

## **Transfer functions: learning about a lagged exposure-outcome association in time-series data**

Hiroshi Mamiya<sup>1</sup>, Alexandra M. Schmidt<sup>1</sup>, Erica E. M. Moodie<sup>1</sup>, David L. Buckeridge<sup>1</sup>

<sup>1</sup> School of Population and Global Health, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 772 Sherbrooke West, Montreal, QC, Canada, H3A 1G1

**Keywords:** Distributed lag model, Time-series analysis, Dynamic regression, Display promotion, Food environment, Sugary yogurt.

### **Contact information:**

Hiroshi Mamiya,

11-4119 Ave Madison, Montreal, QC, H4B 2T8, Canada

Tel: 778-977-7110

Fax: Currently unavailable

Email: [hiroshi.mamiya@mail.mcgill.ca](mailto:hiroshi.mamiya@mail.mcgill.ca)

**Funding:** This study was funded by an Institut de valorisation des données (IVADO) post-doctoral fellowship awarded to author HM. AMS acknowledges funding from the Natural Sciences and Engineering Research Council (NSERC) of Canada. EEMM is supported by a chercheur de mérite career award from the Fonds de recherche du Québec – Santé (FRQS) and a Canada Research Chair (Tier 1) in Statistical Methods for Precision Medicine. DLB is supported by a Canada Research Chair (Tier 1) in Population Health Informatics.

**Author contributions:** The study was conceived and designed by HM and was reviewed and approved by the other authors. AMS and EEMM provided inputs on the statistical analysis and interpretation of the results. DLB provided the data and computational resources. Data analysis and drafting of manuscript was led by HM. All authors reviewed, provided critical comments to the manuscript, and approved the final version of the manuscript for submission.

**Ethics of human subject participation:** This study used secondary data that are aggregated store-level measurements of consumer purchasing, rather than individual consumer level data. The study therefore did not require a separate written or verbal consent from human subjects.

## **Abstract**

Many population exposures in time-series analysis, including food marketing, exhibit a time-lagged association with population health outcomes such as food purchasing. A common approach to measuring patterns of associations over different time lags relies on a finite-lag model, which requires correct specification of the maximum duration over which the lagged association extends. However, the maximum lag is frequently unknown due to the lack of substantive knowledge or the geographic variation of lag length. We describe a time-series analytical approach based on an infinite lag specification under a transfer function model that avoids the specification of an arbitrary maximum lag length. We demonstrate its application to estimate the lagged exposure-outcome association in food environmental research: display promotion of sugary beverages with lagged sales.

## **Introduction**

Many population exposures have time-lagged association with population health outcomes, cumulating their impact over time<sup>1,2</sup>. An often-overlooked aspect of public health nutrition, for example, is the impact of unhealthy food promotions which frequently exhibit time-lagged associations with outcomes such as unhealthy food purchasing. In other words, exposures show a persistent association with outcomes in future times periods. Distributed lag models capture such lagged patterns by incorporating a series of lagged values of an exposure, whose coefficients typically are characterized by a functional structure such as low-order polynomials or splines<sup>1,2</sup>. Aside from the specification of the degree of polynomials, these models require a priori knowledge about the maximum number of lags (hereafter referred to as lag length), since incorrect specification of the lag length may fail to capture the lag structure in terms of shape and

magnitude<sup>3-5</sup>. However, the plausible lag length for a given exposure-outcome association is often unknown due to the lack of substantive knowledge and, in the case of food promotions as an exposure, the lag length may vary widely across geography and food categories of varying shelf-life, such as soda vs. yoghurts<sup>3,4,6</sup>.

In this brief report, we describe a distributed lag model with an infinite lag specification under a transfer function as an approach to obviate the need for a priori specification of lag length. We motivate this with an analysis which aims to capture the lagged association between an exposure related to an obesogenic food environment (weekly time-series of within-store display promotion of sugary beverages) on the purchasing ultra-processed food (weekly sales of such beverages).

## Methods

We assume that measurements are available at discrete and equally spaced intervals, and focus on a widely used and interpretable specification of the transfer function called the Koyck decay<sup>7</sup>, a monotonic decrease of association over time capturing the diminishing association of an exposure such as media advertising with food purchasing<sup>4</sup>.

A transfer function is a parsimonious specification of lagged terms and uses a structural variable, denoted  $E_t$  below, to summarize the current (at time  $t$ ) and cumulative association (up to time  $t$ ) between the exposure variable  $x$  and the outcome  $y$ . Using a decay parameter  $\lambda$  and a lag index  $h$ , monotonically decreasing associations are represented as;

$$E_t = \beta x_t + \lambda^1 \beta x_{t-1} + \lambda^2 \beta x_{t-2} + \dots + \lambda^h \beta x_{t-h},$$

which recursively reduces to

$$E_t = \beta x_t + \lambda E_{t-1}.$$

The coefficient  $\beta$  captures the *immediate* association at time  $t$ , while the coefficient  $\lambda$  captures the extent of the lag: a value closer to 1 implies a more persistent lag, while a value nearer to zero indicates a shorter lag. When the structure  $E_t$  is added to a regression model, learning of a lagged association is accomplished by estimating  $\lambda$  and  $\beta$ .

An interpretation of a lagged association combining these two parameters is called an *impulse response function*. This represents the response (change) of  $y_{t+0} + y_{t+1} + y_{t+2} + \dots + y_{t+h}$  to impulse (one-unit increase of  $x$  at time  $t$  only), while all other variables at time  $t$  or earlier are fixed constant<sup>8</sup>. For the Koyck lag, the impulse response function takes the form

$\beta\lambda^0 + \beta\lambda^1 + \beta\lambda^2 + \dots + \beta\lambda^h$  as visualized in Supplementary Figure S1a-b and

Supplementary Appendix S1. Imposing the constraint  $-1 < \lambda < 1$  will ensure that this function decays to zero as  $h \rightarrow \infty$ .

The general specification of the transfer function governing the shape of lag is as follows;

$$E_t = \beta_0 x_{t-0} + \beta_1 x_{t-1} \dots \beta_p x_{t-p} + \lambda_0 E_{t-0} + \lambda_1 E_{t-1} + \dots + \lambda_q E_{t-q},$$

where the Koyck decay is defined by  $p = 0, q = 1$ . More complex shapes can be specified by higher order transfer functions, i.e., using higher values of  $p$  and  $q$ , as long as a large number of  $p$  does not lead to collinearity across variables  $x_{t-p}$ . Possible shapes include a peak after several lags followed by decay. Further discussion of transfer functions is provided elsewhere.<sup>9-12</sup>

Clearly, there is no need to provide a lag length in this infinite lag specification, but a priori knowledge about plausible shapes of the lag is needed, along with constraints on the value of  $\lambda$ : in case of the Koyck lag,  $0 < \lambda < 1$  ensures monotonic decay, while  $-1 < \lambda < 0$  leads to an oscillating lag (Supplementary Figure 1c). The selection of candidate shapes may rely on model selection, for example guided by metrics for model fit (an information criterion) or forecasting error. This is in contrast to a commonly used finite lag model, where the shape of the lag is learned from data and the plausible (or arbitrary) value of lag length is fixed in advance by the investigator.

## **Application**

The exposure of interest is the weekly within-store display promotion of an obesogenic food product (sugary beverages) that potentially exhibits time-lagged association with the outcome, purchasing quantity of sugary beverages in the same store. Display promotion is the temporary placement of selected items in prominent locations and is shown to have a positive association

with the sales of ultra-processed (‘junk’) food, such as sugar sweetened beverages. This exposure is thus increasingly recognized as the policy target to improve population diets<sup>13</sup>. Our food category of interest is sugar sweetened (not plain) drinkable yogurt, which is a hidden and important source of dietary sugar, in particular among children<sup>14,15</sup>. The weekly status of the proportion of display-promoted sweetened yogurt items (continuous exposure) and the weekly aggregated sum of the sales quantity of sugar-sweetened drinkable yogurt items (continuous outcome) were generated by a single large supermarket in Montreal, Canada over  $n=311$  weeks. A detailed description of the promotion exposure and outcome are provided in Supplementary Appendix S2.

Weekly sales were modelled as a function of the structural variable  $E_t$ , covariates, a seasonal term, and an intercept. We selected the Koyck lag specification with  $\lambda$  bounded to the interval  $[0,1]$ , given that the promotion is expected to have a monotonically decaying association with purchasing due to temporarily decreasing awareness among exposed consumers<sup>4</sup>. The time-series regression used in this study is a dynamic linear model, which allows regression parameters to vary smoothly over time<sup>10,16</sup>. The model was fit under the Bayesian framework. The details of the dynamic linear model, model specification, and prior specification are provided in Supplementary Appendix S3. The parameters of interest are  $\beta$  and  $\lambda$ , from which the impulse response function was estimated. Code to perform the analysis is provided at [https://github.com/hiroshimamiya/promotionLag/blob/main/discountLag\\_KoyckTransfer.stan](https://github.com/hiroshimamiya/promotionLag/blob/main/discountLag_KoyckTransfer.stan)

## Results

The estimated posterior mean of  $\beta$  was 0.68 (95% Credible Interval [CI]: 0.39-0.96), an association implying  $\exp(0.68)$  higher absolute (i.e., non-log) quantity of sugar-sweetened drinkable yogurt at week  $t$  if all yogurt items were display promoted in that week. The posterior

summary of the decay coefficient  $\lambda$  was modestly strong: 0.47 (95% CI 0.20-0.72), as shown by the distinct lag in the estimated impulse response function (Figure 1). The posterior median of the excess quantity of sales associated with lag alone on the absolute (*non-log*) scale was 11,890 (95% CI 2,594-32,051) servings for the entire 6 years period, while that of excess sales related to the immediate association alone was 13,506 (95% CI: 7,726-19,750). An autocorrelation function plot of residuals over time suggests that temporal autocorrelation in the fitted model was largely absent (Supplementary Figure S4).

## Discussion

Time-lagged associations between exposures and outcomes are ubiquitous and, therefore, are important components to decompose in time-series studies. We described the transfer function as an approach to capture lagged association when the lag length is unknown, with an application in food environment research.

The transfer function approach is particularly useful when lag length may vary across many food categories and/or store-neighborhood socio-economic status, the latter being particularly relevant when investigating the role of food promotions in driving geographic and socio-economic disparities of diets. The transfer function can also be used in climate-related environmental time-series analysis to investigate spatially heterogeneous persistence of lagged association between temperature and arbovirus incidence<sup>17</sup>. The relatively simple Koyck lag function may also be used to capture the monotonically decaying lagged association of high ambient temperature with mortality and morbidity, although there may be a need to account for mortality displacement in some geographic regions<sup>6</sup>.

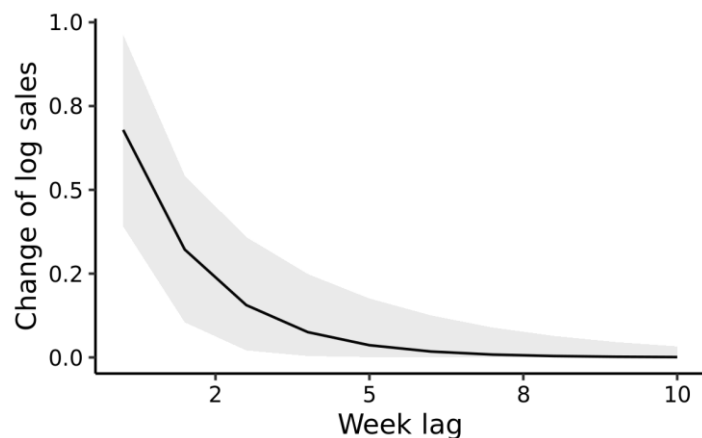
The added value of the proposed approach in estimating a valid lag structure relies on researcher's prior knowledge of plausible lag shapes i.e., values of  $p$  and  $q$  in the transfer



function. Thus, when such knowledge is lacking, existing finite lag models using polynomial or basis function will allow data-driven estimation of the approximate shape of lag, conditional on the availability of theoretical or empirical evidence about lag length. However, the finite lag approach requires other knowledge, specifically the lag length.

Limitations of the current approach include challenges in selecting the most appropriate shape of transfer function, for example when competing shapes show similar model fit and thus the underlying process generating lagged association is undetermined.

Modeling exposure-lag association using a transfer function addresses one of the current methodological gaps in epidemiological time-series analysis, which is the potential bias in lagged association due to incorrect specification of lag length in finite lag model. The proposed approach shown here can be readily implemented using time-series regression and codes provided.



**Figure 1:** The estimated impulse response function of display promotion on the natural log sales of sugar-sweetened drinkable yogurt, based on the lag parameter  $\beta$  and  $\lambda$  learned from the time-series of sales data from a single store in Montreal ( $n=311$  weeks, 2008-2013). The grey band indicates pointwise 95% posterior credible interval. The immediate association is displayed at lag 0 and is 0.68 (95% Posterior Credible Interval: 0.39-0.96), indicating that the immediate impact of display promotion is a doubling of sales, since  $\exp(0.68) = 1.97$ .

## References

1. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. *Int J Epidemiol*. 2013;42(4):1187-1195. doi:10.1093/ije/dyt092
2. Gasparrini A. Modelling lagged associations in environmental time series data: a simulation study. *Epidemiology*. 2016;27(6):835-842. doi:10.1097/EDE.0000000000000533
3. Wang P, Zhang X, Hashizume M, Goggins WB, Luo C. A systematic review on lagged associations in climate–health studies. *Int J Epidemiol*. 2021;50(4):1199-1212. doi:10.1093/ije/dyaa286
4. Hanssens DM, Parsons LJ, Schultz RL, eds. Design of dynamic response models. In: *Market Response Models: Econometric and Time Series Analysis*. International Series in Quantitative Marketing. Springer US; 2001:139-182. doi:10.1007/0-306-47594-4\_4
5. Rushworth A. Bayesian Distributed Lag Models. *arXiv:180106670 [stat]*. Published online January 20, 2018. Accessed September 5, 2021. <http://arxiv.org/abs/1801.06670>
6. Guo Y, Gasparrini A, Armstrong BG, et al. Heat wave and mortality: a multicountry, multicomunity study. *Environ Health Perspect*. 2017;125(8):087006. doi:10.1289/EHP1026
7. Koyck, L. M. Distributed Lags and Investment Analysis. In: *Advanced Macroeconomics*. 4th edition. North- Holland Publishing Co; 1954.
8. Hamilton JD. Chapter 11. Vector Autoregressions. In: *Time Series Analysis*. Princeton University Press; 1994:291-349. doi:10.1515/9780691218632-095
9. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Chapter 11: Transfer function models. In: *Time Series Analysis: Forecasting and Control*. 5th ed. Wiley; 2015.
10. West M, Harrison J. *Bayesian Forecasting and Dynamic Models*. Second edition. Springer; 1997.
11. Alves MB, Gamerman D, Ferreira MA. Transfer functions in dynamic generalized linear models. *Statistical Modelling*. 2010;10(1):03-40. doi:10.1177/1471082X0801000102
12. Ravines RR, Schmidt AM, Migon HS. Revisiting distributed lag models through a Bayesian perspective. *Appl Stoch Models Bus Ind*. 2006;22(2):193-210. doi:<https://doi.org/10.1002/asmb.628>
13. Hecht AA, Perez CL, Polascek M, Thorndike AN, Franckle RL, Moran AJ. Influence of food and beverage companies on retailer marketing strategies and consumer behavior. *Int J Environ Res*. 2020;17(20):7381. doi:10.3390/ijerph17207381

14. Moore JB, Horti A, Fielding BA. Evaluation of the nutrient content of yogurts: a comprehensive survey of yogurt products in the major UK supermarkets. *BMJ Open*. 2018;8:e021387. doi:10.1136/bmjopen-2017-021387
15. Langlois K, Garriguet D, Gonzalez A, Sinclair S, Colapinto CK. Change in total sugars consumption among Canadian children and adults. *Health Rep*. 2019;30(1):10-19.
16. Petris G, Petrone S, Campagnoli P. *Dynamic Linear Models with R*. Springer-Verlag; 2009. doi:10.1007/b135794
17. Freitas LP, Schmidt AM, Cossich W, Cruz OG, Carvalho MS. Spatio-temporal modelling of the first Chikungunya epidemic in an intra-urban setting: The role of socioeconomic status, environment and temperature. *PLOS Neglected Tropical Diseases*. 2021;15(6):e0009537. doi:10.1371/journal.pntd.0009537