

## Modelagem simultânea da média e da dispersão da produção nacional de leite<sup>1</sup>

João Pedro Junqueira Schettino <sup>2</sup>  
Glauco Rodrigues Carvalho <sup>3,4</sup>

**Resumo:** O Brasil é um dos maiores produtores de leite do mundo, esse trabalho estuda a dinâmica da oferta de leite, e como ela se associa com o preço pago ao produtor, os custos de produção, a evolução do setor e suas características peculiares de sazonalidade. Estudos de autocorrelação, correlação defasada entre as variáveis são realizados. A componente da sazonalidade é estimada a partir do método Loss, e, em seguida, modelada por um modelo linear generalizado duplo. As previsões realizadas pelo modelo foram bastante satisfatórias, sendo capaz de realizar boas previsões mesmo em momentos de incerteza

**Palavras-chave:** Loss, modelo linear generalizado duplo, oferta de leite, previsão, PTL

**Simultaneous modeling of the average and dispersion of national milk production.**

**Abstract:** Brazil is one of the largest milk producers in the world, this work studies the dynamics of milk supply, and how it is associated with the farm gate milk price, the production cost, the evolution of the sector, and its peculiar characteristics of seasonality. Autocorrelation studies and lame correlation between the variables are performed. The seasonality component is estimated from the Loss method, and then modeled by a double generalized linear model. The predictions made by the model quite satisfactory, and this was able to carry out assertive predictions even in times of uncertainty

**Keywords:** forecast, generalized linear model, Loss, milk supply, PTL

### Introdução

A pesquisa trimestral do leite (PTL), realizada pelo IBGE, fornece dados mensais que dizem respeito ao volume de leite adquirido por laticínios (IBGE,2022). Entretanto, o tempo de divulgação dos resultados muitas vezes impossibilita a utilização da informação na tomada de decisões, já que as informações possuem uma defasagem temporal de 3 meses. A inexistência de dados atualizados em tempo real sobre a oferta de leite no Brasil gera dificuldades para os tomadores de decisão na cadeia produtiva do leite. As informações divulgadas por empresas públicas ou privadas possuem defasagem temporal, gerando incertezas sobre o volume de leite que está sendo ofertado no País. Neste contexto, o objetivo do trabalho foi definir um modelo estatístico capaz de antever a quantidade de leite captada pela indústria de laticínios nos próximos meses, bem como estimar a quantidade já captada, mas ainda não reportada pela PTL. Neste sentido, a pesquisa tomou um caráter teórico, no que tange a compreensão dos possíveis mecanismos associados a variações da oferta de leite, e um caráter técnico e prático, no que tange à procura do método estatístico ideal para a modelagem.

<sup>1</sup> O presente trabalho foi realizado com o apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil.

<sup>2</sup> Graduando em Estatística – UFJF/Juiz de Fora. e-mail: schettino.joaopedro@gmail.com

<sup>3</sup> Economista, Ph.D. Economia Agrícola, Pesquisador da Embrapa Gado de Leite

<sup>4</sup> Orientador

## Material e Métodos

Como enunciado por COWLING & GARDNER (1964), a resposta da oferta de leite pode ser dividida em duas facetas. Variações na produtividade por vaca e variações no número de vacas leiteiras. A produtividade por vaca, por sua vez, poderia ser destrinchada nos seguintes fatores:

- Variações no custo da ração.
- Variação na composição do rebanho nacional.
- Desenvolvimentos técnico científicos, bem como a evolução genética do rebanho.

O número de vacas, por sua vez, poderia ser caracterizado por:

- Variação no consumo de novilhas.
- Taxa de abate de vacas.

Somado a isso, NERLOVE (1979) caracteriza ainda outros fatores influentes, em casos de países em desenvolvimento, uma vez que neste caso, o desenvolvimento do setor agropecuário, o investimento público em infraestrutura e o desenvolvimento de mercados é mais acentuado. Assinala ainda que nesses casos, o preço pago ao produtor se torna um fator ainda mais relevante, no que tange a determinação da oferta de leite e influenciando tanto na produtividade por vaca quanto no número total de vacas leiteiras. Neste contexto, CARVALHO (2015) modela esses dois aspectos individualmente. O número de vacas é associado ao número de vacas no ano anterior e a probabilidade de estas produzirem leite, dado o preço do leite. A produtividade de leite por vaca, por sua vez, é determinada pela tendência temporal de desenvolvimento do setor, e por custos associados à produção de leite. A produção total de leite seria dada pelo número de vacas multiplicado pela produtividade por vaca. No caso deste trabalho, seguiu-se os mesmos princípios, mas com uma abordagem distinta. A modelagem dos dados foi feita levando em consideração o total de leite ofertado por mês. Dessa forma, nossa variável dependente segue a distribuição de probabilidade do produto entre o número de vacas e sua produtividade média. A produção total de leite em determinado mês, a partir do modelo apresentado, estaria associada aos seguintes fatores:

- Uma tendência temporal, que incorpora o desenvolvimento do setor no longo prazo, no que tange a variação da produtividade das vacas, e o tamanho do rebanho.
- A produção de leite no mesmo mês do ano anterior, incorporando a sazonalidade da oferta de leite.
- O índice nominal de custo de produção do leite (ICP) com defasagem de 3 meses,
- O preço pago ao produtor de leite, deflacionado pelo ICP, com defasagem de 3 meses.
- PIB real trimestral, com defasagem de 3 meses, representando a saúde da economia do país.

Após definir as defasagens ideais, com base na literatura Carvalho (2015) analisa as correlações e faz comparações entre distintos modelos estimados. Neste caso, o comportamento do resíduo ainda indicou heterocedasticidade, pois seu desvio variou com a tendência de crescimento do setor. Sabe-se que a sazonalidade da produção de leite é um fenômeno relevante para toda cadeia produtiva. A queda do volume de leite gerado no período de entressafra diminui a receita do produtor, que, por sua vez, vê o custo de produção aumentar, como descrito por Junqueira (2013). Entretanto, com o avanço dos

sistemas de produção em confinamento, como free stall e compost barn, existe uma tendência de diminuição do efeito da sazonalidade, uma vez que nesses sistemas a vaca especializada produz mais nos meses de maior frio, como descrito por Napolitano (2021). Sendo assim, com auxílio de um método robusto de decomposição de séries temporais por regressões polinomiais locais (Wen *et al.*, 2019) foi possível estudar esse fenômeno da mudança da variabilidade sazonal do volume de leite adquirido pela indústria. Para modelar a relação entre as variáveis explicativas de interesse, enunciadas anteriormente, e a variável resposta (Volume total de leite adquirido pela indústria), foi utilizado um Modelo Linear Generalizado Duplo (SMYTH, 1999). Esta abordagem é utilizada nos casos em que o erro idiossincrático do modelo, e, conseqüentemente, a variável dependente, não só não apresentam aspectos característicos de uma distribuição gaussiana, como são caracterizadas por sua heterocedasticidade.

Este modelo pode ser definido da seguinte forma:

$$Y_t \sim G(\mu_t, \phi^{-1}_t) \quad (1)$$

No nosso caso, assumimos que  $Y_t$ , o volume de leite adquirido pela indústria no ano t, segue uma distribuição gama, onde  $\mu_t$  e  $\phi^{-1}$  são, respectivamente, a média e dispersão da distribuição de  $Y_t$  no ano t. E podem ser modeladas separadamente, como segue:

$$\hat{\mu}_t = \beta_0 + \beta_1 Pre_{t-3} + \beta_2 ICP_{t-3} + \beta_3 Y_{t-12} + \beta_4 T + \beta_5 PIB_{t-3} + \varepsilon_t \quad (2)$$

$$\log(\widehat{\phi^{-1}_t}) = \gamma_0 + \gamma_1 T + \tau_t \quad (3)$$

Onde  $Pre_{t-3}$  é o preço pago ao produtor, deflacionado pelo ICP, com 3 meses de defasagem.

$ICP_{t-3}$  é o índice de custo de produção ICP, também com 3 meses de defasagem.  $Y_{t-12}$  é o volume de leite adquirido pela indústria no mesmo mês no ano anterior, nossa componente sazonal. T é nossa variável de tendência temporal, presente em ambas as equações. E,  $PIB_{t-3}$  é a variação real do PIB trimestral no trimestre anterior. Dessa forma, o modelo para a média é caracterizado pela equação (2) e o modelo para a dispersão no entorno da média, pela equação (3). Esse método nos possibilita gerar intervalos de confiança mais estreitos, uma vez que a dispersão da variável de interesse também é estimada.

## Resultados e Discussão

Utilizando o procedimento de decomposição de séries temporais mencionado anteriormente, foi possível analisar e comparar a amplitude da sazonalidade da produção de leite no Brasil com os estados de MG, RS, GO, SC e PR no tempo, identificando que, de fato, a componente sazonal tem se tornado cada vez menos significativa no tempo, apesar de ainda sofrer resistência em alguns estados (Figura 1). Isto acaba influenciando na variância de nossa variável de interesse no tempo, uma vez que a amplitude de uma amostra é um estimador de sua variância populacional.

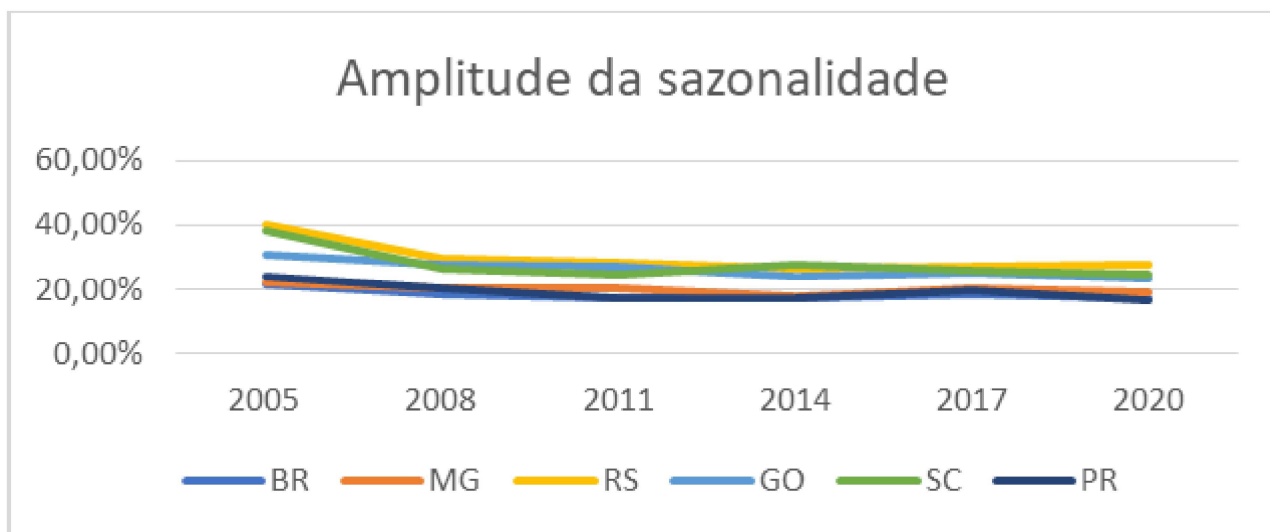


Figura 1. – Amplitude da sazonalidade no tempo.

Fonte: IBGE (2022). Elaboração dos autores

Neste contexto, o modelo é estimado como descrito na seção de materiais e métodos. Seus coeficientes são apresentados na Tabela 1.

Tabela 1. Coeficientes estimados para o modelo

Coefficiente	Estimativa	Erro padrão	Estatística t	Pr(> t )
$\beta_0$	252.103,264	65450,303	3,852	0,000
$\beta_1$	143.726,609	17496,539	8,215	0,000
$\beta_2$	-1.662,023	177,654	-9,355	0,000
$\beta_3$	0,763	0,035	21,775	0,000
$\beta_4$	3.104,038	361,912	8,577	0,000
$\beta_5$	3.052,337	1217,269	2,508	0,013
$\gamma_0$	-5,649	0,311	-18,184	0,000
$\gamma_1$	-0,010	0,003	-3,326	0,001

Fonte: IBGE (2022). Elaboração dos autores

As previsões para o ano de 2021 e primeiro trimestre de 2022, em comparação com os valores reais, são mostradas na Tabela 2.

Tabela 2. Previsões para além da amostra

DATA	Volume previsto (Milhões de litros)	Volume real (Milhões de litros)	Erro percentual	DATA	Volume previsto (Milhões de litros)	Volume real (Milhões de litros)	Erro percentual
jan/21	2.361.986	2.348.087	-0,59%	jan/22	2.156.735	2.075.711	-3,90%
fev/21	2.103.644	2.050.830	-2,58%	fev/22	1.891.149	1.868.943	-1,19%
mar/21	2.115.607	2.176.123	2,78%	mar/22	1.977.000	1.939.145	-1,95%
abr/21	1.974.814	1.945.779	-1,49%	abr/22	1.767.885	-	-
mai/21	1.924.444	1.960.000	1,81%	mai/22	1.773.578	-	-
jun/21	1.896.628	1.932.339	1,85%	jun/22	1.736.690	-	-
jul/21	2.090.852	2.039.552	-2,52%	jul/22	1.834.905	-	-
ago/21	2.111.668	2.087.985	-1,13%	ago/22	1.903.695	-	-
set/21	2.089.068	2.078.884	-0,49%	set/22	1.948.383	-	-
out/21	2.123.517	2.130.461	0,33%	out/22	2.068.177	-	-
nov/21	2.102.831	2.135.312	1,52%	nov/22	2.124.400	-	-
dez/21	2.194.446	2.193.986	-0,02%	dez/22	2.181.023	-	-

Fonte: IBGE (2022). Elaboração dos autores

Para o ano de 2021, o erro percentual médio foi de -0,04% e o erro absoluto percentual médio, 1,43%. O modelo apresentado neste trabalho realizou as melhores previsões, quando comparado com os outros modelos testados, no que tange ao volume total anual de leite, como indicado pelo erro percentual médio. Entretanto, no que se refere a acurácia mês a mês, o melhor modelo não levou em consideração o PIB trimestral, realizando previsões para 2021 com erro absoluto percentual médio de 1,19%, mas com erro médio percentual maior, de -0,44%.

## Conclusões

Após extensa investigação, foi possível estabelecer um modelo capaz de realizar previsões assertivas do volume de leite adquirido pela indústria. Ademais, pode-se concluir que, modelar a dispersão da oferta de leite como função da tendência de crescimento do setor se mostrou como uma importante ferramenta, reduzindo o intervalo de confiança das previsões realizadas pelo modelo. O comportamento sazonal, bem como eventuais outliers continua sendo um empecilho para o aperfeiçoamento do modelo. Modelos que contenham variáveis climáticas podem auxiliar nestes dois quesitos, e devem ser investigados em estudos futuros.

## Referências

CARVALHO, G. R. Brazilian Dairy Model: Impacts of agricultural policies and exogenous shocks. 2015. Dissertação (Doutorado). College Station: Texas A & M University, 2015. 147 p.

COWLING, K., GARDNER, T. W. Milk Supply Response: An Interbreed Analysis. **Journal of the Royal Statistical Society. The Statistician**, v.14, p. 255–266, 1964

IBGE – Instituto Brasileiro de Geografia e Estatística. Pesquisa Trimestral do Leite. 2022. Disponível em: < <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9209-pesquisa-trimestral-do-leite.html?=&t=o-que-e> > Acesso em: 02 fev. 2022.

JUNQUEIRA, R. V. B. Análise da sazonalidade da produção de leite no Brasil. *Pecuária de leite*, 2013. Disponível em: <https://pt.engormix.com/pecuaria-leite/artigos/analise-sazonalidade-producao-leite-t38244.htm>. Acesso em: 05 de mar. de 2022.

NAPOLITANO, M. A sazonalidade da produção de leite ainda é a mesma? *Milkpoint*, 2021. Disponível em: <https://www.milkpoint.com.br/artigos/producao-de-leite/a-sazonalidade-da-producao-ainda-e-a-mesma-228112/>. Acesso em: 05 de mar. de 2022.

NERLOVE, M. The Dynamics of Supply: Retrospect and Prospect. ***American Journal of Agricultural Economics***, v.61, p. 874-888, 1979.

SMYTH, G. K. Generalized linear models with varying dispersion. ***Journal of the Royal Statistical Society: Series B (Methodological)***, v. 51, p. 47-60, 1989.

WEN, Q, GAO, J, SONG, X, SUN, L, XU, H, ZHU, S. Robust STL: A robust seasonal-trend decomposition algorithm for long time series. **In: *Proceedings of the AAAI Conference on Artificial Intelligence***, v. 33, p. 5409-5416, 2019.