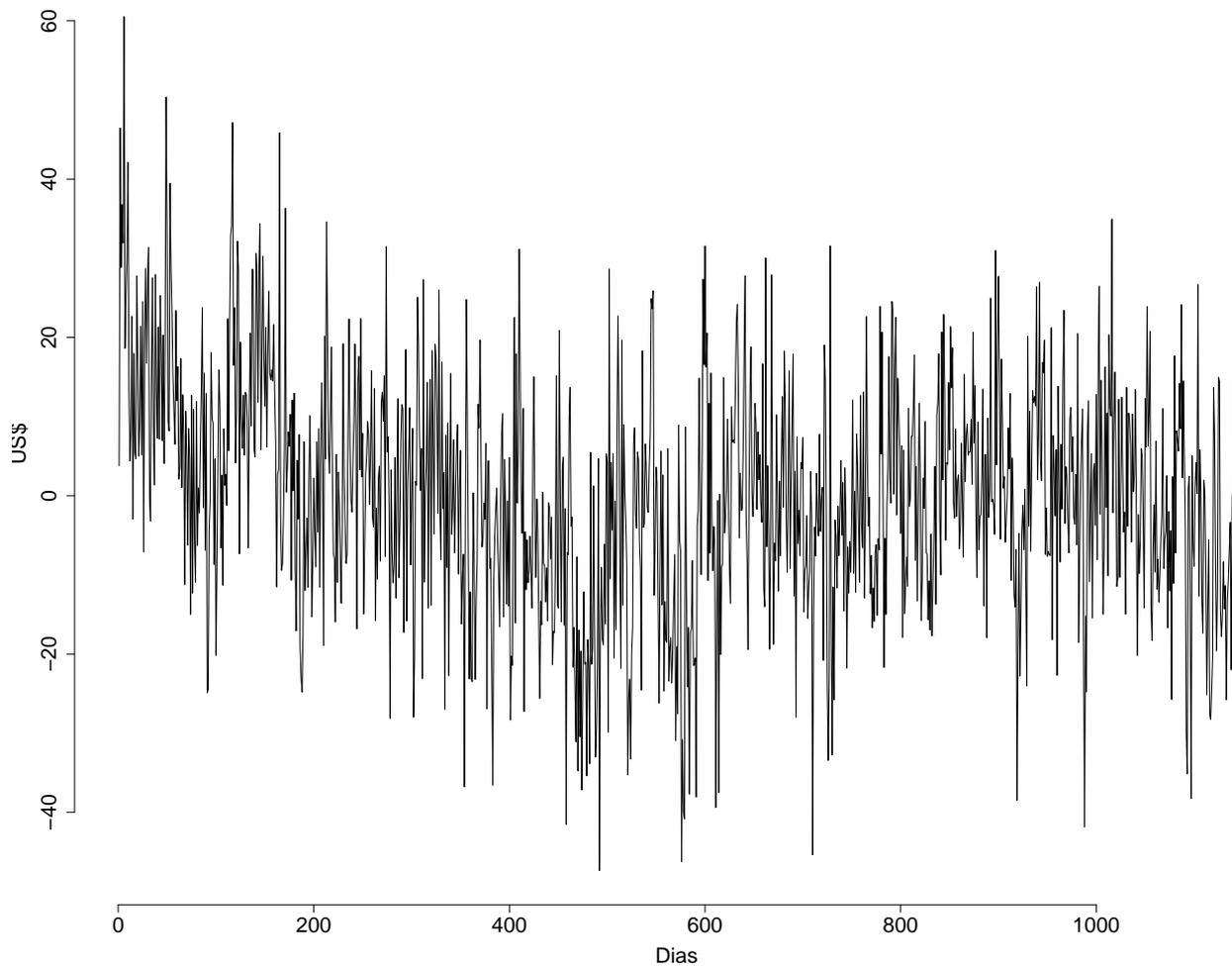


Figura 5.10: Gráfico dos resíduos da receita resgatada

Na sequência, na figura 5.11 temos a ACF e a PACF dos resíduos. Podemos observar que nossos resíduos não tem um comportamento de ruído branco. Vemos na ACF que ainda temos um processo oscilatório nos lags. E, na PACF, é possível ver que elas não estão dentro dos limites. Assim, indicando que nossos resíduos não são independentes. Isto pode ser comprovado ao aplicarmos o teste de Ljung-Box nos resíduos, onde obtemos o  $p$ -valor  $< 2,2e^{-16}$  e o  $\chi^2 = 165$ . Portanto, sendo necessário aplicarmos um outro modelo aos nossos dados.

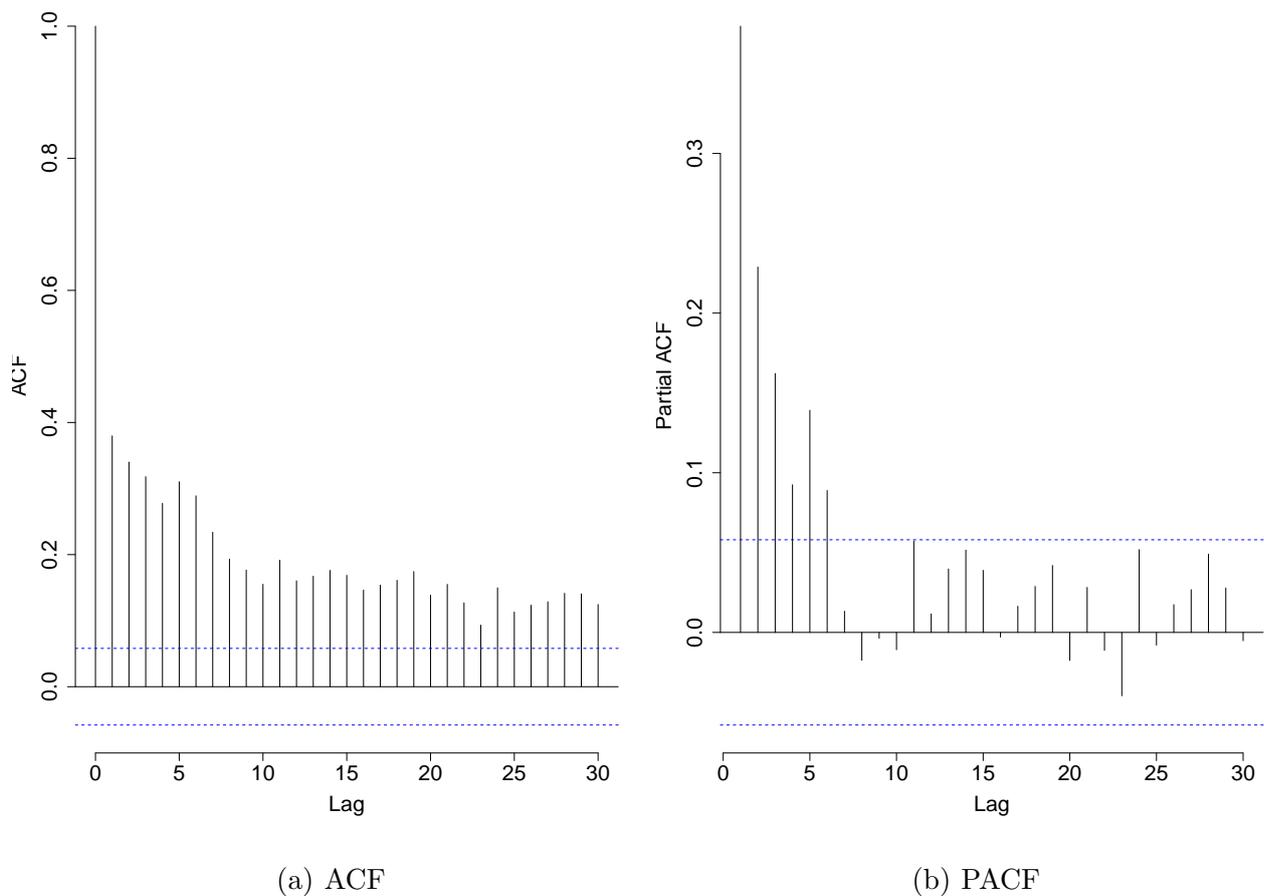


Figura 5.11: ACF e PACF dos resíduos

Observando a figura 5.11, sugerimos o ajuste de ARMA(1,1) para a série temporal dos resíduos juntamente com o modelo de regressão harmônica para os dados. A partir disso, obtivemos um  $AIC = 9109$  e um  $BIC = 9149$ . Assim, abaixo temos a equação do ajuste:

$$Z_t = 0,94Z_{t-1} - 0,75\epsilon_{t-1} + 58,59 - 5,73\cos(2\pi w_1 t) + 1,13\sin(2\pi w_1 t) - 3,34\cos(2\pi w_2 t) - 3,65\sin(2\pi w_2 t). \quad (5.3)$$

Ao ajustarmos os resíduos, esperamos que ele se comporte como um ruído branco. Podemos observar o gráfico dos mesmos na figura 5.12.

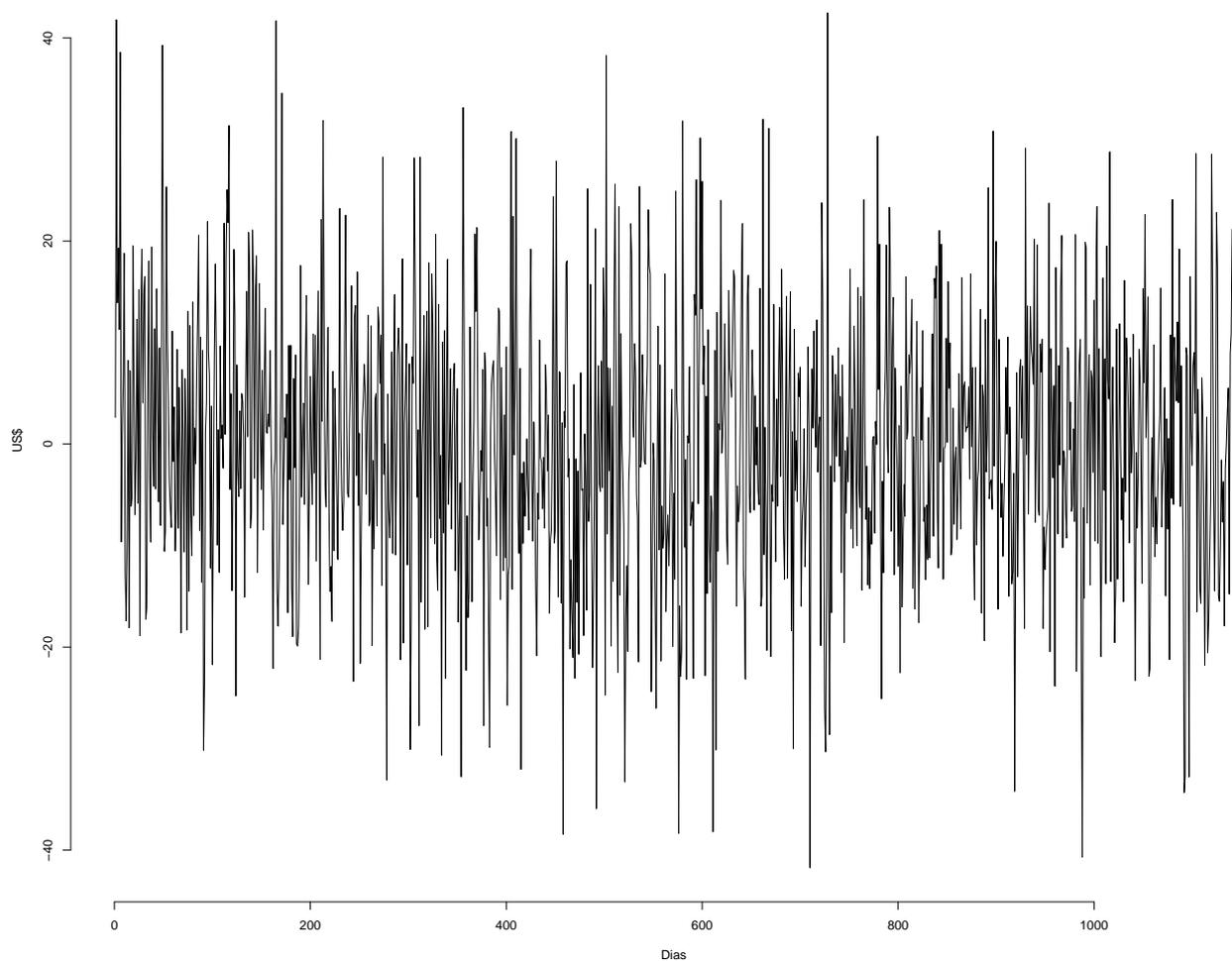


Figura 5.12: Gráfico dos resíduos da receita resgatada

Podemos ver no gráfico 5.12 que os nossos resíduos parecem ser estacionários. Com isso, iremos mostrar a ACF e a PACF dos mesmos na figura 5.13.

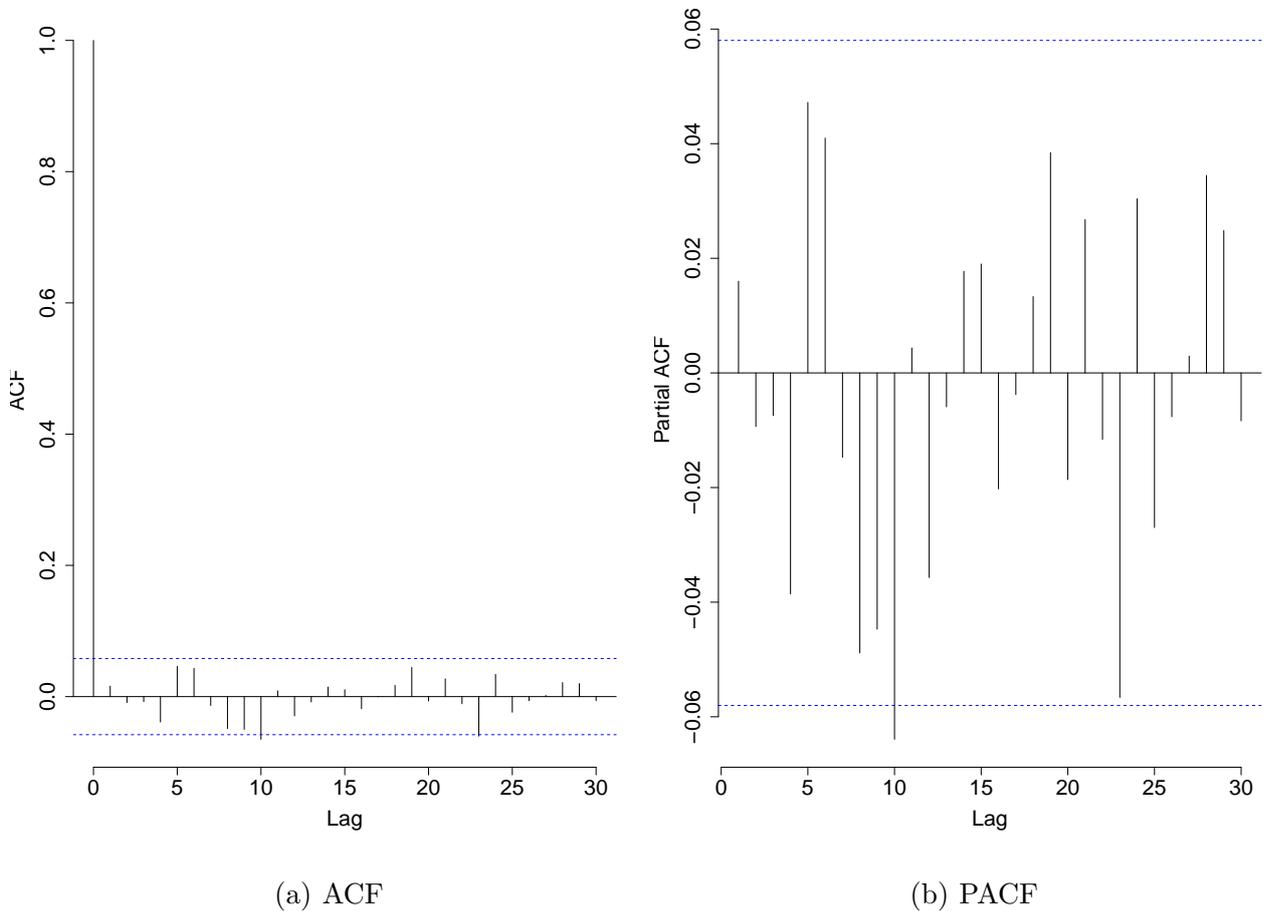


Figura 5.13: ACF e PACF dos resíduos

Aplicamos o teste de Ljung-Box nos resíduos e obtivemos o  $p - valor = 0,5882$  e o  $\chi^2 = 0,2932$ . Esses valores nos mostram que aceitamos a hipótese nula, ou seja, aceitamos que nossos resíduos são independentes.

Portanto, como o nossos resíduos se comportam como ruído branco, o nosso próximo passo é aplicar o teste de Shappiro-Wilk. E, assim, verificar se nossos resíduos seguem uma distribuição normal. Na figura 5.14 temos o gráfico do nosso teste, nele é possível observar que a cauda dos mesmos não estão perto da linha da distribuição normal, o que poderia nos mostrar uma não normalidade dos resíduos.

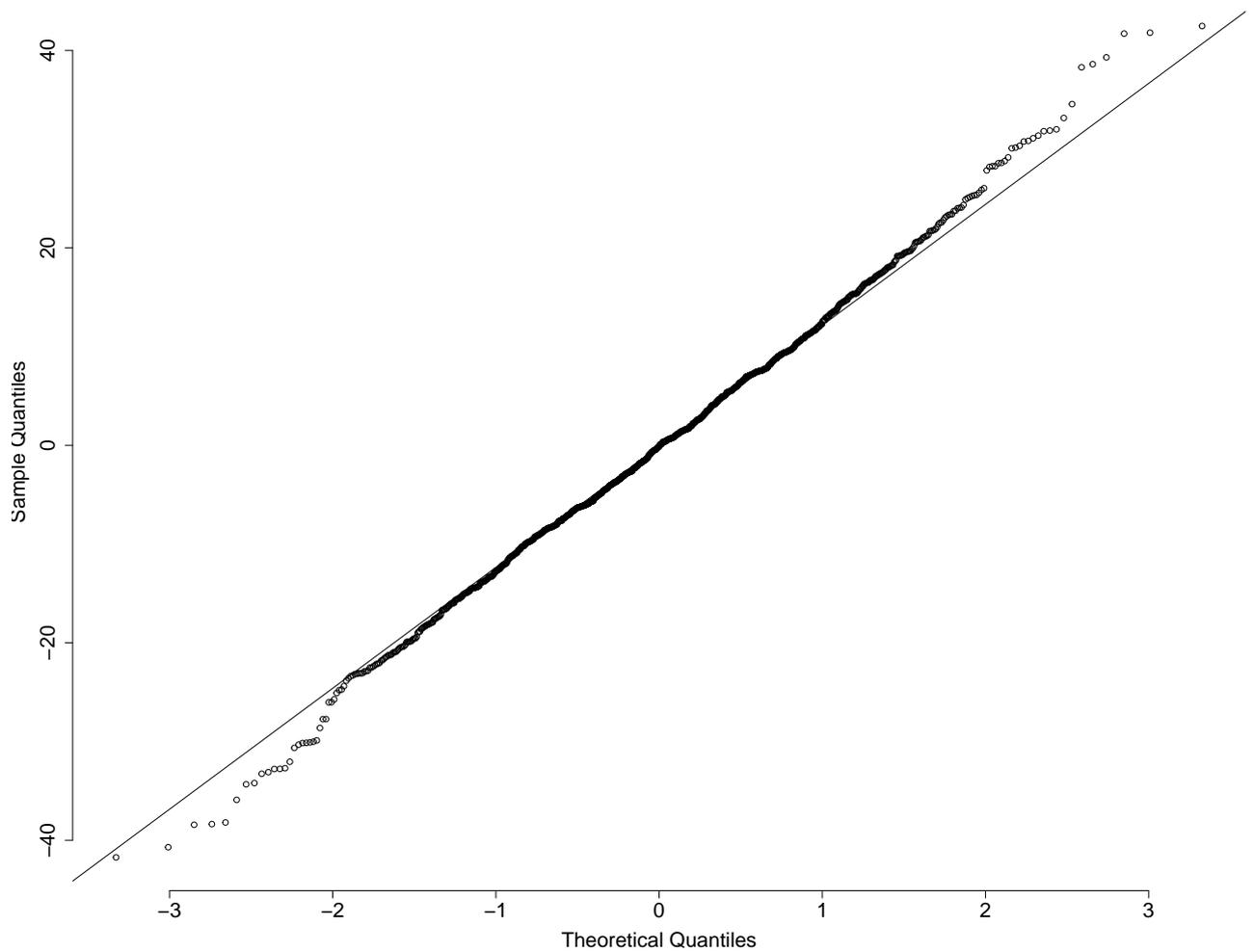


Figura 5.14: Normalidade dos resíduos

Também obtivemos uma estatística de teste  $W = 0,9979$  e um  $p - valor = 0,1546$ . Estes valores nos mostram que não há evidências para rejeitarmos a hipótese nula de normalidade dos resíduos.

Na figura 5.15 é possível observar as sazonalidades estimadas de forma determinística.

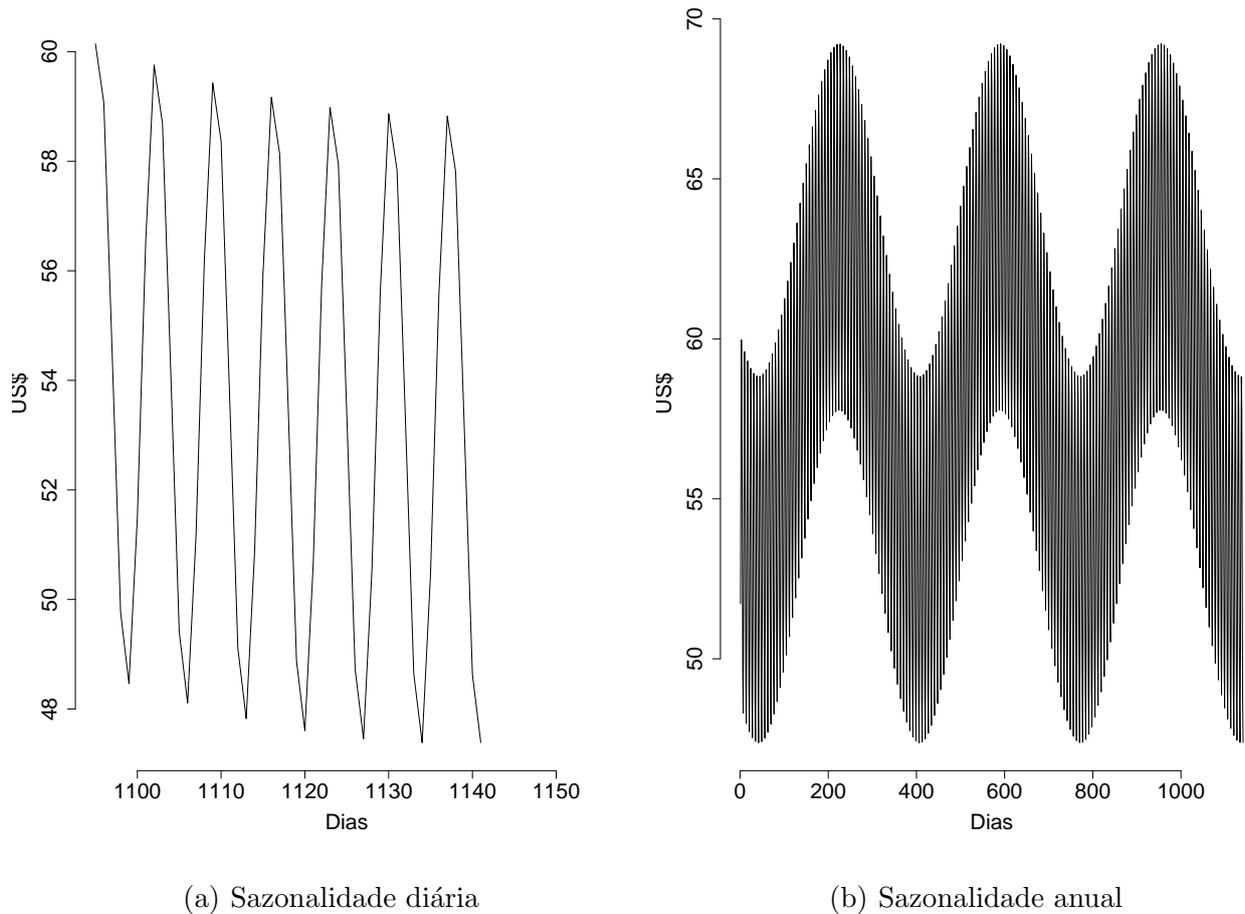


Figura 5.15: Sazonalidade diária e anual estimada

### 5.2.3 Previsão

Por fim, depois de analisarmos e ajustarmos nosso modelo, vamos fazer a previsão da nossa série temporal assim como fizemos com a previsão no modelo SARIMA.

Na figura 5.16 podemos observar, assim como na previsão do modelo SARIMA, os últimos 30 dias da nossa série com a previsão nos últimos 14 dias e seus limites de confiança. Podemos ver que a previsão segue os nossos dados observados. Assim, podemos dizer que esta modelagem também foi adequada à nossa série temporal da receita resgatada.

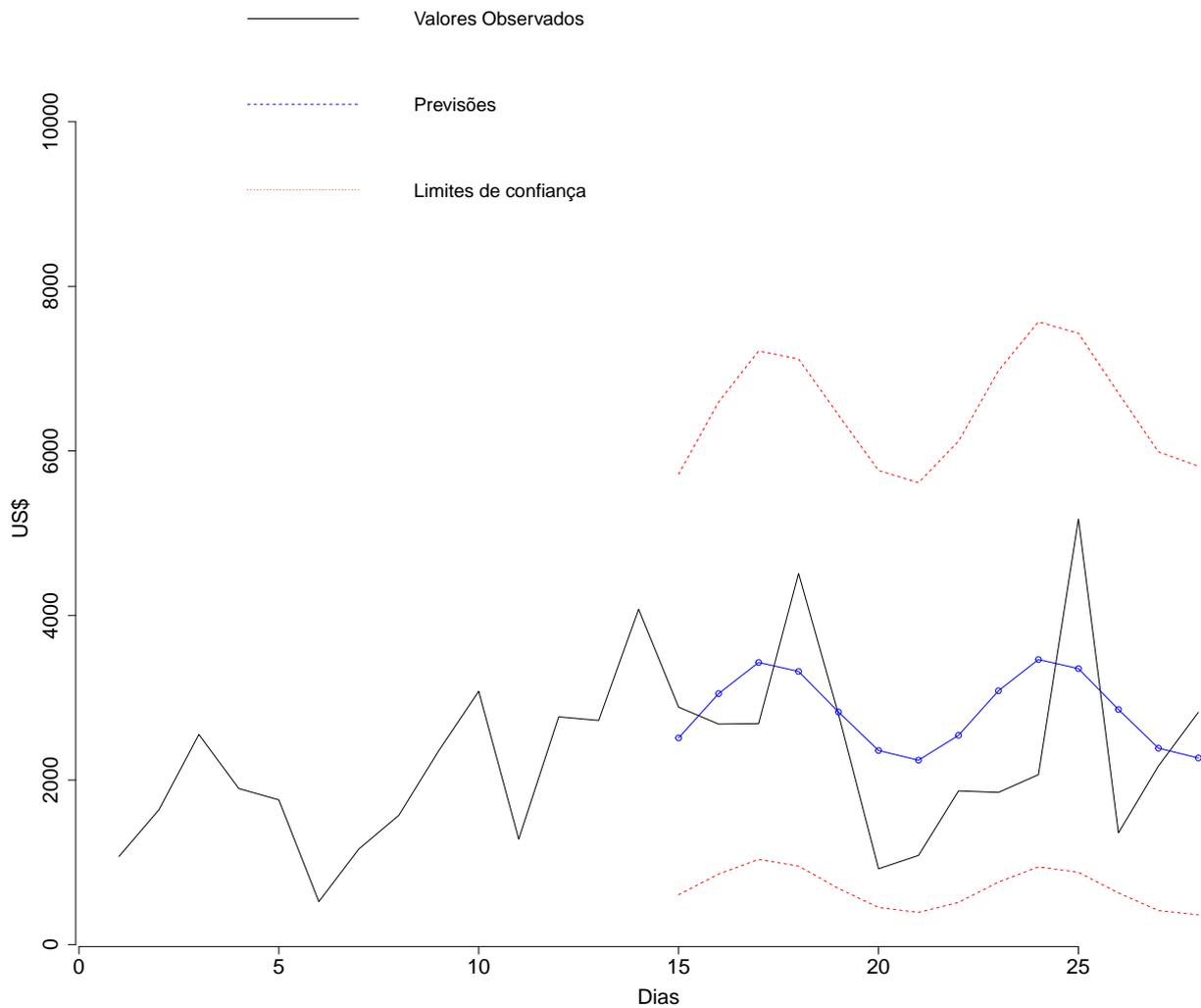


Figura 5.16: Previsão da receita resgatada

# Capítulo 6

## Conclusão

O principal objetivo deste trabalho foi analisar as séries temporais da receita colocada e da receita resgatada de uma empresa do ramo de *travel retail*. Uma vez que estas séries apresentam comportamento sazonal (de ciclo diário e anual), propomos duas abordagens de modelagem: uma via modelos SARIMA, que explica a sazonalidade de forma puramente estocástica; e uma via modelos de regressão harmônica, que permite que a sazonalidade seja tratada de forma determinística e estimada em função de harmônicos. Em ambas foram utilizados os modelos SARIMA e o modelo de regressão harmônica.

Primeiramente, em relação a receita colocada, transformamos os nossos dados tirando a raiz quadrada dos mesmos. E a partir disso vimos que o modelo estocástico que melhor se ajustou aos nossos dados foi o SARIMA(1,1,2)(1,1,1)<sub>7</sub>. Neste modelo foi possível observar o menor valor de AIC e BIC. Depois disso, verificamos através do teste de Ljung-Box que os nossos resíduos eram independentes, ou seja, eram ruído branco. E, através do teste de Shappiro-Wilk, assumimos que os resíduos seguiam uma distribuição normal. Com isso, partimos para a previsão com um intervalo de 95% de confiança. Foi possível observar que a nossa previsão foi adequada aos nossos dados, pois elas estavam próximas dos valores observados.

Ainda sobre a receita colocada, o próximo passo foi ajustar um modelo de regressão harmônica para a nossa série temporal. Utilizamos neste modelo dois harmônicos de

forma a explicar a sazonalidade diária e anual. Ao ajustarmos o modelo, podemos verificar que os nossos resíduos não se comportavam como um ruído branco. Com isso, foi preciso modelar através de um modelo ARMA(1,1) os resíduos para que eles fossem independentes. Em seguida, aplicamos o teste de LJung-Box onde obtivemos um  $p$ -valor pequeno e foi possível comprovar que novos resíduos eram independentes. Por isso, em sequência testamos o Shappiro-Wilk e também assumimos que os nossos resíduos seguiam uma distribuição normal. Portanto, fizemos a previsão da nossa série com um intervalo de 95% de confiança e vimos que a nossa previsão também foi adequada aos nossos dados.

Em segundo lugar, analisamos a receita resgatada. Assim como na receita colocada, tiramos a raiz quadrada dos nossos dados para diminuirmos a escala dos mesmos. E, vimos que o modelo que melhor se ajustou aos nossos dados também foi o SARIMA(1,1,2)(1,1,1)<sub>7</sub>, onde seu AIC e BIC obtiveram valores menores, o que é compreensível pois os valores de resgate dependem dos valores que foram colocados. Os testes de LJung-Box e de Shappiro-Wilk corroboraram com a hipótese de um ruído branco normal para os erros. Em seguida, calculamos a previsão para os últimos 14 dias dos nossos dados com um intervalo de 95% de confiança, e ao compararmos com os dados observados desse mesmo período vimos que os nossos valores estimados foram adequados.

Em seguida, ainda sobre a receita resgatada, ajustamos nossa série com o modelo de regressão harmônica com dois harmônicos de forma a explicar a sazonalidade diária e anual. Ao ajustarmos o modelo e analisarmos os resíduos, foi possível verificar que os mesmos não se comportavam como um ruído branco. Com isso, foi preciso modelá-los através de um modelo proposto ARMA(1,1). E, assim, ao aplicarmos os testes de LJung-Box e Shappiro-Wilk foi possível afirmar que os nossos resíduos eram independentes e assumimos que eles seguiam uma distribuição normal. Por fim, fizemos a previsão da mesma forma que no modelo SARIMA. E, é possível verificar que a mesma foi adequada aos nossos dados.

A partir da análise realizada neste trabalho, podemos concluir que tanto os modelos SARIMA quanto os modelos de regressão harmônica se mostraram adequados para a análise das séries temporais estudadas. Eles permitiram, de maneiras diferentes, que entendêssemos o comportamento das série temporais da receita colocada e da receita resgatada. Além disso, os resultados mostraram que os modelos sob ambas as abordagens fornecem boas previsões, o que faz com que sejam ferramentas úteis para utilização de tomada de decisão das empresas do ramo considerados neste trabalho.

Por fim, como trabalhos futuros, podemos analisar as séries temporais sob o enfoque bayesiano por meio de modelos dinâmicos sazonais [West e Harrison \(1997\)](#), em que permite-se que os parâmetros do modelo variem no tempo de forma dinâmica. A análise sob esta abordagem já foi iniciada, mas os resultados ainda são parciais e por este motivo não foram descritos neste documento.

# Referências Bibliográficas

- Box, G. E. P. e Jenkins, G. M. (1970) *Time series analysis: forecasting and control*. San Francisco: Holden-Day.
- Brockwell, P. J. e Davis, R. A. (2010) *Introduction to time series and Forecasting*. New York: Springer Verlag.
- Cryer, J. D. e Chan, K. S. (2008) *Time Series Analysis - With Applications in R*. New York: Springer Verlag.
- DeGroot, M. H. e Schervish, M. J. (2003) *Probability and Statistics*. London: Pearson.
- Morettin, P. A. e Tolo, C. M. C. (2006) *Análise de Séries Temporais*. São Paulo: Blusher.
- R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- West, M. e Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag, 2nd edn.