

Estimation of Causal Functional Linear Regression Models

J.C.S.de Miranda¹

¹ Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

Correspondence: J.C.S.de Miranda, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil.

E-mail: simon@ime.usp.br

Received: October 7, 2017 Accepted: November 3, 2017 Online Published: November 6, 2017

doi:10.5539/jmr.v9n6p106

URL: <https://doi.org/10.5539/jmr.v9n6p106>

This work was partially supported by FAPESP grant 2013/00506-1.

Abstract

We present a methodology for estimating causal functional linear models using orthonormal tensor product expansions. More precisely, we estimate the functional parameters α and β that appear in the causal functional linear regression model:

$$\mathcal{Y}(s) = \alpha(s) + \int_a^b \beta(s, t) \mathcal{X}(t) dt + \mathcal{E}(s),$$

where $\text{supp } \beta \subset \mathfrak{T}$, and \mathfrak{T} is the closed triangular region whose vertexes are (a, a) , (b, a) and (b, b) . We assume we have an independent sample $\{(\mathcal{Y}_k, \mathcal{X}_k) : 1 \leq k \leq N, k \in \mathbb{N}\}$ of observations where the \mathcal{X}_k 's are functional covariates, the \mathcal{Y}_k 's are time order preserving functional responses and \mathcal{E}_k , $1 \leq k \leq N$, is i.i.d. zero mean functional noise.

Keywords: causal, functional linear model, orthonormal series expansions, tensor product.

1. Introduction

Functional linear models are generalizations of linear models to functional data. There are three basic functional linear models according to the functional data being the response, the covariates or both. They are written:

$$\mathcal{Y}(t) = \alpha(t) + x\beta(t) + \mathcal{E}(t),$$

the functional response model with scalar covariate,

$$y = \alpha + \int_a^b \beta(t) \mathcal{X}(t) dt + \mathcal{E}(t),$$

the scalar response model with functional covariate, and

$$\mathcal{Y}(s) = \alpha(s) + \int_a^b \beta(s, t) \mathcal{X}(t) dt + \mathcal{E}(s),$$

the functional response model with functional covariate which is also called the fully functional model. They, as well as their generalizations, have a wide range of applications. See all the listed references in this note.

Concerning the fully functional model, with the additional assumption that \mathcal{Y} and \mathcal{X} are both functions of *time* and are both defined on the *same* time interval, $[a, b]$, then, as already remarked in Functional Data Analysis literature, either we admit that the values of the covariates at future times impact the values of the response at past and present times or we admit that the support of β is contained in the closed triangular region \mathfrak{T} whose vertexes are (a, a) , (b, a) and (b, b) .

The later alternative is equivalent to the

“Volterra type” model

$$\mathcal{Y}(s) = \alpha(s) + \int_a^s \beta(s, t) \mathcal{X}(t) dt + \mathcal{E}(s) \quad (1)$$

and to the “Fredholm type” model

$$\mathcal{Y}(s) = \alpha(s) + \int_a^b \beta(s, t) \chi_{\mathfrak{T}}(s, t) \mathcal{X}(t) dt + \mathcal{E}(s) \quad (2)$$

where β is a function defined on a domain that contains \mathfrak{T} and $\chi_{\mathfrak{T}}$ is its indicator function. We will call this model *Causal Functional Linear Model*.

Model (2) may be *directly* estimated with expansions in tensor product basis. This is possible by expanding the function $\tilde{\beta}$, defined on the square $[a, b]^2$ by $\tilde{\beta}(x, y) = \beta(x, y)$ if $(x, y) \in \mathfrak{T}$ and $\tilde{\beta}(x, y) = 0$ if $(x, y) \in [a, b]^2 \setminus \mathfrak{T}$, in a series of tensor products and using equation (2). However, we expect the convergence to be slow and also the existence of Gibbs phenomena at the border of the triangle, more precisely, at all points (u, v) of the line segment joining (a, a) to (b, b) for which there exists $\epsilon > 0$, which may depend on (u, v) , such that for every $r > 0$ the Lebesgue measure of the set $\{(x, y) \in B((u, v), r) : |(\beta\chi_{\mathfrak{T}})(x, y)| > \epsilon\}$ is strictly positive. The issue here is that causality imposes the slope function to be identically zero on the interior of the upper triangular region with vertexes on points (a, a) , (b, b) and (a, b) .

It is possible to obtain estimators of causal functional linear models via finite elements methods as done in (Malfait & Ramsay, 2003) for the case of historical functional linear models.

The aim of this note is to show that we still can obtain good estimates using expansions in tensor products. This is made possible by the construction of a symmetric extension of the slope function and the use of an algebraic trick. No additional Gibbs phenomena is expected and convergence is faster than that associated to the direct expansion method, since now it will be similar to that associated to the application of the direct tensor product expansion method to estimate the slope function of a “standard” functional linear model, i.e., one whose slope function is defined on the whole square $[a, b]^2$ and does not present an abrupt behavior change between the lower and upper triangular regions. We remind that the direct estimation of the slope function using tensor product expansions for causal functional linear models was unsuitable because of the possible, very probable indeed, existence of a sharp discontinuity of the slope function at the hypotenuse of the triangular region \mathfrak{T} .

2. Estimator Construction

Let $\{\phi_i : i \in \mathcal{I}\}$ be an orthonormal basis of $L^2([a, b])$. We write

$$\alpha = \sum_{i \in \mathcal{I}} a_i \phi_i, \quad (3)$$

and

$$\beta\chi_{\mathfrak{T}} = \sum_{i, j \in \mathcal{I}} b_{ij} \phi_i \otimes \phi_j. \quad (4)$$

Now, define \mathfrak{B} as the symmetric extension of $\beta|_{\mathfrak{T}}$ to $[a, b]^2$, i.e., $\forall (s, t) \in \mathfrak{T}$, $\mathfrak{B}(t, s) = \mathfrak{B}(s, t) = \beta(s, t)\chi_{\mathfrak{T}}(s, t)$.

The expansion of \mathfrak{B} is written

$$\mathfrak{B} = \sum_{i, j \in \mathcal{I}} b_{ij} \phi_i \otimes \phi_j. \quad (5)$$

From the causal model definition we have

$$\begin{aligned} \mathcal{Y}(s) &= \alpha(s) + \int_a^s \beta(s, t)X(t)dt + \mathcal{E}(s) = \\ &= \alpha(s) + \int_a^b \mathfrak{B}(s, t)\chi_{\mathfrak{T}}(s, t)X(t)dt + \mathcal{E}(s), \end{aligned} \quad (6)$$

and, multiplying by $X(s)$, we obtain

$$\begin{aligned} \mathcal{Y}(s)X(s) &= \\ &= \alpha(s)X(s) + \int_a^b \mathfrak{B}(s, t)\chi_{\mathfrak{T}}(s, t)X(t)X(s)dt + X(s)\mathcal{E}(s). \end{aligned} \quad (7)$$

Integrating this equation for $s \in [a, b]$ gives

$$\begin{aligned} \int_a^b \mathcal{Y}(s)X(s)ds &= \int_a^b \alpha(s)X(s)ds + \\ &+ \int_a^b \int_a^b \mathfrak{B}(s, t)\chi_{\mathfrak{T}}(s, t)X(t)X(s)dt ds + \int_a^b X(s)\mathcal{E}(s)ds. \end{aligned} \quad (8)$$

The following simple fact will permit us to continue the calculations and make the desired tensor expansions:

$$\begin{aligned} \int_a^b \int_a^b \mathfrak{B}(s, t) \chi_{\mathfrak{T}}(s, t) \mathcal{X}(t) \mathcal{X}(s) dt ds = \\ \frac{1}{2} \int_a^b \int_a^b \mathfrak{B}(s, t) \mathcal{X}(t) \mathcal{X}(s) dt ds. \end{aligned} \quad (9)$$

Equation (8) now is written

$$\begin{aligned} \int_a^b \mathcal{Y}(s) \mathcal{X}(s) ds = \int_a^b \alpha(s) \mathcal{X}(s) ds + \\ \frac{1}{2} \int_a^b \int_a^b \mathfrak{B}(s, t) \mathcal{X}(t) \mathcal{X}(s) dt ds + \int_a^b \mathcal{X}(s) \mathcal{E}(s) ds. \end{aligned} \quad (10)$$

Now, writing $\mathcal{Y} = \sum_{i \in I} y_i \phi_i$ and $\mathcal{X} = \sum_{i \in I} x_i \phi_i$, we have

$$\begin{aligned} \int_a^b \mathcal{Y}(s) \mathcal{X}(s) ds = \int_a^b \sum_{i \in I} y_i \phi_i(s) \sum_{j \in I} x_j \phi_j(s) ds = \\ \sum_{i \in I} \sum_{j \in I} y_i x_j \int_a^b \phi_i(s) \phi_j(s) ds = \sum_{i \in I} \sum_{j \in I} y_i x_j \delta_{i,j} = \sum_{i \in I} y_i x_i. \end{aligned} \quad (11)$$

Analogously, we have

$$\int_a^b \alpha(s) \mathcal{X}(s) ds = \sum_{i \in I} a_i x_i, \quad (12)$$

and, expanding $\mathcal{E} = \sum_{i \in I} e_i \phi_i$,

$$\int_a^b \mathcal{E}(s) \mathcal{X}(s) ds = \sum_{i \in I} e_i x_i. \quad (13)$$

Now,

$$\begin{aligned} \int_a^b \int_a^b \mathfrak{B}(s, t) \mathcal{X}(t) \mathcal{X}(s) dt ds = \\ \int_a^b \int_a^b \sum_{i,j \in I} b_{ij} \phi_i(s) \phi_j(t) \sum_{k \in I} x_k \phi_k(t) \sum_{l \in I} x_l \phi_l(s) dt ds = \\ \sum_{i,j \in I} \sum_{k \in I} \sum_{l \in I} b_{ij} x_k x_l \int_a^b \int_a^b \phi_i(s) \phi_j(t) \phi_k(t) \phi_l(s) dt ds = \\ \sum_{i,j \in I} \sum_{k \in I} \sum_{l \in I} b_{ij} x_k x_l \int_a^b \phi_i(s) \phi_l(s) ds \int_a^b \phi_j(t) \phi_k(t) dt = \\ \sum_{i,j \in I} \sum_{k \in I} \sum_{l \in I} b_{ij} x_k x_l \delta_{i,l} \delta_{j,k} = \sum_{i,j \in I} b_{ij} x_j x_i. \end{aligned} \quad (14)$$

Substituting (11), (12), (13), and (14) in (10) we get

$$\sum_{i \in I} y_i x_i = \sum_{i \in I} a_i x_i + \frac{1}{2} \sum_{i,j \in I} b_{ij} x_j x_i + \sum_{i \in I} e_i x_i. \quad (15)$$

Note that, from (15) we have

$$\left| \sum_{i \in I} y_i x_i - \sum_{i \in I} a_i x_i - \frac{1}{2} \sum_{i,j \in I} b_{ij} x_j x_i \right| = \left| \sum_{i \in I} e_i x_i \right| \leq \sqrt{\sum_{i \in I} e_i^2} \sqrt{\sum_{i \in I} x_i^2}. \quad (16)$$

Equivalently, if $X \neq 0$, we write

$$\frac{\left(\sum_{i \in I} y_i x_i - \sum_{i \in I} a_i x_i - \frac{1}{2} \sum_{i,j \in I} b_{ij} x_j x_i \right)^2}{\sum_{i \in I} x_i^2} \leq \sum_{i \in I} e_i^2. \quad (17)$$

Let us write the sample $\{\mathcal{Y}_k, \mathcal{X}_k\}$, and noise \mathcal{E}_k as

$$\mathcal{Y}_k = \sum_{i \in I} y_{ki} \phi_i, \quad \mathcal{X}_k = \sum_{i \in I} x_{ki} \phi_i,$$

and

$$\mathcal{E}_k = \sum_{i \in I} e_{ki} \phi_i.$$

Now, using (17) for each k and summing over k for $1 \leq k \leq N$, we get

$$\sum_{k=1}^N \frac{\left(\sum_{i \in I} y_{ki} x_{ki} - \sum_{i \in I} a_i x_{ki} - \frac{1}{2} \sum_{i,j \in I} b_{ij} x_{kj} x_{ki} \right)^2}{\sum_{i \in I} x_{ki}^2} \leq \sum_{k=1}^N \sum_{i \in I} e_{ki}^2 = \sum_{k=1}^N \|\mathcal{E}_k\|_2^2. \quad (18)$$

Note that inequality (18) must be fulfilled whatever the noise energy is. We do not know how high this energy is. Thus, to guarantee its fulfillment, we will minimize its left hand side. This is similar to minimizing the energy of the residuals when fitting a regression model. This leads us to seek the minimum of

$$\mathcal{F}(\vec{a}, \vec{b}) = \sum_{k=1}^N \frac{\left(\sum_{i \in I} y_{ki} x_{ki} - \sum_{i \in I} a_i x_{ki} - \frac{1}{2} \sum_{i,j \in I} b_{ij} x_{kj} x_{ki} \right)^2}{\sum_{i \in I} x_{ki}^2} \quad (19)$$

with respect to a_i and b_{ij} , where $\vec{a} = (a_i)_{i \in I} \in \ell^2(I)$ and $\vec{b} = (b_{ij})_{(i,j) \in I^2} \in \ell^2(I^2)$.

Letting $\frac{\partial}{\partial a_i} \mathcal{F}(\vec{a}, \vec{b}) = 0$ and $\frac{\partial}{\partial b_{ij}} \mathcal{F}(\vec{a}, \vec{b}) = 0$ we obtain the following system of equations:

For all $l \in I$,

$$\sum_{k=1}^N \frac{\left(\sum_{i \in I} y_{ki} x_{ki} - \sum_{i \in I} a_i x_{ki} - \frac{1}{2} \sum_{i,j \in I} b_{ij} x_{kj} x_{ki} \right) x_{kl}}{\sum_{i \in I} x_{ki}^2} = 0 \quad (20)$$

and, for all $(l, m) \in I^2$,

$$\sum_{k=1}^N \frac{\left(\sum_{i \in I} y_{ki} x_{ki} - \sum_{i \in I} a_i x_{ki} - \frac{1}{2} \sum_{i,j \in I} b_{ij} x_{kj} x_{ki} \right) x_{kl} x_{km}}{\sum_{i \in I} x_{ki}^2} = 0 \quad (21)$$

Since for all $(\vec{c}, \vec{d}) \in \ell^2(I) \times \ell^2(I^2)$ we have

$$\begin{aligned} \frac{d^2}{dt^2} \mathcal{F}(\vec{a}, \vec{b}) + t(\vec{c}, \vec{d}) &= \\ 2 \sum_{k=1}^N \frac{\left(\sum_{i \in I} c_i x_{ki} + \frac{1}{2} \sum_{i,j \in I} d_{ij} x_{kj} x_{ki} \right)^2}{\sum_{i \in I} x_{ki}^2} &\geq 0, \end{aligned} \quad (22)$$

every solution of the simultaneous system of equations (20) and (21) is a point of local minimum of \mathcal{F} . Moreover, since (22) is true for every $(\vec{a}, \vec{b}) \in \ell^2(I) \times \ell^2(I^2)$, every solution of this system of equations is also a global minimum of \mathcal{F} .

Now, observe that I is an enumerable infinite set and both (20) and (21) have infinitely many equations and unknowns. In general, for every finite sample size N , the solutions to (20) and (21) will yield estimators for α and \mathfrak{B} that perfectly fit the data. See (Ramsay & Silverman, 2005). In practice, we will consider a finite subset \mathfrak{F} of I to perform the expansions of the covariates, the responses and the parameters. This will lead to the finite version of equations (20) and (21) from which we will obtain our estimators \hat{a}_i and \hat{b}_{ij} of a_i and b_{ij} . Observe that, in general, these estimators depend on the choice of $\mathfrak{F} \subset I$. They are based on the *approximate* relation

$$\sum_{i \in \mathfrak{F}} y_i x_i \approx \sum_{i \in \mathfrak{F}} a_i x_i + \frac{1}{2} \sum_{i, j \in \mathfrak{F}} b_{ij} x_j x_i + \sum_{i \in \mathfrak{F}} e_i x_i, \quad (23)$$

which comes from (15).

The estimators of the parameters α and \mathfrak{B} will be given by

$$\hat{\alpha} = \sum_{i \in \mathfrak{F}} \hat{a}_i \phi_i \quad (24)$$

and

$$\hat{\mathfrak{B}} = \sum_{i, j \in \mathfrak{F}} \hat{b}_{ij} \phi_i \otimes \phi_j \quad (25)$$

where \hat{a}_i and \hat{b}_{ij} are the solutions of the finite system of equations:

For all $l \in \mathfrak{F}$,

$$\sum_{k=1}^N \frac{(\sum_{i \in \mathfrak{F}} y_{ki} x_{ki} - \sum_{i \in \mathfrak{F}} \hat{a}_i x_{ki} - \frac{1}{2} \sum_{i, j \in \mathfrak{F}} \hat{b}_{ij} x_{kj} x_{ki}) x_{kl}}{\sum_{i \in \mathfrak{F}} x_{ki}^2} = 0 \quad (26)$$

and for all $(l, m) \in \mathfrak{F}^2$,

$$\sum_{k=1}^N \frac{(\sum_{i \in \mathfrak{F}} y_{ki} x_{ki} - \sum_{i \in \mathfrak{F}} \hat{a}_i x_{ki} - \frac{1}{2} \sum_{i, j \in \mathfrak{F}} \hat{b}_{ij} x_{kj} x_{ki}) x_{kl} x_{km}}{\sum_{i \in \mathfrak{F}} x_{ki}^2} = 0. \quad (27)$$

Observe that this is a linear system in \hat{a}_i and \hat{b}_{ij} . If the cardinality of \mathfrak{F} is p then this is a system of $p^2 + p$ equations on $p^2 + p$ unknowns.

Finally, the estimated model will be written

$$\mathcal{Y}(s) = \sum_{i \in \mathfrak{F}} \hat{a}_i \phi_i(s) + \int_a^s \sum_{i, j \in \mathfrak{F}} \hat{b}_{ij} \phi_i(s) \phi_j(t) \mathcal{X}(t) dt + \mathcal{E}(s) \quad (28)$$

or, equivalently,

$$\mathcal{Y}(s) = \sum_{i \in \mathfrak{F}} \hat{a}_i \phi_i(s) + \int_a^b \sum_{i, j \in \mathfrak{F}} \hat{b}_{ij} \phi_i(s) \phi_j(t) \chi_{\mathfrak{T}}(s, t) \mathcal{X}(t) dt + \mathcal{E}(s) \quad (29)$$

3. Final Remarks

The linearity of systems (26) and (27) is of great importance for computational reasons.

The choice of the basis is an important practical issue. As a general guideline, the choice of Fourier sine cosine basis is recommended when we have non localized frequency behavior of covariates and responses, and wavelet basis shall be chosen for the localized frequency covariates and responses. The case where we have opposite localized frequency behaviors of covariates and responses is probably the most difficult to deal with.

Choosing a convenient finite linearly independent orthonormal set $\{\phi_i : i \in \mathfrak{F} \subset I\}$ to expand the data and the functional parameters will furnish, for suitable bases, smooth estimates of the parameters for the Causal Linear Functional Model. This is a kind of regularization that is obtained by the simple reduction the number of basis functions used in the estimation process. See (Ramsay & Silverman, 2005). Other methods for regularization are available, such as the use of penalties given by some functional applied to the parameters, and the use of threshold techniques on the set of estimated coefficients

for the intercept and slope functions. The weighted sum of the integrated square of some measure of curvature for both the mean and the slope functions can be used as a penalty functional.

In this work we focused on *constructing* suitable estimators of the functional parameters of causal functional linear regression models using series expansions in a general tensor product basis. More detailed studies of these estimators is the subject of future works.

Acknowledgements

The author thanks NSSJC.

References

- Dabo-Niang, S. & Ferraty, F. (Editors). (2008). *Functional and Operatorial Statistics, Contributions to Statistics*. Heidelberg: Physica-Verlag, a Springer company.
- Cardot, H. & Sarda, P. *Varying-Coefficient Functional Linear Regression Models*. <http://cardot.perso.math.cnrs.fr>
- Ferraty, F. (Editor). (2011). *Recent Advances in Functional Data Analysis and Related Topics, Contributions to Statistics*. Berlin-Heidelberg: Physica-Verlag, Springer. <https://doi.org/10.1007/978-3-7908-2736-1>
- Ferraty, F. & Vieu, P. (2006). *Nonparametric functional data analysis*. New York: Springer-Verlag.
- Horváth, L. & Kokoszka, P. (2012). *Inference for Functional Data with Applications*. New York: Springer. <https://doi.org/10.1007/978-1-4614-3655-3>
- Malfait, N. & Ramsay, J. O. (2003). The historical functional linear model. *The Canadian Journal of Statistics*, 31(2), 115 - 128.
- Müller H-G. & Zhang, Y. (2005). Time-Varying Functional Regression for Predicting Remaining Lifetime Distributions from Longitudinal Trajectories. *Biometrics*, 61(4), 1064 - 1075. <https://doi.org/10.1111/j.1541-0420.2005.00378.x>
- Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*. (2nd ed.). New York: Springer- Verlag.
- Ramsay, J. O. & Silverman, B. W. (2002). *Applied Functional Data Analysis*. New York: Springer- Verlag. <https://doi.org/10.1002/0470013192.bsa239>
- Wu, Y. , Fan, J., & Müller H-G. (2010). Varying-coefficient functional linear regression. *Bernoulli*, 16(3), 730 - 758. <https://doi.org/10.3150/09-BEJ231>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).