پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

# Application of Functional Linear Regression with Impact Points to Real Data: A Case Study

Alireza Shirvani*[1]

[1]Department of Statistics, Velayat University, Iranshahr, Iran

**Abstract**

This study demonstrates the practical application of the methodology introduced in Shirvani *et al.* (2024) for detecting and testing the significance of impact points within a functional linear regression framework. Using data from a Google AdWords campaign, the analysis confirms that the functional predictor at the identified impact points significantly influences the response variable, highlighting the utility of the approach in real-world settings. The results underscore the importance of impact point detection in functional data modeling.

**Keywords:** Functional linear regression, Impact point, Hypothesis test.
**Mathematics Subject Classification (2020):** 62G10.

## 1 Introduction

Functional Data Analysis (FDA) is a powerful statistical tool for modeling data observed over a continuum, such as time or space (for more details, see Ramsay and Silverman (2005)). A key topic within FDA is the identification of impact points, discrete observation times at which the corresponding functional values have a significant influence on the response variable, considered by Kneip *et al.* (2016).

In Shirvani *et al.* (2024), a rigorous hypothesis testing approach was proposed to evaluate whether the functional predictor evaluated at specific impact points in the domain has a statistically significant effect on the response. The framework involves constructing an appropriate test statistic whose distribution under the null hypothesis is well characterized.

The investigation of impact points in functional linear models has received considerable attention. Cardot *et al.* (2003) pioneered early work on hypothesis testing in this context, providing foundational tools. Crambes *et al.* (2009) and Kneip *et al.* (2016) further refined estimation procedures and impact detection techniques. Recent developments, such as those by Liebl *et al.* (2020), have enhanced impact point identification through improved estimation methods and bootstrap approaches. The testing procedure outlined in Shirvani *et al.* (2024) builds upon these advancements, offering a statistically rigorous and practically applicable framework.

*Corresponding author, a.shirvani@velayat.ac.ir

This paper applies the methodology from Shirvani *et al.* (2024) to a real dataset obtained from a Google AdWords campaign, aiming to demonstrate its effectiveness in practical data analysis and its capacity to identify meaningful impact points.

## 2 Model

We analyze a functional linear model where the response variable $Y$ is scalar, and the predictor variable $X$ belongs to $L^2(\mathcal{I})$, with $\mathcal{I} = [a, b] \subset \mathbb{R}$ being a bounded interval, and $L^2(\mathcal{I})$ representing the space of all square-integrable functions on $\mathcal{I}$. The data consist of independent observations $(X_i, Y_i)$, for $i = 1, \ldots, n$, of the pair $(X, Y)$. The functional predictor $X$ over $\mathcal{I}$ satisfies

$$E\left[\int_{\mathcal{I}} X^2(t)dt\right] < \infty.$$

For simplicity, we assume the variables are centered, meaning

$$E(Y) = 0 \quad \text{and} \quad E[X(t)] = 0 \quad \text{for a.e. } t \in [a, b].$$

Then, the $Y_i$'s are generated according to the following functional linear regression model with points of impact:

$$Y_i = \int_{\mathcal{I}} \beta(t)X_i(t)dt + \sum_{r=1}^{S} \beta_r X_i(\tau_r) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where the errors $\varepsilon_i$ are i.i.d. with zero mean and finite variance $\sigma^2$, the slope function $\beta$ is an unknown element of $L^2(\mathcal{I})$, and the errors $\varepsilon_i$ are independent of the $X_i$.

Note that the overall effect of the entire trajectory $X_i(\cdot)$ on $Y_i$ is captured by

$$\int_{\mathcal{I}} \beta(t)X_i(t)dt$$

in Model (1). The model also incorporates an unknown number $S \in \mathbb{N}$ of specific time points $\tau_1, \ldots, \tau_S$, called "points of impact," such that the response $Y_i$ is significantly influenced by the values $X_i(\tau_1), \ldots, X_i(\tau_S)$.

Model (1) involves the unknown parameters: the function $\beta(t)$, the number of impact points $S \geq 0$, and the impact points $\tau_r$, as well as the coefficients $\beta_r$, for $r = 1, \ldots, S$, all of which must be estimated from the data.

## 3 Data Description and Application

The dataset under consideration includes daily impressions and clicks. This real-world dataset, obtained from a Google AdWords campaign-an essential platform of Alphabet (Google's parent company)-provides a practical example for applying the proposed impact point detection method. Google AdWords operates on a Pay-Per-Click (PPC) mechanism, where advertisers bid for sponsored impressions that appear alongside search results when users enter relevant keywords. The auction system allocates a limited number of sponsored impressions based on an ad-rank procedure that accounts for bid amounts and quality scores. Crucially, advertisers only pay when a user clicks on their ad.

To model the relationship between the total annual clicks and the trajectories of daily impressions, we used a functional linear regression framework with points of impact. Specifically, the response variable is the logarithm of the total annual clicks:

$$Y_i = \log\left(\sum_{t=1}^{365} \text{clicks}_{it}\right),$$

and the functional predictor is given by the logarithm of daily impressions:

$$X_i(t) = \log(\text{impressions}_{it}),$$

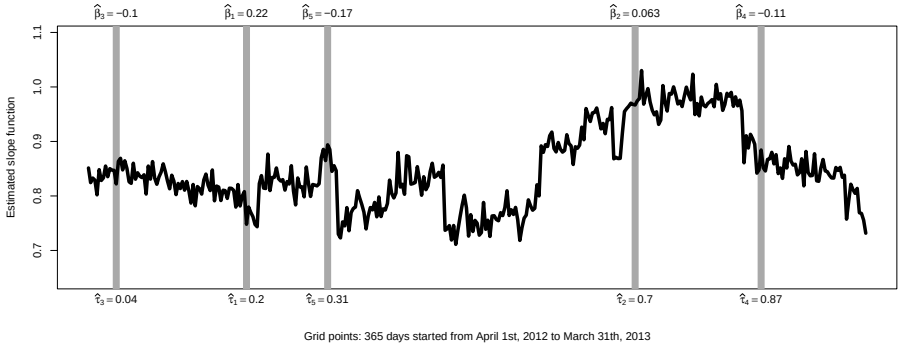



Figure 1: Estimated slope function and points of impact using Kneip *et al.* (2016) method. The horizontal axis shows 365 days of the year, and the vertical axis displays the estimated slope function. The impact points are indicated by vertical lines.

for $i = 1, \ldots, 903$ and $t = 1, \ldots, 365$. The data originate from a Google AdWords campaign conducted between April 1, 2012, and March 31, 2013.

Applying the method proposed by Kneip *et al.* (2016), the estimated impact points are summarized in Figure 1. Five impact points were identified at the following dates: June 14 ($\hat{\tau}_1$, $\hat{\beta}_1 = 0.22$), December 13 ($\hat{\tau}_2$, $\hat{\beta}_2 = 0.06$), April 14 ($\hat{\tau}_3$, $\hat{\beta}_3 = -0.10$), February 10 ($\hat{\tau}_4$, $\hat{\beta}_4 = -0.11$), and July 22 ($\hat{\tau}_5$, $\hat{\beta}_5 = -0.17$). These impactful days could correspond to specific seasonal effects, holiday periods, or promotional spikes, providing valuable insights in marketing and campaign planning. To assess whether including these impact points significantly enhances the model, the hypothesis testing procedure suggested by Shirvani *et al.* (2024) was employed. The resulting p-value was zero, indicating that the impact points have a statistically significant effect on the response variable.

# 4 Conclusion

This application illustrates that the impact points approach from Kneip *et al.* (2016) effectively identifies significant impactful days in a functional predictor. The results align with domain expectations and provide valuable interpretability for practical applications. Importantly, the hypothesis test introduced by Shirvani *et al.* (2024) assesses the overall significance of the entire set of impact points collectively, rather than testing each coefficient individually. The fact that the p-value is zero indicates that not all impact point coefficients are zero-some have a meaningful effect-confirming the presence of significant impact points and emphasizing the importance of this test in validating the collective impact of those points.

In the future, it will be important to develop a hypothesis test that can assess the significance of individual impact points separately, rather than evaluating the collective set. Such a test would allow for a more detailed understanding of which specific points have a meaningful influence on the response variable, providing greater specificity and interpretability in impact point detection.

# References

Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003), Testing hypotheses in the functional linear model, *Scandinavian Journal of Statistics*, **30**, 241–255.

Crambes, C., Kneip, A., and Sarda, P. (2009), Smoothing splines estimators for functional linear regression, *Annals of Statistics*, **37**, 35–72.

Kneip, A., Dette, H., and Gertheiss, J. (2016), Variable selection in functional linear models, *Annals of Statistics*, **44**(3), 1060–1088.

Liebl, D., Rameseder, S. and Rust, C. (2020), Improving estimation in functional linear regression with points of impact: insights into Google AdWords, *Journal of Computational and Graphical Statistics,* **29**, 814–826.

Ramsay, J., and Silverman, B. (2005), *Functional Data Analysis*, 2nd Ed., Springer.

Shirvani, A., Khademnoe, O. and Hosseini-Nasab, M. (2024), Hypothesis testing for points of impact in functional linear regression, *Computational and Applied Mathematics*, **43**, 201.