



Impacto causal de uma visita técnica no processo de fornecimento de uma fazenda via modelo estrutural temporal Bayesiano

Vinícius Basseto Félix¹, Beatriz Marquezini Alvarenga² e Larissa Bueno Fernandes³

¹ Mestrando em Bioestatística - Universidade Estadual de Maringá - felix_prot@hotmail.com

² Consultora de inteligência de mercado em Zoocenia - Gestão Agropecuária

³ Mestranda em Bioestatística - Universidade Estadual de Maringá

RESUMO

A fim de avaliar e demonstrar seus resultados, uma empresa que presta consultoria técnica em fazendas tem por necessidade a mensuração do impacto de suas visitas. Desta forma, o trabalho mostra como é possível a obtenção desta medida através de um modelo estado-espaco com abordagem Bayesiana, bem com sua significância e comportamento. A metodologia aplicada a um conjunto de dados real mostra claramente o efeito da intervenção, de forma que a visita técnica da empresa Gestão Agropecuária proveu uma melhora na produção da fazenda auditada.

Palavras chave: Impacto causal, modelo espaço-estado, intervenção.

1 INTRODUÇÃO

O processo de produção de carne bovina é dividido em diversas etapas, desde a fabricação de ração até o abate do animal. Cada uma destas fases por mais automatizada que seja, possui envolvimento humano, de forma que para se alcançar uma maior eficiência, é necessário estabelecer processos.

Assim, uma das formas de auditar a qualidade de uma fazenda é a visita técnica, em que especialistas acompanham e analisam o processo, buscando falhas eminentes ou potenciais melhorias. Entretanto, é difícil a mensuração do impacto por meio desta visita, e uma solução é a mensuração do impacto causal, ou seja, utilizar-se da série histórica da fazenda e entender sua eficiência ao longo do tempo, e dado a intervenção técnica, prever como se daria se a mesma não tivesse acontecido, permitindo assim o cálculo do efeito da visita.

2 METODOLOGIA

2.1 Modelo

Segundo [1], modelos estruturais temporais são modelos de estado-espaco para séries temporais. Assim, seja Y_t um processo estocástico, a representação espaço-estado implica na decomposição com processos estocásticos não-observados, ou seja, latentes, desta forma pode-se expressá-lo como um par de equações:

$$y_t = Z_t^T \alpha_t + \varepsilon_t, \quad (1)$$

em que y_t é uma observação escalar, Z_t é um vetor, α_t é um vetor latente, e $\varepsilon_t \sim N(0, \sigma_t^2)$ uma variável aleatória (v.a) independente. Já para 2º equação,

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad (2)$$

em que T_t é uma matriz de transição, R_t é uma matriz de controle, e $\eta_t \sim N(0, Q_t)$ é uma v.a independente.

A equação (1) remete aos dados observados (y_t) relacionados ao vetor latente de estado (α_t), enquanto a (2) condiz com a evolução do vetor latente ao longo do tempo. As matrizes Z_t , T_t , R_t e Q_t permitem a inclusão de componentes para capturar características como tendência ou até efeitos sazonais.

Neste trabalho utiliza-se de um modelo de nível local, expresso como:

$$y_t = \mu_t + \beta^T x_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (3)$$

em que β^T são coeficientes associados as covariáveis, e

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad (4)$$

em que é necessário a especificação de μ_1 .

2.2 Inferência Bayesiana

Uma prática comum nos modelos estruturais temporais, via abordagem bayesiana, é o uso da priori *spike-and-slab* [3], tal que há muitos preditores e acredita-se que a maioria serão nulos, este tipo de priori atribui uma probabilidade de que todo coeficiente pode ser nulo. Assim, seja $\gamma_k = 1$ se $\beta_k \neq 0$, e $\gamma_k = 0$ se $\beta_k = 0$, tem-se que a priori pode ser expressa como

$$p(\beta, \gamma, \sigma_\varepsilon^2) = p(\beta_\gamma | \gamma, \sigma_\varepsilon^2) p(\sigma_\varepsilon^2 | \gamma) p(\gamma), \quad (5)$$

em que β_γ é o subconjunto de β , tal que $\beta_k \neq 0$. Além disso, toma-se que a priori de γ é dada por

$$\gamma \sim \prod_{k=1}^K \pi_k^{\gamma_k} (1 - \pi_k)^{1 - \gamma_k}, \quad (6)$$

ou seja, parte-se da distribuição de *Bernoulli*. Ainda, para simplificação pode-se tomar $\pi_k = \pi \forall k$, neste trabalho utilizou-se também que $\pi = 1$.

Além da priori de γ , as prioris condicionais de β e σ_ε^2 , dado γ , podem ser expressas como

$$\beta_\gamma | \sigma_\varepsilon, \gamma \sim N(0, \sigma_\varepsilon^2 (\Omega_\gamma^{-1})^{-1}), \quad (7)$$

em que Ω_γ^{-1} corresponde as linhas e colunas da matriz simétrica Ω^{-1} quando $\gamma_k = 1$. Além disso,

$$\frac{1}{\sigma_\varepsilon^2} | \gamma \sim \text{Gamma}(0, 0.5; 5s_y), \quad (8)$$

em que s_y é o desvio padrão marginal da resposta.

Utiliza-se do método *Markov chain Monte Carlo* (MCMC) [2], para os parâmetros vistos acima e outros (θ), tal que

- Para α , simula-se através de $p(\alpha | y, \beta, \theta, \sigma_\varepsilon^2)$;
- Para θ , simula-se através de $p(\theta | y, \alpha, \beta, \sigma_\varepsilon^2)$;
- Para β e σ_ε^2 , simula-se através de $p(\beta, \sigma_\varepsilon^2 | y, \alpha, \theta)$.

Em que utiliza 1000 iterações, em que as 200 primeiras são descartadas e métodos visuais são utilizados para avaliação da convergência das cadeias do método MCMC. Por fim, como interesse está na previsão dos dados pré-intervenção, utiliza-se da distribuição a posteriori da predição, tal que a mesma pode ser expressa por

$$p(\tilde{y} | y) = \int p(\tilde{y} | \theta, \beta, \sigma_\varepsilon^2, \alpha) p(\theta, \beta, \sigma_\varepsilon^2, \alpha | y) d(\theta, \beta, \sigma_\varepsilon^2, \alpha), \quad (9)$$

em que \tilde{y} é o conjunto de valores a ser predito. Toda avaliação de significância será feita via intervalo de 95% credibilidade.

2.3 Dados

Para aplicação, utilizou-se um banco de dados com 7260 observações, referente ao período de 03/03/2017 a 09/10/2017, em que se fez a soma dos valores de ração fornecida e predita por dia, e posteriormente calculado a diferença percentual diária, dada por: $100 * (\text{fornecido} - \text{previsto}) / \text{previsto}$.

Assim para análise de intervenção utiliza-se do fornecimento obtido via o sistema automatizado de coletado de dados, em que se faz uma análise diária, avaliando a diferença percentual diária entre o previsto e realizado.

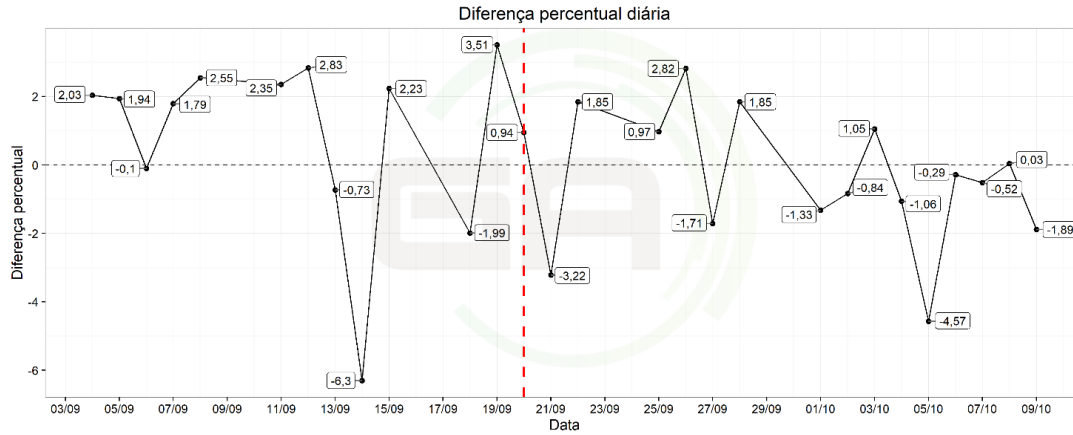


Figura 1: Diferença percentual diária ao longo do período de 03/09 a 09/10, com marcação da intervenção no dia 20/09.

Vê-se na Figura 1 o comportamento da diferença percentual diária, em que a linha vermelha assinala a visita técnica realizada pela Gestão Agropecuária (GA). A fim de avaliar esta intervenção faz-se um ajuste de um modelo de séries temporais para prever o período pós-intervenção, dado a previsão de como se a mesma não tivesse ocorrido.

3 RESULTADOS E DISCUSSÕES

Para o ajuste do modelo, fez-se a adição da covariável peso inicial do lote, tal que se obteve a estimativa pontual e o respectivo intervalo de credibilidade (ICr), 95%, para cada parâmetro.

Parâmetro	Estimativa	ICr 95%
β	0,981	[0,358; 1,58]
σ_{ϵ}^2	0,617	[0,522; 0,735]
σ_{η}^2	0,016	[0,013; 0,021]

Tabela 1: Estimativas dos parâmetros do modelo de nível local.

Vê-se pela Tabela 1 que os intervalos não contém o valor 0, logo há significância dos parâmetros estimados.

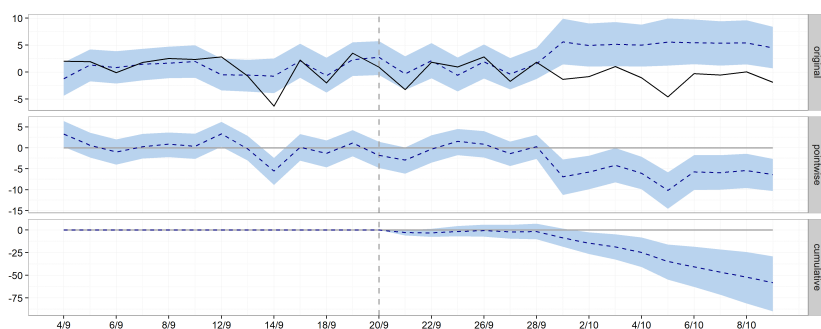


Figura 2: Gráfico da série prevista sem intervenção (*original*), diferença entre o previsto e observado (*pointwise*) e efeito cumulativo da intervenção (*cumulative*).

A Figura 2 apresenta 3 painéis, o 1º, intitulado *original*, representa a série histórica original, como na Figura 1, (em preto) e a previsão da série sem a intervenção (linha pontilhado, em azul). O 2º painel (*pointwise*) mostra a diferença entre a série original e a previsão. Por fim, o 3º painel (*cumulative*) mostra o efeito cumulativo do 2º painel pós-intervenção.

	Efeito médio	Efeito acumulado
Observado	-0,45	-6,80
Predito (erro padrão)	3,3 (0,95)	49,7 (14,27)
ICr 95%	[1,4, 5,2]	[21,5, 77,9]
Efeito absoluto (erro padrão)	-3,8 (0,95)	-56,5 (14,27)
ICr 95%	[-5,6, -1,9]	[-84,7, -28,3]
Efeito relativo (erro padrão)	-114% (29%)	-114% (29%)
ICr 95%	[-170%, -57%]	[-170%, -57%]

Tabela 2: Estimativas dos efeitos mensurados representados pela Figura 2.

Através da Tabela 2 que após a visita da GA a diferença percentual tinha uma média aproximada de -0,45. Porém, na ausência da intervenção fez-se uma previsão que a diferença média percentual seria próxima de 3,8. Assim, com a intervenção mostrou-se um decréscimo de 114%, e visto que a probabilidade de se obter este efeito por acaso é inferior a 0,001, há evidências de que o efeito da visita da GA foi significante, provocando um decréscimo na diferença percentual.

4 CONCLUSÃO

Conclui-se que o modelo permite uma previsão interessante, bem como a inclusão de priori informativas e covariáveis. Além disso, o método mostra claramente o efeito da intervenção através do cálculo do impacto causal, de forma que a visita técnica da GA proveu uma melhora na produção da fazenda auditada.

5 AGRADECIMENTOS

Agradecemos a empresa Gestão Agropecuária por ceder os dados.

Referências

- [1] BRODERSEN, K. H. et al. Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, v. 9, n. 1, p. 247-274, 2015.
- [2] JOHANNES, M.; POLSON, N. MCMC methods for continuous-time financial econometrics. *Handbook of Financial Econometrics*, v. 2, n. 1, p. 12-26, 2009.
- [3] SCOTT, Steven L.; VARIAN, Hal R. Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, v. 5, n. 1-2, p. 4-23, 2014.