**PTL Minas e Escolha de Parâmetros para a função ARIMA no R**

**1 – Análise dos Gráficos das Séries PTL-MG, PLP E ICP Leite.**

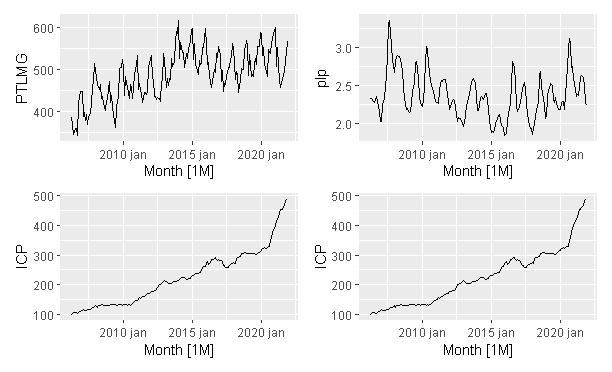


Gráfico do PTL-MG com os anos[[1]](#footnote-1). Observa-se:

* **Tendência** **de crescimento**;
* **Sazonalidade forte** a cada ano;
* Nenhum ciclo muito evidente.

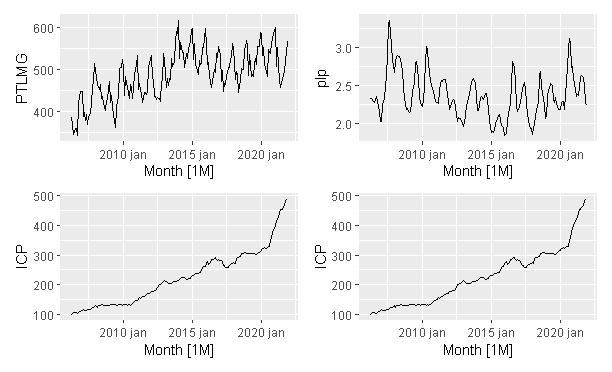


Gráfico do PLP[[2]](#footnote-2) com os anos.

Observa-se:

* Fraca **tendência de** **queda** até 2015, seguida por uma **tendência de crescimento**;
* **Sazonalidade forte** a cada ano.
* Nenhum ciclo evidente.

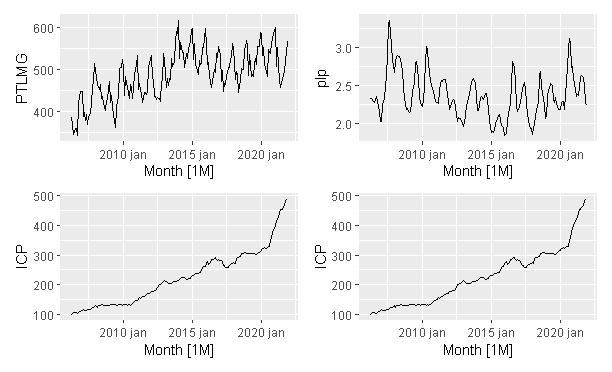


Gráfico do ICP com os anos.

Observa-se:

* **Tendência de crescimento.**
* Nenhum ciclo ou sazonalidade.

**2 – Gráficos de Correlação.**

Abaixo: Imagens com as saídas da função lag1.plot() – aplicada a PTL-MG em função dela própria – e lag2.plot() – PTL-MG em função de ICP e PLP.

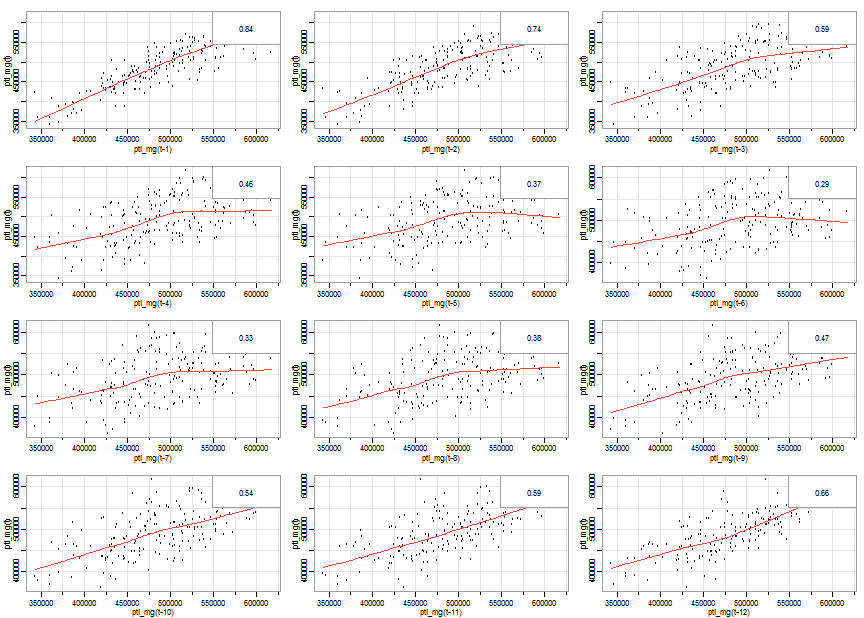


Imagem 1 - Variável PTL MG em função dela própria defasada + correlações.

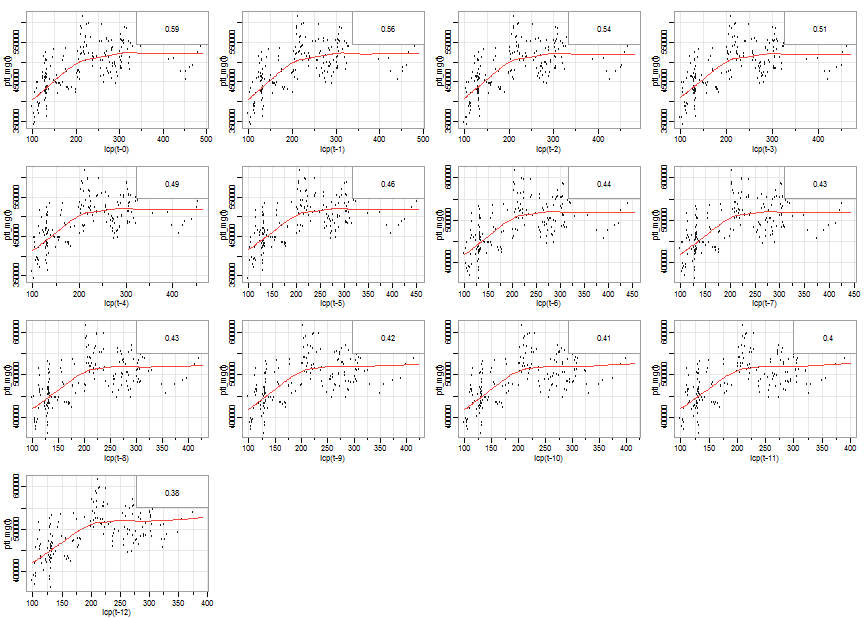


Imagem 2 - Variável PTL MG em função de ICP + correlações.

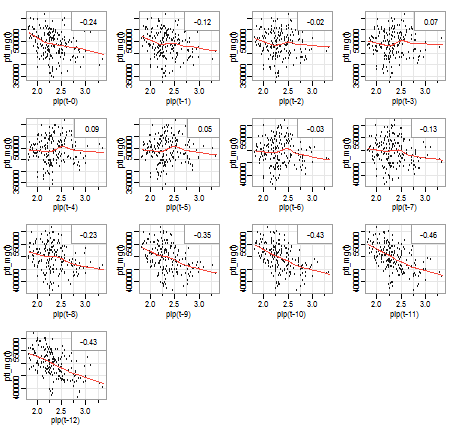


Imagem 3 - Variável PTL MG em função de PLP + correlações.

*Observações*

* As variáveis *preço do leite ao produtor* (PLP) e *índice de custos de produção do leite* (ICP) não apresentaram correlações significativas com o preço trimestral do leite em Minas Gerais (PTL-MG).
* Na *Imagem 1*, onde apresentam-se gráficos de dispersão (*scatterplot*) da variável PTL-MG *versus* ela própria defasada, os que se mostraram mais significativos foram o primeiro (lag = 1) e o último (lag = 12), respectivamente. Logo, faremos um modelo ARIMA utilizando apenas a PTL-MG inicialmente.

**3 – ARIMA(p,d,q)(P,D,Q)[s]**

O modelo utilizado inicialmente, na tentativa de prever a variável PTL-MG em três meses, é o **ARIMA(p, d, q)(P,D,Q)[s]**, onde:

* p – ordem da autoregressão (defasagem);
* d – ordem de diferenciação;
* q – média móvel;
* P, D, Q – os mesmos parâmetros anteriores, só que **sazonais**.
* s – número de períodos por ano ou sazonalidade (ex: se a série for mensal, s = 12);

O método de previsão ARIMA é baseado na ideia de que a informação dos valores passados de uma série temporal podem, por si próprios, serem utilizados para prever os valores futuros da mesma série. Qualquer série temporal não-sazonal que exibe padrões e não é considerada *white noise* pode ser modelada com ARIMA.

**4 – Pressupostos para Estacionariedade**

Antes de aplicarmos o ARIMA à nossa série temporal, devemos verificar se a série é **estacionária** (não apresenta padrões como tendência, ciclo e sazonalidade). Caso não seja, devemos transformá-la.

Como sabemos, PTL-MG é não-estacionária pois apresenta forte sazonalidade e tendência de crescimento bem evidente (como visto na seção 1).

**4.1 – Diferenciação (parâmetros d / D)**

A abordagem mais comum para tornar uma série não-estacionária em estacionária consiste em **diferenciá-la**, isto é, subtrair de cada uma de suas observações a observação imediatamente anterior. Foi o que foi feito com PTL-MG, obtendo-se, para essa variável, os valores contidos na *tabela 1* (as observações após a décima foram omitidas).



Tabela 1 – Valores de PTLMG, PTLMG\_diff [[3]](#footnote-3), ICP e PLP para as dez primeiras observações.

O valor escolhido para o parâmetro *d* deverá ser o número de diferenciações necessárias para que a série vire estacionária. Para PTL-MG, apenas uma diferenciação simples e outra sazonal parecem ser suficientes - ou seja, *d* = 1 e D = 1. A escolha desses valores pode ser confirmada com as funções *unitroot\_ndiffs()* e *unitroot\_nsdiffs()*.

Abaixo, gráficos de ACF e PACF para PTL-MG antes (esquerda) e após (direita) a primeira diferenciação.

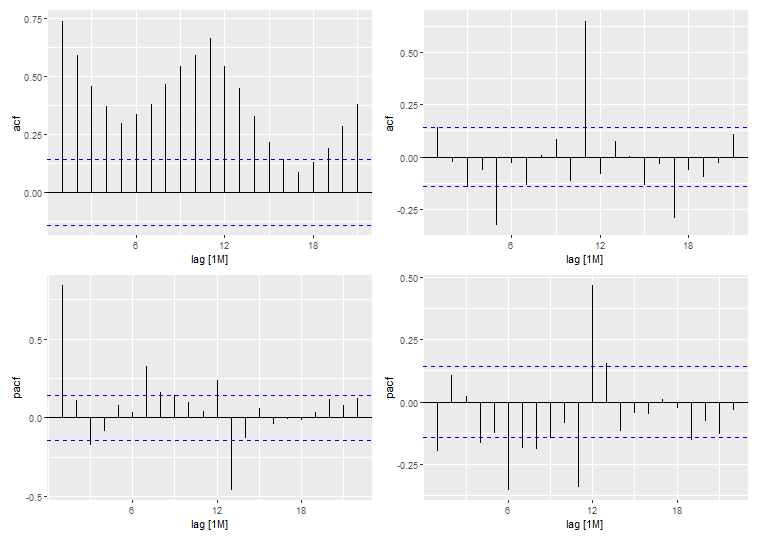


Imagem 4 - Gráficos ACF e PACF para PTL-MG e PTLMG\_diff.

Ainda é possível perceber um viés de sazonalidade nos gráficos apresentados na Imagem 4, logo, aplica-se a diferenciação sazonal, obtendo-se os seguintes resultados para ACF e PACF:

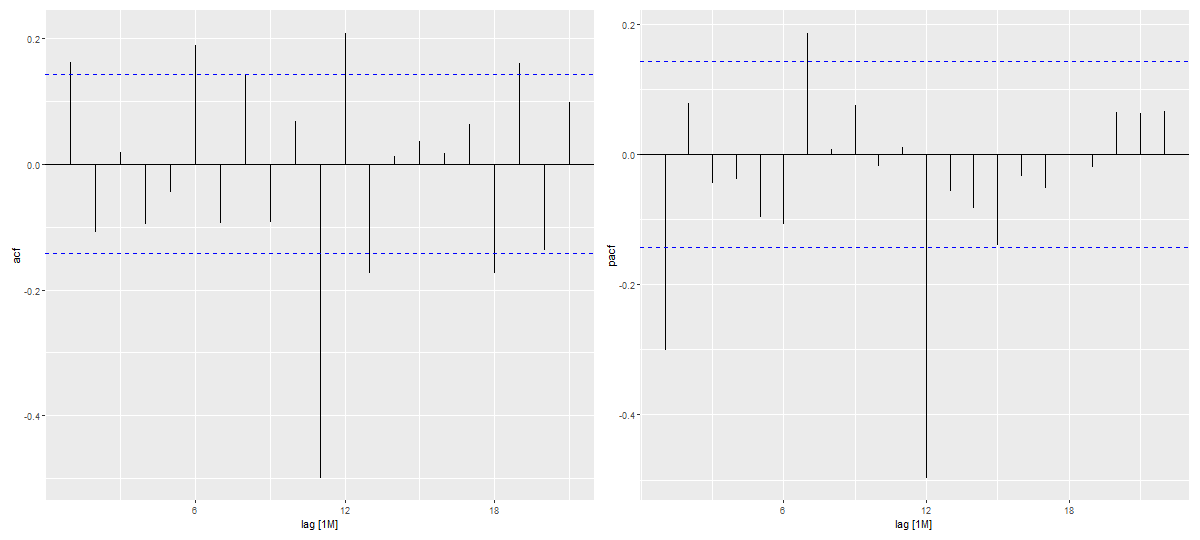


Imagem 5 - Gráficos ACF e PACF para PTLMG\_diff.

Em seguida, gráficos da função lag1.plot() aplicada a PTLMG\_diff (PTLMG diferenciada de forma simples e sazonal). Se a autocorrelação for positiva para um número alto de defasagens (até lag = 10 ou mais), então a série precisa de mais diferenciações. Caso contrário, se as autocorrelações forem muito negativas, então provavelmente foram feitas mais diferenciações que o necessário.

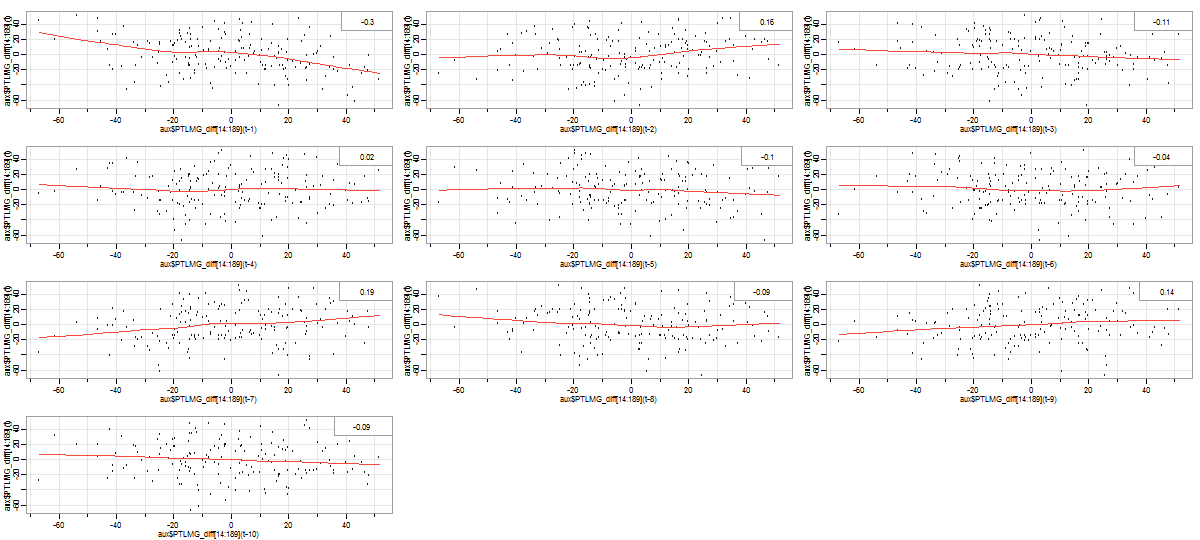


Imagem 6 - lag1.plot() aplicada à PTLMG\_diff (PTLMG com d =1 e D = 1).

Considerando a imagem 6, é possível perceber que 4 das 10 *lags* apresentam autocorrelações positivas. Além disso, considerando-se as autocorrelações negativas, o maior módulo obtido foi 0,3. Logo, admitimos que a série não precisa de mais diferenciações e nem foi diferenciada além do necessário.

Mas será que a série agora é, de fato, estacionária? Ao checar essa hipótese com a função que realiza o teste Augmented Dickey Fuller, obtemos p-valor = 0,01. Como a estacionariedade é atribuída à hipótese alternativa do teste, rejeitamos a hipótese nula – o que significa que não temos evidências fortes o suficiente para afirmar que a série não seja estacionária.

**4.2 – Defasagem (parâmetros p / P)**

O próximo passo é identificar se o modelo precisa de algum grau *p* de defasagem. Uma forma de verificar isso é por meio da análise do gráfico de autocorrelação parcial (PACF *plot*s) da série estacionária – inicialmente, tomamos a ordem de *p* como igual ao número de *lags* que ultrapassam o limite de significância no gráfico PACF. Para a nossa variável, tal gráfico pode ser visualizado na Imagem 5. Observa-se que, em três lags, a autocorrelação calculada ultrapassou a margem de significância (lags 1, 7 e 12). Mas podemos ser conservadores e adotar p = 1. O mesmo procedimento pode ser feito para achar P se levarmos em consideração as lags dos períodos que são múltiplos da sazonalidade (provavelmente, olhar até a terceira *lag* será suficiente). No segundo gráfico da imagem 5, vemos que a lag = 12 ultrapassou em muito o limite de significância. Logo, podemos adotar P = 1.

**4.3 – Média Móvel (parâmetros q / Q)**

Da mesma forma que olhamos o gráfico PACF para identificar o valor de p/P, podemos olhar o gráfico ACF para determinar os parâmetros q/Q. Nisso, verificamos que 5 *lags* ultrapassam os limites de significância, mas adotarei q = 1. Além disso, como *lag* = 12 também ultrapassa o limite, adotarei Q = 1.

**5 - Modelo ARIMA(1,1,1)(1,1,1)[12]**

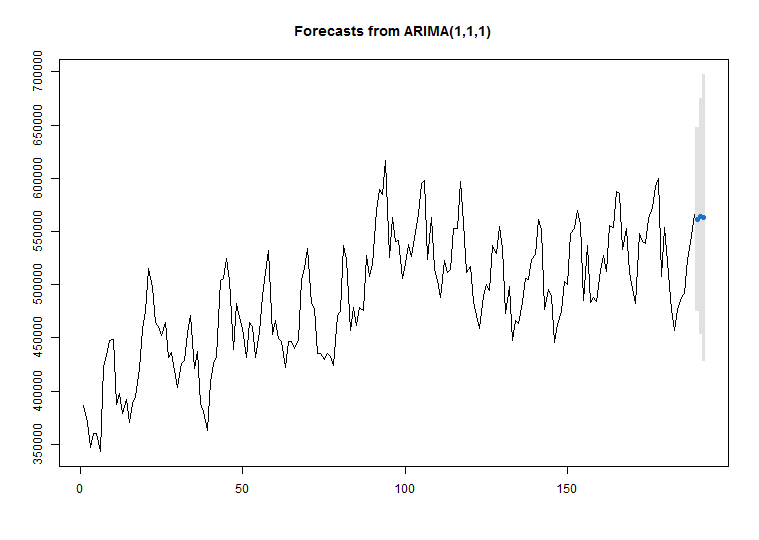


Imagem 7 – Projeção em 3 meses para PTLMG com Arima(1,1,1)(1,1,1)[12].



Imagem 8 - Resultados da previsão para três meses de PTLMG com Arima(1,1,1)(1,1,1)[12].

**6 – Resíduos de ARIMA(1,1,1)(1,1,1)[12]**

testar para media movel q = 2 e Q = 12 (?) (recomendado para dados mensais com sazonalidade anual).

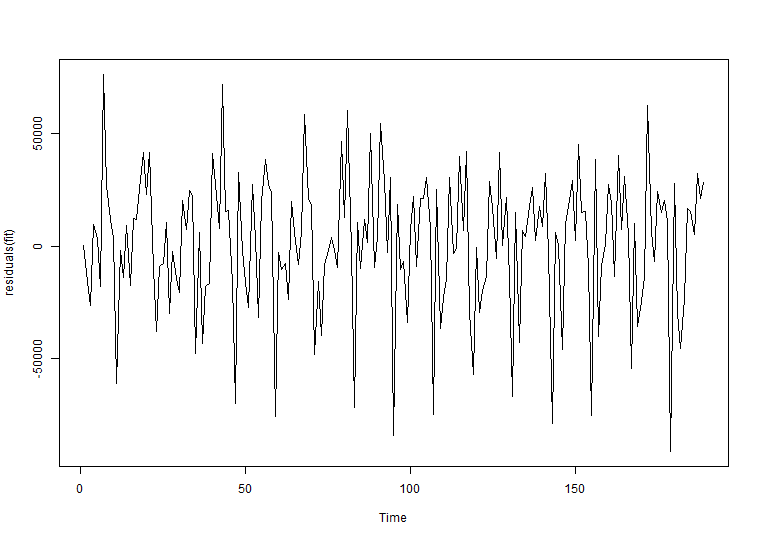


Imagem 9 - Gráfico dos Resíduos do modelo Arima(1,1,1)(1,1,1)[12] ajustado à PTLMG.

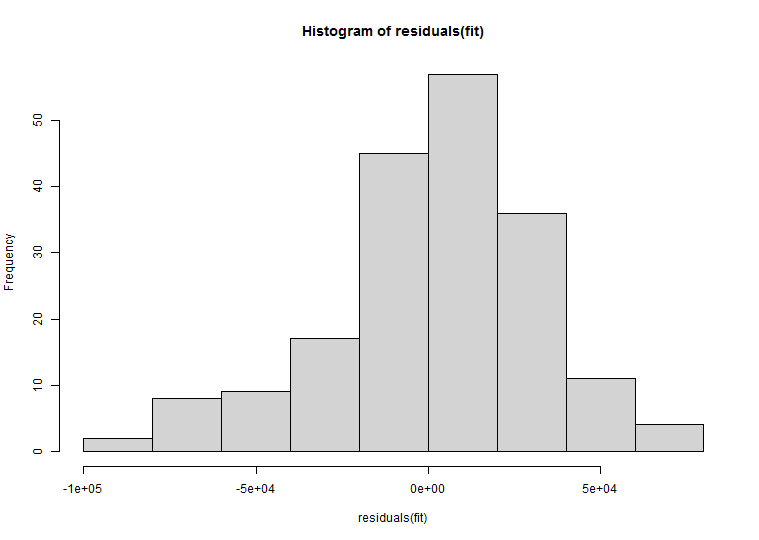


Imagem 10 - Histograma dos Resíduos do modelo Arima(1,1,1)(1,1,1)[12] ajustado à PTLMG.

Na imagem 9 (gráfico dos resíduos de Arima(1,1,1)(1,1,1)[12]), podemos perceber que os resíduos não apresentam nenhum padrão significativo e estão bem espalhados – conforme o esperado. Além disso, na imagem 10 (histograma dos resíduos), temos que esses dados seguem uma distribuição que lembra a normal – outro pressuposto que deve ser atendido para os resíduos.

**7 – Verificação da Precisão do Modelo**

a ser feito.

1. Os dados utilizados são referentes ao período compreendido entre abril de 2006 e dezembro de 2021. [↑](#footnote-ref-1)
2. PLP – Preço do leite ao produtor nacional. [↑](#footnote-ref-2)
3. PTLMG\_diff é a variável PTLMG diferenciada. A fim de não perder a variável PTLMG com suas características originais, cria-se essa variável auxiliar para realizar o processo de tornar a série estacionária (e armazená-la). [↑](#footnote-ref-3)