Journal of Econometrics 201 (2017) 269–291

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Functional linear regression with functional response[☆]

David Benatia [a], Marine Carrasco [a,*], Jean-Pierre Florens [b]

[a] Université de Montréal, Canada
[b] Toulouse School of Economics, France

ARTICLE INFO

Article history:
Available online 19 August 2017

JEL classification:
C13

Keywords:
Functional regression
Instrumental variables
Linear operator
Tikhonov regularization

ABSTRACT

In this paper, we develop new estimation results for functional regressions where both the regressor $Z(t)$ and the response $Y(t)$ are functions of Hilbert spaces, indexed by the time or a spatial location. The model can be thought as a generalization of the multivariate regression where the regression coefficient is now an unknown operator Π. We propose to estimate the operator Π by Tikhonov regularization, which amounts to apply a penalty on the L^2 norm of Π. We derive the rate of convergence of the mean-square error, the asymptotic distribution of the estimator, and develop tests on Π. As trajectories are often not fully observed, we consider the scenario where the data become more and more frequent (infill asymptotics). We also address the case where Z is endogenous and instrumental variables are used to estimate Π. An application to the electricity consumption completes the paper.

1. Introduction

With the increase of storage capability, continuous time data are available in many fields including finance, medicine, meteorology, and microeconometrics. Researchers, companies, and governments look for ways to exploit this rich information. In this paper, we develop new estimation results for functional regressions where both the regressor $Z(t)$ and the response $Y(t)$ are functions of an index such as the time or a spatial location. Both $Z(t)$ and $Y(t)$ are assumed to belong to Hilbert spaces. The model can be thought as a generalization of the multivariate regression where the regression coefficient is now an unknown operator Π. An interesting feature of our model is that $Y(t)$ depends not only on contemporaneous $Z(t)$ but also on past and future values of Z.

We propose to estimate the operator Π by Tikhonov regularization, which amounts to apply a penalty on the L^2 norm of Π. The choice of a L^2 penalty, instead of L^1 used in Lasso, is motivated by the fact that – in the applications we have in mind – there is no reason to believe that the relationship between Y and Z is sparse. We derive the rate of convergence of the mean-square error (MSE) and the asymptotic distribution of the estimator for a fixed regularization parameter α and develop tests on Π. In some applications, it would be interesting to test whether $Y(t)$ depends only on the past values of Z or only on contemporaneous values of Z. If the application is on network and t refers to the spatial location, our model could describe how the behavior of a firm $Y(t)$ depends on the decision of neighboring firms $Z(s)$. Testing properties of Π will help to characterize the strategic response of firms.

Often, the full trajectories are not observed but only a discretized version is available. This case raises specific challenges which will be addressed in the scenario where the data become more and more frequent (infill asymptotics).

We also consider the case where Z is endogenous and instrumental variables are used to estimate Π. To the best of our knowledge, the model with functional response and endogenous functional regressor has never been studied before. We derive an estimator based on Tikhonov regularization and show its rate of convergence.

There is a large body of work done on linear functional regression where the response is a scalar variable Y and the regressor is a function. Some recent references include Cardot et al. (2003), Hall and Horowitz (2007), Horowitz and Lee (2007), Darolles et al. (2011), Crambes et al. (2009) and Florens and Van Bellegem (2015). In contrast, only a few researchers have tackled the functional linear regression in which both the predictor Z and the response Y are random functions. The object of interest is the estimation of the conditional expectation of Y given Z. In this setting, the unknown parameter is an integral operator. This model is discussed in the monographs by Ramsay and Silverman (2005), Ferraty and Vieu (2006) and Horvath and Kokoszka (2012). Cuevas et

☆ The authors thank the editors, two referees, Stéphane Bonhomme, Samuele Centorrino, Jan Johannes, and the participants of the 6th French Econometrics Conférence, CESG 2016, and the Workshop on Inverse Problems (Heidelberg, 2016) for helpful comments. Carrasco gratefully acknowledges financial support from NSERC grant RGPIN-2015-03798.

* Correspondence to: Université de Montréal, Département de Sciences Economiques, CP 6128, succ Centre Ville, Montréal, QC H3C3J7, Canada.

E-mail addresses: david.benatia@umontreal.ca (D. Benatia), marine.carrasco@umontreal.ca (M. Carrasco), jean-pierre.florens@tse-fr.eu (J.-P. Florens).

http://dx.doi.org/10.1016/j.jeconom.2017.08.008
0304-4076/

al. (2002) consider a fixed design setting and propose an estimator of Π based on interpolation. Yao et al. (2005) consider the case where both predictor and response trajectories are observed at discrete and irregularly spaced times. Their estimator is based on nonparametric estimators of the principal components. Park and Qian (2012) use functional principal components to estimate a regression where both the response and independent variables are densities. Crambes and Mas (2013) consider also a spectral cut-off regularized inverse and derive the asymptotic mean square prediction error which is then used to derive the optimal choice of the regularization parameter. Our model is also related to the functional autoregressive (FAR) model studied by Bosq (2000), Kargin and Onatski (2008) and Aue et al. (2015), among others. The estimation methods used in these papers are based on functional principal components and differ from ours. Antoch et al. (2010) use a FAR to forecast the electricity consumption. In their model, the weekday consumption curve is explained by the curve from the previous week. The authors use B-spline to estimate the operator.

Our contribution is to develop an estimator which can be expressed as products of matrices and vectors. It does not require choosing a basis or estimating eigenfunctions. It involves only one smoothing parameter. We derive the rate of convergence of our estimator under some source condition which is common in the inverse problem literature. Interestingly, we do not need to impose any restriction on the multiplicity of eigenvalues.

The paper is organized as follows. Section 2 introduces the model and the estimators. Section 3 derives the rate of convergence of the MSE. Section 4 presents the asymptotic normality of the estimator for a fixed regularization parameter. Issues relative to the choice of the regularization parameter are discussed in Section 5. Discrete observations are addressed in Section 6. Section 7 considers an endogenous regressor. Section 8 presents simulation results. Section 9 presents an application to the electricity market where the dependent variable is the electricity consumption and the independent variable is the temperature. The proofs are collected in Appendix. Data and Matlab programs used in this paper are available at: https://davidbenatia.wordpress.com/.

2. The model and estimator

2.1. The model

We consider a regression model where both the predictor and response are random functions. We observe pairs of random trajectories (y_i, z_i) $i = 1, 2, \ldots, n$ with square integrable predictor trajectories z_i and response trajectories y_i. They are realizations of random processes (Y, Z) with zero mean functions and unknown covariance operators. The extension to the case, where the mean is unknown but estimated, is straightforward. The arguments of Y and Z are denoted t which may refer to the time, a location or a characteristic such as the age or income of an agent.

We assume that Y belongs to a separable real Hilbert space $\mathcal{E}$ equipped with an inner product $\langle , \rangle$ and Z belongs to a separable real Hilbert space $\mathcal{F}$ equipped with an inner product $\langle , \rangle$ (to simplify notations, we use the same notation for both inner products even though they usually differ).

The model is

$$Y = \Pi Z + U \tag{1}$$

where U is a zero mean random element of $\mathcal{E}$ and Π is a nonrandom Hilbert–Schmidt[1] operator from $\mathcal{F}$ to $\mathcal{E}$. The regressor Z is said to be exogenous if $cov\,(Z, U) = 0$. Z will be assumed exogenous in Sections 2–6. This assumption will be relaxed in Section 7.

[1] K is Hilbert–Schmidt if $\sum_j \langle K\phi_j, K\phi_j\rangle < \infty$ for any basis (ϕ_j).

For illustration, consider the following example

$$\mathcal{E} = \left\{ g : \int_{\mathcal{S}} g(t)^2 dt < \infty \right\},$$
$$\mathcal{F} = \left\{ f : \int_{\mathcal{T}} f(t)^2 dt < \infty \right\}$$

where $\mathcal{S}$ and $\mathcal{T}$ are some intervals of $\mathbb{R}$. An operator Π from $\mathcal{E}$ to $\mathcal{F}$ is a Hilbert–Schmidt if and only if it admits a representation as an integral operator such that

$$(\Pi\varphi)(s) = \int_{\mathcal{T}} \pi(s,t)\,\varphi(t)\,dt$$

for any $\varphi \in \mathcal{F}$ and

$$\int_{\mathcal{S}} \int_{\mathcal{T}} |\pi(s,t)|^2 dtds < \infty.$$

π is referred to as the kernel of the operator Π. Note that our theory covers the case where $\mathcal{S}$ and $\mathcal{T}$ are subsets of $\mathbb{R}^p$ for $p > 1$ so that the index s and t may be vectors. Moreover, we are not limited to L^2 spaces.

Consider the case where $\mathcal{E}$ and $\mathcal{F}$ are Sobolev spaces:

$$\mathcal{E} = \left\{ g \in \mathcal{C}^1(\mathcal{S}) : \int_{\mathcal{S}} g(t)^2 dt + \int_{\mathcal{S}} g(t)'^2 dt < \infty \right\},$$
$$\mathcal{F} = \left\{ f \in \mathcal{C}^1(\mathcal{T}) : \int_{\mathcal{T}} f(t)^2 dt + \int_{\mathcal{S}} f(t)'^2 dt < \infty \right\}$$

where $\mathcal{C}^1(\mathcal{S})$ denotes the space of all complex valued functions defined on $\mathcal{S}$ with continuous first derivatives. $\mathcal{E}$ is a Hilbert space with inner product $\langle g_1, g_2\rangle = \int_{\mathcal{S}} \left(g_1(t)\,g_2(t) + g_1'(t)\,g_2'(t)\right) dt$. The same is true for $\mathcal{F}$. Let Π be an integral operator from $\mathcal{F}$ to $\mathcal{E}$, $(\Pi\varphi)(s) = \int_{\mathcal{T}} \pi(s,t)\,\varphi(t)\,dt$ for all $\varphi \in \mathcal{F}$. Then Π is a Hilbert–Schmidt operator if and only if

$$\int_{S}\int_{\mathcal{T}} \left(|\pi|^2 + \left|\frac{\partial\pi}{\partial s}\right|^2 + \left|\frac{\partial\pi}{\partial t}\right|^2 + \left|\frac{\partial^2\pi}{\partial s\partial t}\right|^2 \right) dtds < \infty.$$

Our theory covers the case of Sobolev spaces, the main difference with L^2 spaces is the way the inner products are defined.

The main model we have in mind is the model where Π is an integral operator and (1) takes the form:

$$Y(s) = \int_{\mathcal{T}} \pi(s,t)\,Z(t)\,dt + U(s).$$

In this model, $Y(s)$ depends not only on $Z(s)$ but also on all the $Z(t)$, for $t \neq s$. The object of interest is the estimation of the operator Π using a panel of observations (Y_i, Z_i), $i = 1, 2, \ldots, n$. Below, we describe three examples of applications of Model (1).

Example of application 1. Electricity market.

Let $Y_i(t)$ be the electricity consumption for the province of Ontario in Canada for day i at hour t, and $Z_i(t)$ be the average temperature for the same province for day i at hour t. Recent research (McLaughlin et al., 2011; Yu et al., 2013) reveals that the optimal consumption path of electricity over one day may depend not only on the current temperature but also on past and future temperatures. We could model this relationship as a multivariate regression where the dependent variable is the 24×1 vector of electricity consumptions at different hours of the day and the independent variable is a 72×1 vector of temperatures. Estimating this model by ordinary least squares (OLS) would require inverting a 72×72 matrix which would likely be near singular. The variance of the resulting OLS estimator would be very large. Moreover, if the frequency of the observations increases, it is expected that the successive observations will become more and more correlated. It is therefore natural to assume that the observations result from the discretization of a curve. Model (1) seems to be an attractive alternative to the multivariate regression in this setting. This application will be continued in Section 9.

Example of application 2. Technological spillovers in productivity.

There has been an increasing interest in the economic literature on quantifying interactions and spillovers among firms (see Manresa, 2015 and references therein). Assume that $Y(s)$ is the log of the average output of firms with industry code s and $Z(t)$ is the log of the average R&D expenditures for firms with industry code t. Model (1) would permit to characterize how firms from one sector benefit from R&D advancements done in adjacent sectors. From a practical point of view, note that industry code goes from 1 to 6 digits. A higher number of digits correspond to a thinner grid on the continuum of industry codes. To obtain an index between 0 and 1, we normalize the industry codes t by dividing them by the largest existing industry code with the same number of digits. As in Manresa (2015), we suggest using repeated observations over time to estimate the model, so that the index i corresponds to a time period. This example will not be pursued here.

Example of application 3. Risk neutral density.[2]

The risk neutral density (RND) plays a crucial role in the pricing of financial derivatives. Let S_t denote the value of an asset at time t. The price $g_t(K,T)$ of a European option with the strike price K and the maturity T is equal to the expected present value of the future payoff $g(S_T,K)$ under the risk-neutral density:

$$g_t(K,T) = \frac{1}{R_{t,T}} \int_0^{\infty} g(S_T,K) f(S_T,T;S_t,t)\,dS_T \tag{2}$$

where $f(S_T,T;S_t,t)$ is the conditional RND and $R_{t,T} = e^{\int_t^T r_s ds}$ where r_t is the risk-free rate. For a European call option, we have

$$g_t(K,T) = C_t(K,T),$$
$$g(S_T,K) = \max(S_T - K, 0).$$

It is well known that (2) does not hold exactly because of measurement errors and the incompleteness of the market (see Gouriéroux and Jasiak, 2001, p.322), therefore we add an error term to (2) and obtain an equation equivalent to (4). Using a set of call prices with different strike prices K and maturity times T, we can write (2) as

$$Y_t(T,S_t) = \int_0^{\infty} \pi(T,S_t,s) Z(s)\,ds + U(T,S_t) \tag{3}$$

where $Y_t(T,S_t) = R_{t,T} C_t(K,T)$, $\pi(T,S_t,s) = f(s,T;S_t,t)$, $Z(s) = \max(s-K,0)$. To estimate this model, one could use the same data as in Ait-Sahalia and Lo (1998). They focus on S&P 500 index options traded at the Chicago Board Options Exchange. They use price data on every option traded during 1993, which results in a sample of 14,431 options after cleaning the data. By observing so many options and maturity times, we should be able to recover most of the supports of S_t and T. Note that the first index in π is bidimensional (T,S_t). Our theory covers this case. Here $Z(s)$ is deterministic, so we are in a fixed design setting. The error term U may be correlated across observations but this issue is not addressed in the present paper.[3]

Many papers have proposed nonparametric estimators of the RND before, among others Jackwerth and Rubinstein (1996), Ait-Sahalia and Lo (1998), Garcia and Gençay (2000), Ait-Sahalia et al. (2001) and Bondarenko (2003). Most of these papers start with the relation

$$f(s,T;S_t,t) = \frac{1}{R_{t,T}} \frac{\partial^2 C_t}{\partial x^2}\bigg|_{x=S_T}$$

and then rely on some smoothing. Here, we propose to use regularization to deal with the ill-posedness of the inverse problem (2) directly.

2.2. The estimator

We denote V_Z the operator from $\mathcal{F}$ to $\mathcal{F}$ which associates to functions $\varphi \in \mathcal{F}$:

$$V_Z\varphi = E[Z\langle Z,\varphi\rangle].$$

Note that, as Z is centered, V_Z is the covariance operator of Z. We denote C_{YZ} the covariance operator of (Y,Z). It is the operator from $\mathcal{F}$ to $\mathcal{E}$ such that

$$C_{YZ}\varphi = E[Y\langle Z,\varphi\rangle]$$

Using (1), we have

$$\begin{aligned} cov(Y,Z) &= cov(\Pi Z + U, Z) \\ &= \Pi cov(Z,Z) + cov(U,Z). \end{aligned}$$

Hence, we have the following relationships:

$$C_{YZ} = \Pi V_Z, \tag{4}$$

$$C_{ZY} = V_Z \Pi^* \tag{5}$$

where Π^* is the adjoint of Π and C_{ZY} is defined as the operator from $\mathcal{E}$ to $\mathcal{F}$ such that

$$C_{ZY}\psi = E[Z\langle Y,\psi\rangle]$$

for any ψ in $\mathcal{E}$. Note that C_{ZY} is the adjoint of C_{YZ}, C_{YZ}^*.

First we describe how to estimate Π^* using (5). The unknown operators V_Z and C_{ZY} are replaced by their sample counterparts. The sample estimate of V_Z is

$$\hat{V}_Z\varphi = \frac{1}{n}\sum_{i=1}^{n} z_i\langle z_i,\varphi\rangle$$

for $\varphi \in \mathcal{F}$. The sample estimate of C_{ZY} is

$$\hat{C}_{ZY}\psi = \frac{1}{n}\sum_{i=1}^{n} z_i\langle y_i,\psi\rangle$$

for $\psi \in \mathcal{E}$. An estimator of Π^* cannot be obtained directly by solving $\hat{C}_{ZY} = \widehat{V}_Z\Pi^*$ because the initial equation $C_{ZY} = V_Z\Pi^*$ is an ill-posed problem in the sense that V_Z is invertible only on a subset of $\mathcal{E}$ and its inverse is not continuous. Note that $\hat{V}_Z$ has finite rank equal to n and hence is not invertible. A Moore–Penrose generalized inverse would not help because it would not be continuous. To stabilize the inverse, we need to use some regularization scheme. We adopt Tikhonov regularization (see Kress, 1999; Carrasco et al., 2007).

The estimator of Π^* is defined as

$$\hat{\Pi}^*_\alpha = \left(\alpha I + \hat{V}_Z\right)^{-1}\hat{C}_{ZY} \tag{6}$$

and that of Π is defined by

$$\hat{\Pi}_\alpha = \hat{C}_{YZ}\left(\alpha I + \hat{V}_Z\right)^{-1} \tag{7}$$

where α is some positive regularization parameter which will be allowed to converge to zero as n goes to infinity. The estimators (6) and (7) can be viewed as generalization of ordinary least-squares estimators. They also have an interpretation as the solution to an inverse problem.

[2] We thank a referee for suggesting this example.

[3] Another issue is that the data may not be stationary because the stock price S_t is not stationary. One way to make it stationary is to divide by the strike price K, see Garcia et al. (2010).

2.3. Link with inverse problems

At this stage, it is useful to make the link with the inverse problem literature. Let $\mathcal{H}$ be the Hilbert space of linear Hilbert–Schmidt operators from $\mathcal{F}$ to $\mathcal{E}$. The inner product on $\mathcal{H}$ (see Pedersen, 1989) is

$$\langle \Pi_1, \Pi_2 \rangle_{\mathcal{H}} = tr\left(\Pi_1^* \Pi_2\right) = tr\left(\Pi_2^* \Pi_1\right).$$

Dropping the error term in (1), we obtain, for the sample, the equation

$$\hat{r} = K\Pi \tag{8}$$

where $\hat{r} = (y_1, \ldots, y_n)'$ and K is the operator from $\mathcal{H}$ to $\mathcal{E}^n$ such that $K\Pi = (\Pi z_1, \ldots, \Pi z_n)'$. The inner product on $\mathcal{E}^n$ is

$$\langle f, g \rangle_{\mathcal{E}^n} = \frac{1}{n} \sum_{i=1}^{n} \langle f_i, g_i \rangle_{\mathcal{E}}$$

with $f = (f_1, \ldots, f_n)'$ and $g = (g_1, \ldots, g_n)'$. Let us check that $\hat{\Pi}_\alpha$ is a classical Tikhonov regularized inverse of the operator K:

$$\hat{\Pi}_\alpha = \left(\alpha I + K^* K\right)^{-1} K^* \hat{r}.$$

We need to find K^*. We look for the operator B from $\mathcal{F}$ to $\mathcal{E}$ solution of

$$\langle K\Pi, f \rangle_{\mathcal{E}^n} = \langle \Pi, B \rangle_{\mathcal{H}}. \tag{9}$$

Note that

$$\begin{aligned} \langle \Pi, B \rangle_{\mathcal{H}} &= tr\left(B^* \Pi\right) \\ &= \sum_j \left\langle B^* \Pi \varphi_j, \varphi_j \right\rangle_{\mathcal{F}} \\ &= \sum_j \left\langle \Pi \varphi_j, B \varphi_j \right\rangle_{\mathcal{E}} \end{aligned}$$

where $\left(\varphi_j\right)$ is an orthonormal basis of $\mathcal{F}$. On the other hand,

$$\langle K\Pi, f \rangle_{\mathcal{E}^n} = \frac{1}{n} \sum_i \langle \Pi z_i, f_i \rangle_{\mathcal{E}}.$$

Using $z_i = \sum_j \left\langle z_i, \varphi_j \right\rangle \varphi_j$, we obtain

$$\begin{aligned} \langle K\Pi, f \rangle_{\mathcal{E}^n} &= \frac{1}{n} \sum_i \left\langle \sum_j \left\langle z_i, \varphi_j \right\rangle \Pi \varphi_j, f_i \right\rangle_{\mathcal{E}} \\ &= \sum_j \left\langle \Pi \varphi_j, \frac{1}{n} \sum_i \left\langle z_i, \varphi_j \right\rangle f_i \right\rangle_{\mathcal{E}}. \end{aligned}$$

It follows from (9) that $B\varphi_j = \frac{1}{n} \sum_i \left\langle z_i, \varphi_j \right\rangle f_i$ for all j and hence

$$B\varphi = \left(K^* f\right) \varphi = \frac{1}{n} \sum_i \langle z_i, \varphi \rangle f_i$$

for all φ in $\mathcal{E}$.

Now, we have

$$K^* K \Pi = \frac{1}{n} \sum_i \langle z_i, \varphi \rangle \Pi z_i = \Pi \widehat{V}_Z$$

and

$$K^* \hat{r} = \frac{1}{n} \sum_i \langle z_i, . \rangle y_i = \hat{C}_{YZ}.$$

It follows that

$$\begin{aligned} \hat{\Pi}_\alpha &= \left(\alpha I + K^* K\right)^{-1} K^* \hat{r} \\ &= \hat{C}_{YZ} \left(\alpha I + \hat{V}_Z\right)^{-1}. \end{aligned}$$

In summary, we established that $\hat{\Pi}_\alpha$ is the Tikhonov regularized inverse of K in Eq. (8).

The estimator $\hat{\Pi}_\alpha$ is also a penalized least-squares estimator:

$$\begin{aligned} \hat{\Pi}_\alpha &= \arg\min_{\Pi} \sum_{i=1}^{n} \|y_i - \Pi z_i\|^2 + \alpha \|\Pi\|_{HS}^2 \\ &= \arg\min_{\Pi} \sum_{i=1}^{n} \|y_i - \Pi z_i\|^2 + \alpha \sum \tilde{\mu}_j^2 \end{aligned}$$

where $\tilde{\mu}_j$ are the singular values of the operator Π.

2.4. Identification

It is easier to study identification from the viewpoint of Eq. (5). Let $\mathcal{H}$ be the space of Hilbert–Schmidt operators from $\mathcal{E}$ to $\mathcal{F}$. Let T be the operator from $\mathcal{H}$ to $\mathcal{H}$ defined as

$$TH = V_Z H \text{ for } H \text{ in } \mathcal{H}.$$

According to (5), Π^* is identified if and only if T is injective.

V_Z injective implies T injective. Indeed, we have

$$\begin{aligned} TH &= 0 \\ &\Leftrightarrow V_Z H = 0 \\ &\Leftrightarrow V_Z H \psi = 0, \forall \psi \\ &\Leftrightarrow H \psi = 0, \forall \psi \end{aligned}$$

by the injectivity of V_Z. Hence $H = 0$. It turns out that T is injective if and only if V_Z is injective. This can be shown by deriving the spectrum of T.

First, we show that T is self-adjoint. The adjoint T^* of T satisfies

$$\langle TH, K \rangle = \left\langle H, T^* K \right\rangle$$

for arbitrary operators H and K of $\mathcal{H}$. We have

$$\begin{aligned} \langle TH, K \rangle &= tr\left((TH)^* K\right) \\ &= tr\left((V_Z H)^* K\right) \\ &= tr\left(H^* V_Z K\right). \end{aligned}$$

Hence, $T^* K = V_Z K = TK$. Therefore, T is self-adjoint.

The spectrum of T is also closely related to that of V_Z. Let $\left(\mu_j, H_j\right)_{j=1,2,\ldots}$ denote the eigenvalues and eigenfunctions of T and $\left(\lambda_j, \varphi_j\right)_{j=1,2,\ldots}$ be the eigenvalues and eigenfunctions of V_Z so that $V_Z \varphi_j = \lambda_j \varphi_j$. H_j is necessarily of the form, $H_j = \varphi_j \langle \iota, . \rangle$ where ι is the 1 function in $\mathcal{E}$. Then,

$$\begin{aligned} TH_j &= V_Z \varphi_j \langle \iota, . \rangle \\ &= \lambda_j \varphi_j \langle \iota, . \rangle \\ &= \lambda_j H_j. \end{aligned}$$

So that the eigenvalues of T are the same as those of V_Z.

In summary, a necessary and sufficient condition for the identification of Π is that V_Z is injective. When V_Z is not injective, our estimator will typically converge to $\Pi_{N^\perp}$, the projection of Π on the orthogonal to the null space of T. This case is discussed in Section 3.

An example of non injective V_Z is given by an integral operator with degenerate kernel:

$$\left(V_Z \varphi\right)(s) = \int_0^1 v(s, t)\, \varphi(t)\, dt$$

where

$$v(s, t) = \sum_{j=1}^{J} a_j(s)\, b_j(t)$$

for some finite J and a_j linearly independent. Then, φ belongs to the null space of V_Z if and only if $\int_0^1 b_j(t)\varphi(t)\,dt = 0$. 0 is an eigenvalue of V_Z associated with the elements of the null space of V_Z, hence V_Z is not injective.

2.5. Computation of the estimator

To show how to compute $\hat{\Pi}_\alpha^*$ explicitly, we multiply the left and right hand sides of (6) by $\left(\alpha I + \hat{V}_Z\right)$ to obtain

$$\hat{C}_{ZY}\psi = \left(\alpha I + \hat{V}_Z\right)\hat{\Pi}_\alpha^*\psi \Leftrightarrow$$

$$\frac{1}{n}\sum_{i=1}^n z_i\langle y_i, \psi\rangle = \alpha\hat{\Pi}_\alpha^*\psi + \frac{1}{n}\sum_{i=1}^n z_i\left\langle z_i, \hat{\Pi}_\alpha^*\psi\right\rangle. \quad (10)$$

Then, we take the inner product with $z_l, l = 1, 2, \ldots, n$ on the left and right hand side of (10), to obtain n equations:

$$\frac{1}{n}\sum_{i=1}^n \langle z_l, z_i\rangle\langle y_i, \psi\rangle = \alpha\left\langle z_l, \hat{\Pi}_\alpha^*\psi\right\rangle + \frac{1}{n}\sum_{i=1}^n \langle z_l, z_i\rangle\left\langle z_i, \hat{\Pi}_\alpha^*\psi\right\rangle,$$
$$l = 1, 2, \ldots, n, \quad (11)$$

with n unknowns $\left\langle z_i, \hat{\Pi}_\alpha^*\psi\right\rangle$, $i = 1, 2, \ldots, n$. Let M be the $n \times n$ matrix with (l, i) element $\langle z_l, z_i\rangle/n$, v the $n-$vector of $\left\langle z_i, \hat{\Pi}_\alpha^*\psi\right\rangle$ and w the $n-$vector of $\langle y_i, \psi\rangle$. (11) is equivalent to

$$Mw = (\alpha I + M)\,v.$$

And $v = (\alpha I + M)^{-1}Mw = M(\alpha I + M)^{-1}w$. For a given ψ, we can compute

$$\begin{aligned}\hat{\Pi}_\alpha^*\psi &= \frac{1}{\alpha n}\sum_{i=1}^n z_i\left(\langle y_i, \psi\rangle - \left\langle z_i, \hat{\Pi}_\alpha^*\psi\right\rangle\right)\\ &= \frac{1}{\alpha n}\underline{z}'\left(I - M(\alpha I + M)^{-1}\right)w\\ &= \frac{1}{n}\underline{z}'(\alpha I + M)^{-1}w \quad (12)\end{aligned}$$

where $\underline{z}$ is the $n-$vector of z_i.

Now, we explain how to estimate $\Pi\varphi$ for any $\varphi \in \mathcal{F}$. Taking the inner product with φ in the left and right hand sides of (12), we obtain

$$\left\langle \varphi, \hat{\Pi}_\alpha^*\psi\right\rangle = \frac{1}{\alpha n}\sum_{i=1}^n \langle\varphi, z_i\rangle\left(\langle y_i, \psi\rangle - \left\langle z_i, \hat{\Pi}_\alpha^*\psi\right\rangle\right) \Leftrightarrow$$

$$\left\langle \hat{\Pi}_\alpha\varphi, \psi\right\rangle = \frac{1}{\alpha n}\sum_{i=1}^n \langle\varphi, z_i\rangle\left\langle y_i - \hat{\Pi}_\alpha z_i, \psi\right\rangle$$

for all $\psi \in \mathcal{E}$. This implies

$$\hat{\Pi}_\alpha\varphi = \frac{1}{\alpha n}\sum_{i=1}^n \langle\varphi, z_i\rangle\left(y_i - \hat{\Pi}_\alpha z_i\right). \quad (13)$$

Hence, to compute $\hat{\Pi}_\alpha\varphi$, we need to know $\hat{\Pi}_\alpha z_i$. From (7), we have

$$\alpha\hat{\Pi}_\alpha + \hat{\Pi}_\alpha\hat{V}_Z = \hat{C}_{YZ}.$$

Applying the l.h.s and r.h.s to $z_i, i = 1, 2, \ldots, n$, we obtain

$$\alpha\hat{\Pi}_\alpha z_i + \hat{\Pi}_\alpha\hat{V}_Z z_i = \hat{C}_{YZ}z_i \Leftrightarrow$$

$$\alpha\left(\hat{\Pi}_\alpha z_i\right)(t) + \frac{1}{n}\sum_{j=1}^n\left(\hat{\Pi}_\alpha z_j\right)(t)\left\langle z_j, z_i\right\rangle$$
$$= \frac{1}{n}\sum_{j=1}^n y_j(t)\left\langle z_j, z_i\right\rangle, i = 1, 2, \ldots, n. \quad (14)$$

For each t, we can solve the n equations with n unknowns $\left(\hat{\Pi}_\alpha z_j\right)(t)$ given by (14) and deduct $\hat{\Pi}_\alpha\varphi$ from (13). Let us denote $\underline{y}(t)$, $\underline{z}(t)$, $\underline{\varsigma}(t)$ the $n \times 1$ vectors with ith element $y_i(t)$, $z_i(t)$, and $\hat{\Pi}_\alpha z_i(t)$ respectively. The solution $\underline{\varsigma}(t)$ of Eq. (14) is equal to

$$\underline{\varsigma}(t) = (\alpha I + M)^{-1}M\underline{y}(t).$$

It follows from (13) that $\hat{\Pi}_\alpha$ is an integral operator with degenerate kernel equal to

$$\hat{\pi}_\alpha(s, t) = \frac{1}{n}\underline{y}(s)'(\alpha I + M)^{-1}\underline{z}(t). \quad (15)$$

Note that this calculation is exact, we did not rely on any discretization. We see that to compute the kernel of $\hat{\Pi}_\alpha$, we only need to compute the elements $\langle z_l, z_i\rangle/n$ of the matrix M.

The prediction of Y_i is given by

$$\hat{y}_i = \hat{\Pi}_\alpha z_i.$$

2.6. Alternative estimators

In this section, we briefly expose the existing alternative estimators of the operator Π in a functional linear regression with functional response. In these papers, the operator kernel is defined slightly differently from ours. To fix the notation, we consider the estimation of β in the following model

$$Y(t) = \int \beta(s, t)\, Z(s)\, ds + U(s). \quad (16)$$

In our notation, $\Pi(t, s) = \beta(s, t)$.

Ramsay and Silverman (2005) propose two ways for estimating β. Both methods rely on an approximation of β using basis functions $\{\eta_j : j \geq 1\}$ and $\{\theta_l : l \geq 1\}$ which could be for instance spline or Fourier bases. In the first version, β is estimated by

$$\beta_{JL}(s, t) = \sum_{j=1}^J\sum_{l=1}^L b_{jl}\eta_j(s)\,\theta_l(t) = \eta(s)'B\theta(t)$$

for some small numbers J and L. Truncating the η basis permits to avoid over-fitting whereas truncating the θ basis insures that the prediction is smooth. The estimator of b_{jl} is obtained by minimizing

$$\begin{aligned}&\|y - \int z(s)\,\beta_{JL}(s, .)\,ds\|^2\\ &\quad = \int\sum_{i=1}^n\left[y_i(t) - \int z_i(s)\,\beta_{JL}(s, .)\,ds\right]^2 dt.\end{aligned}$$

In the second approach proposed by Ramsay and Silverman (2005), J and L are chosen to be large and smoothness is ensured by adding some roughness penalties to the objective function:

$$\begin{aligned}\min_\beta &\|y - \int z(s)\,\beta_{JL}(s, .)\,ds\|^2 + \lambda_s\int\int\left[L_s\beta_{JL}(s, t)\right]^2 dsdt\\ &+ \lambda_t\int\int\left[L_t\beta_{JL}(s, t)\right]^2 dsdt\end{aligned}$$

where λ_s and λ_t are the penalization parameters and L_k denotes a differential operator to be applied to $\beta(s, t)$ with respect to $k \in \{s, t\}$ only. Ramsay and Silverman (2005) use cross-validation methods for choosing the penalization parameters λ_s and λ_t.

Instead of using arbitrary bases $\{\eta_j\}$, $\{\theta_l\}$, a popular approach consists in using the functional principal components which correspond to the eigenfunctions of the covariance operators V_Z and V_Y of Z and Y respectively. Let ψ_j and ϕ_l be the eigenfunctions of V_Z and V_Y respectively. Let λ_j be the eigenvalues of V_Z. Estimators $\hat{\lambda}_j$, $\hat{\psi}_j$ and $\hat{\phi}_l$ of λ_j, ψ_j and ϕ_l respectively are given by the eigenvalues and eigenfunctions of the sample counterparts of V_Z and V_Y. Then, an estimator of $\beta(s,t)$ is given by

$$\hat{\beta}(s,t)=\sum_{j=1}^{J}\sum_{l=1}^{L}\frac{\hat{\rho}_{jl}}{\hat{\lambda}_j}\hat{\psi}_j(s)\hat{\phi}_l(t) \tag{17}$$

where

$$\hat{\rho}_{jl}=\int\int \hat{C}_{ZY}(s,t)\hat{\psi}_j(s)\hat{\phi}_l(t)\,dsdt$$

and $\hat{C}_{ZY}(s,t)$ is an estimator of $E(Z(s)Y(t))$. The simplest estimator is given by the empirical covariance

$$\hat{C}_{ZY}(s,t)=\frac{1}{n}\sum_{i=1}^{n}Z_i(s)Y_i(t).$$

The resulting kernel is

$$\hat{\beta}(s,t)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J}\frac{\langle Z_i,\hat{\psi}_j\rangle}{\hat{\lambda}_j}\hat{\psi}_j(s)\sum_{l=1}^{L}\langle Y_i,\hat{\phi}_l\rangle\hat{\phi}_l(t). \tag{18}$$

Yao et al. (2005) rely on scatterplot smoothing to compute $\hat{\rho}_{jl}$ in (17) and use leave-one-out cross-validation to select the number of included eigenfunctions J and L. Park and Qian (2012) and Crambes and Mas (2013) use a slightly different estimator:

$$\hat{\beta}(s,t)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J}\frac{\langle Z_i,\hat{\psi}_j\rangle}{\hat{\lambda}_j}\hat{\psi}_j(s)Y_i(t).$$

For comparison, using Eq. (6), our estimator of β can be written as

$$\hat{\beta}(s,t)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\langle Z_i,\hat{\psi}_j\rangle}{\hat{\lambda}_j+\alpha}\hat{\psi}_j(s)Y_i(t).$$

So the only differences are that the sum over j is not truncated and $\hat{\lambda}_j$ in the denominator is replaced by $\hat{\lambda}_j+\alpha$. Crambes and Mas (2013) derive the rate of convergence of their estimator under the assumption that λ_j is a convex function of j.

Bosq (2000) considers a model similar to (16) where the regressor is the lagged value of the dependent variable. Let Z_i be a stationary process in a Hilbert space. The autoregressive model of order one writes

$$Z_{i+1}(t)=\int \beta(s,t)Z_i(s)\,ds+U_i(t). \tag{19}$$

Bosq (2000) proposes an estimator of the form (18) and establishes its consistency under the assumption that all eigenvalues are distinct. Its rate of convergence depends on the difference between successive eigenvalues.

In summary, our estimator presents some advantages compared to existing estimators: (a) It involves only one smoothing parameter α. (b) Computing our estimator using formula (15) is simple and does not require the estimation of eigenfunctions. (c) In Section 3, we derive the properties of our estimator under a source condition which implicitly imposes some restrictions on the decay rate of eigenvalues but does not rule out multiple eigenvalues. (d) The extension in Section 7 allows for endogenous regressors.

3. Rate of convergence of the MSE

In this section, we study the rate of convergence of the mean square error (MSE) of $\hat{\Pi}^*_\alpha$. Several assumptions are needed.

Assumption 1. The observations (U_i, Y_i, Z_i), $i=1,2,\ldots,n$ are independent, identically distributed as (U,Y,Z) where Y and U are random processes of a separable Hilbert space $\mathcal{E}$ and Z is a random process of a separable Hilbert space $\mathcal{F}$. Moreover, $E(U_i)=0$, $cov(U_i,Z_i)=0$, $cov(U_i,U_j|Z_1,Z_2,\ldots,Z_n)=0$ for all $i\neq j$ and $=V_U$ for $i=j$ where V_U is a nonrandom trace-class operator.

Assumption 2. Π belongs to $\mathcal{H}(\mathcal{F},\mathcal{E})$ the space of Hilbert–Schmidt operators.

Assumption 3. V_Z is a trace-class operator and $\left\|\hat{V}_Z-V_Z\right\|^2_{HS}=O_p(1/n)$.

Assumption 4. There is a Hilbert–Schmidt operator R from $\mathcal{E}$ to $\mathcal{F}$ and a constant $\beta>0$ such that $\Pi^*=V_Z^{\beta/2}R$.

Assumption 1 imposes that U_i is homoskedastic and Z_i is exogenous. The stronger condition of strict exogeneity, $E(U_i|Z_i)=0$, is not needed here. It will be used only briefly in Section 4.

An operator K is trace-class if $\sum_j\langle K\phi_j,\phi_j\rangle<\infty$ for any basis (ϕ_j). If K is self-adjoint positive definite, it is equivalent to say that the sum of the eigenvalues of K is finite. Given V_U is a covariance operator, V_U is trace-class if and only if $E\left(\|U_i\|^2\right)<\infty$.

The notation $\|\|_{HS}$ refers to the Hilbert–Schmidt norm of operators, $\|K\|^2_{HS}\equiv\sum_j\langle K\phi_j,K\phi_j\rangle<\infty$ for any basis (ϕ_j). An operator K is Hilbert–Schmidt (noted HS) if $\|K\|^2_{HS}<\infty$. If K is self-adjoint positive definite, the condition $\|K\|^2_{HS}<\infty$ is equivalent to the condition that the eigenvalues of K are square summable. A sufficient condition for $\left\|\hat{V}_Z-V_Z\right\|^2_{HS}=O_p(1/n)$ is that Z_i is a i.i.d. random process and $E\left(\|Z_i\|^4\right)<\infty$, see Proposition 5 of Dauxois et al. (1982).

Assumption 4 is a source condition needed to characterize the rate of convergence of the MSE. Moreover, it guarantees that Π^* belongs to the orthogonal of the null space of V_Z denoted $\mathcal{N}(V_Z)$. Given this condition, there is no need to impose $\mathcal{N}(V_Z)=\{0\}$ to get the identification.

The source condition can be rewritten in terms of the spectral decomposition of V_Z. Let (λ_j,φ_j) be the eigenvalues (ordered in decreasing order) and eigenfunctions of V_z. The source condition of Assumption 4 is equivalent to assume that

$$\sum_{j=1}^{\infty}\frac{\langle\Pi^*\varphi,\varphi_j\rangle^2}{\lambda_j^{\beta}}<\infty \text{ for all } \varphi\in\mathcal{E}.$$

As V_Z is a compact operator, λ_j go to zero as j goes to infinity. So this condition imposes that the eigenvalues of λ_j decline to zero not too fast relatively to the Fourier coefficients $\langle\Pi^*\varphi,\varphi_j\rangle$.

Because V_Z is a compact integral operator, its inverse is a differential operator. Hence, the larger β in Assumption 4, the smoother the operator Π^* is. To illustrate the connection between Assumption 4 and smoothness conditions, let us consider two simple examples.

1. Example 1. Let $\mathcal{E}=\mathcal{F}=L^2[0,1]$ and $(V_Z\varphi)(s)=\int_0^1\min(s,t)\varphi(t)\,dt$. If $\beta=2$, then the source condition of Assumption 4 is satisfied if and only if for all $\varphi\in L^2[0,1]$, $(\Pi^*\varphi)(0)=0$, $\left.\frac{\partial(\Pi^*\varphi)(t)}{\partial t}\right|_{t=1}=0$ and $(\Pi^*\varphi)(t)$ is twice differentiable with respect to t.
2. Example 2. We consider now the case where V_Z has a Gaussian kernel, that is $(V_Z\varphi)(s)=\int_{-\infty}^{\infty}\exp\left\{-\frac{(s-t)^2}{\sigma^2}\right\}\varphi(t)\,dt$ and assume $\beta=2$. In this case, $(\Pi^*\varphi)(t)$ is infinitely differentiable with respect to t for all $\varphi\in\mathcal{E}$.

The conditional MSE is defined by

$$MSE=E\left(\left\|\hat{\Pi}_\alpha-\Pi\right\|^2_{HS}|Z_1,\ldots,Z_n\right).$$

Our definition of the MSE makes sense only if $\hat{\Pi}_\alpha$ is a Hilbert–Schmidt operator. We establish this property in the following proposition.

Proposition 1. *Under Assumptions 1 and 3, $\hat{\Pi}_\alpha$ belongs to $\mathcal{H}(\mathcal{F},\mathcal{E})$ for all $\alpha > 0$.*

Now we turn our attention to the rate of convergence of the MSE.

Proposition 2. *Assume Assumptions 1–4 hold.*
If $\beta > 1$, then MSE=$O_p\left(\frac{1}{n\alpha} + \alpha^{\beta\wedge 2}\right)$.
If $\beta < 1$, then MSE=$O_p\left(\frac{\alpha^\beta}{n\alpha^2} + \alpha^{\beta}\right)$.

Remark 3.1. Proposition 2 shows that the MSE exhibits the usual trade-off between the variance decreasing in α and the bias increasing in α. Taking the optimal α which equates the rates of the squared bias and variance, we obtain the following rates:

- for $\beta > 1$, $\alpha \sim n^{-1/(1+\beta\wedge 2)}$, $MSE \sim n^{-\beta\wedge 2/(1+\beta\wedge 2)}$,
- for $\beta < 1$, $n\alpha^2 \to \infty$, $MSE \sim \alpha^\beta$.

Remark 3.2. We see that the rate of convergence of the bias does not improve when $\beta > 2$. This is due to the well-known saturation effect of Tikhonov regularization (see for instance Engl et al., 2000). This rate could be improved by using iterated Tikhonov regularization.

Remark 3.3. Proposition 2 does not impose any restriction on the multiplicity of the eigenvalues. This is a major difference with the estimators based on the truncation. The consistency proof for these estimators requires that the estimated eigenvalues are consistent estimators of the true eigenvalue. When the order of multiplicity of an eigenvalue is greater than one, there is no such consistent estimator. On the other hand, our proof does not rely on the spectral decomposition of the operator. In that sense, Tikhonov regularization is more robust to situations where eigenvalues are multiple than methods based on the truncation.

Now we consider the case where the operator T defined in Section 2.4 is not injective and Π is not identified. We can decompose the operator Π as $\Pi_N + \Pi_{N^\perp}$ where $\Pi_N \in \mathcal{N}(T)$ and $\Pi_{N^\perp} \in \mathcal{N}(T)^\perp$. $\Pi_{N^\perp}$ corresponds to the minimal least squares solution of the inverse problem. We will show below that under appropriate assumptions, $\hat{\Pi}_\alpha$ converges to $\Pi_{N^\perp}$. To derive this result, we replace Assumption 4 by the following assumption where Π^* is replaced by $\Pi^*_{N^\perp}$.

Assumption 4′. There is a Hilbert–Schmidt operator R from $\mathcal{E}$ to $\mathcal{F}$ and a constant $\beta > 0$ such that $\Pi^*_{N^\perp} = V_Z^{\beta/2}R$.

Proposition 3. *Assume Assumptions 1–3 and 4′ hold.*
$$E\left(\left\|\hat{\Pi}_\alpha - \Pi_{N^\perp}\right\|^2_{HS}|Z_1,\ldots,Z_n\right) = O_p\left(\frac{1}{n\alpha^2} + \alpha^{\beta\wedge 2}\right).$$

Remark 3.4. Under our assumptions, Π cannot be estimated consistently because there is no information about Π_N in the data. However, $\Pi_{N^\perp}$ is consistently estimated.

Remark 3.5. Observe that the lack of identification results in a larger bias and hence a slower rate of convergence of the MSE. Here, the optimal α satisfies $\alpha \sim n^{-1/(2+\beta\wedge 2)}$ and the rate for the MSE is $n^{-\beta\wedge 2/(2+\beta\wedge 2)}$. Florens et al. (2011) studied the effect of nonidentification in the context of nonparametric instrumental regression and found a reduced rate of convergence as we found here.

Remark 3.6. The saturation effect of Tikhonov persists in the non identified case.

4. Asymptotic normality for fixed α and tests

In this section, we will derive the asymptotic distribution of $\hat{\Pi}^*_\alpha$ when α is fixed and then develop a test for the null hypothesis $H_0 : \Pi = \Pi_0$ where Π_0 is some known operator.

When α is fixed, $\hat{\Pi}^*_\alpha$ is not consistent and keeps an asymptotic bias. It is useful to define Π^*_α the regularized version of Π^*:

$$\Pi^*_\alpha = (\alpha I + V_Z)^{-1}V_Z\Pi^*.$$

We have

$$\begin{aligned}
\hat{\Pi}^*_\alpha - \Pi^*_\alpha &= \left(\alpha I + \hat{V}_Z\right)^{-1}\hat{C}_{ZU} \\
&\quad + \left(\alpha I + \hat{V}_Z\right)^{-1}\hat{V}_Z\Pi^* - (\alpha I + V_Z)^{-1}V_Z\Pi^* \\
&= (\alpha I + V_Z)^{-1}\hat{C}_{ZU} \\
&\quad + \left[\left(\alpha I + \hat{V}_Z\right)^{-1} - (\alpha I + V_Z)^{-1}\right]\hat{C}_{ZU} \\
&\quad + \alpha\left(\alpha I + \hat{V}_Z\right)^{-1}\left(V_Z - \hat{V}_Z\right)(\alpha I + V_Z)^{-1}\Pi^* \\
&= (\alpha I + V_Z)^{-1}\hat{C}_{ZU} \qquad (20)\\
&\quad + \alpha(\alpha I + V_Z)^{-1}\left(\hat{V}_Z - V_Z\right)(\alpha I + V_Z)^{-1}\Pi^* \\
&\quad + O_p\left(\frac{1}{n}\right). \qquad (21)
\end{aligned}$$

As n goes to infinity, $\hat{\Pi}^*_\alpha - \Pi^*_\alpha$ converges to zero and is $\sqrt{n}-$asymptotically normal. The first two terms of the r.h.s are $O_p\left(1/\sqrt{n}\right)$ and will affect the asymptotic distribution. This distribution is not simple. We are going to characterize it below. The notation $\otimes$ denotes the functional tensor product defined as $(x \otimes y)(f) = \langle x, f\rangle y$.

From Eqs. (20) and (21), neglecting the $O_p(1/n)$ term, we have

$$\begin{aligned}
\hat{\Pi}^\star_\alpha - \Pi^\star_\alpha &= (\alpha I + V_Z)^{-1}\hat{C}_{ZU} + \alpha(\alpha I + V_Z)^{-1}(\hat{V}_Z - V_Z) \\
&\quad (\alpha I + V_Z)^{-1}\Pi^\star \\
&= (\alpha I + V_Z)^{-1}\frac{1}{n}\int_i\sum(u_i \otimes z_i) + \alpha(\alpha I + V_Z)^{-1} \\
&\quad \times \frac{1}{n}\sum_i(z_i \otimes z_i - V_Z)(\alpha I + V_Z)^{-1}\Pi^\star \\
&= \frac{1}{n}\sum_i\Big(u_i \otimes (\alpha I + V_Z)^{-1}z_i + \alpha\Pi(\alpha I + V_Z)^{-1}z_i \\
&\quad \otimes(\alpha I + V_Z)^{-1}z_i\Big) \\
&\quad - \alpha(\alpha I + V_Z)^{-1}V_Z(\alpha I + V_Z)^{-1}\Pi^\star \\
&= \frac{1}{n}\sum_i\Big(u_i \otimes (\alpha I + V_Z)^{-1}z_i + \alpha\Pi(\alpha I + V_Z)^{-1}z_i \\
&\quad \otimes(\alpha I + V_Z)^{-1}z_i\Big) \\
&\quad -\alpha \mathrm{E}[\Pi(\alpha I + V_Z)^{-1}Z \otimes (\alpha I + V_Z)^{-1}Z] \\
&= \frac{1}{n}\sum_i\left(u_i \otimes \tilde{z}_i + \alpha\Pi\tilde{z}_i \otimes \tilde{z}_i - \alpha\mathrm{E}[\Pi\tilde{Z} \otimes \tilde{Z}]\right) \\
&= \frac{1}{n}\sum_i\left((u_i + \alpha\Pi\tilde{z}_i) \otimes \tilde{z}_i - \alpha C_{\tilde{Z}\Pi\tilde{Z}}\right),
\end{aligned}$$

where $\tilde{Z} \equiv (\alpha I + V_Z)^{-1}Z$ and $\tilde{z}_i \equiv (\alpha I + V_Z)^{-1}z_i$. The first equality makes use of the definition of the empirical covariance operators using tensor products. The second line uses the elementary properties $K(Y \otimes X) = Y \otimes KX$ and $(Y \otimes X)K = K^\star Y \otimes X$ for $X \in \mathcal{F}$, $X \in \mathcal{E}$ and $K \in \mathcal{H}$. The third line uses the definition of V_Z. The interchange of the expectation operator and $(\alpha I + V_Z)^{-1}$ is allowed since the latter is a bounded linear operator. By Banach inverse theorem, the inverse of a bounded linear operator is itself linear and bounded (see, for instance Rudin, 1991). The last equality holds since the functional tensor product distributes over addition.

The covariance operator of $\hat{\Pi}^\star_\alpha - \Pi^\star_\alpha$ is an operator which maps the space of Hilbert–Schmidt operators from $\mathcal{E}$ to $\mathcal{F}$, denoted $\mathcal{G}$, into itself. Such an operator may be difficult to write explicitly. Fortunately, the properties of tensor products of infinite-dimensional Hilbert–Schmidt operators defined on separable Hilbert spaces are well-known,[4] and may be used like in Dauxois et al. (1982) to write explicitly the covariance operator of an infinite-dimensional Hilbert–Schmidt random operator. The tensor product $\Pi_1\tilde{\otimes}\Pi_2$ for $(\Pi_1,\Pi_2) \in \mathcal{G}^2$ is a mapping from $\mathcal{G}$ into itself, hence $\Pi_1\tilde{\otimes}\Pi_2$ is an element of the Hilbert space of Hilbert–Schmidt operators from $\mathcal{G}$ to $\mathcal{G}$ equipped with the Hilbert–Schmidt inner product. For $T = \varphi\otimes\psi \in \mathcal{G}$, $\Pi_1 = X\otimes Z \in \mathcal{G}$ and $\Pi_2 = Y\otimes W \in \mathcal{G}$, this tensor product is equivalently defined as :

(i) $(\Pi_1\tilde{\otimes}\Pi_2)T = \langle T, \Pi_1\rangle_{\mathcal{G}}\Pi_2 \in \mathcal{G}$

(ii) $\Big((X\otimes Z)\tilde{\otimes}(Y\otimes W)\Big)(\varphi\otimes\psi) = \Big((X\otimes Y)\varphi\Big) \otimes\Big((Z\otimes W)\psi\Big)$, $\forall \varphi, X, Y \in \mathcal{E}, \psi, Z, W \in \mathcal{F}$.

Based upon definition (i), the covariance operator of Π_1 and Π_2 naturally writes as

$$E\Big[\langle\cdot, \Pi_1 - \mathrm{E}[\Pi_1]\rangle_{\mathcal{G}}\big(\Pi_2 - \mathrm{E}[\Pi_2]\big)\Big] = \mathrm{E}\Big[\big(\Pi_1 - \mathrm{E}[\Pi_1]\big)\tilde{\otimes}\big(\Pi_2 - \mathrm{E}[\Pi_2]\big)\Big]. \tag{22}$$

Furthermore, to show asymptotic normality, we shall use the classical central limit theorem for i.i.d. processes in separable Hilbert spaces. The following is stated as Theorem 2.7 in Bosq (2000) and is reproduced here for clarity.

Theorem 4 (*Bosq, 2000*)**.** *Let $(Z_i, i \geq 1)$ be a sequence of i.i.d. $\mathcal{F}$-valued random variables, where $\mathcal{F}$ is a separable Hilbert space, such that $E\|Z_i\|^2 < \infty$, $E(Z_i) = \bar{Z}$ and $V_Z = V$, then one has*

$$\frac{1}{\sqrt{n}}\sum_i (Z_i - \bar{Z}) \xrightarrow{d} \mathcal{N}(0, V),$$

We are now geared to derive the asymptotic covariance operator of interest in its general form, under some standard assumptions.

Proposition 5. *Assume (U_i, Z_i) i.i.d., $\Omega_\alpha < \infty$, $E\|Z_i\|^4 < \infty$, $E\|U_i\|^2\|Z_i\|^2 < \infty$, then*

$$\sqrt{n}(\hat{\Pi}^\star_\alpha - \Pi^\star_\alpha) \xrightarrow{d} \mathcal{N}(0, \Omega_\alpha), \tag{23}$$

where the asymptotic covariance operator[5] *Ω_α for fixed α is given by*

$$\Omega_\alpha = E\Big[\Big((U + \alpha\Pi\tilde{Z})\otimes\tilde{Z}\Big)\tilde{\otimes}\Big((U + \alpha\Pi\tilde{Z})\otimes\tilde{Z}\Big)\Big] - \alpha^2 C_{\tilde{Z}\Pi\tilde{Z}}\tilde{\otimes}C_{\tilde{Z}\Pi\tilde{Z}}, \tag{24}$$

which simplifies to

$$\Omega_0 = E\Big[\big(U\otimes V_Z^{-1}Z\big)\tilde{\otimes}\big(U\otimes V_Z^{-1}Z\big)\Big], \tag{25}$$

when $\alpha \to 0$.

Remarks. The results of Proposition 5 are derived without imposing the strict exogeneity assumption, $E[U_i|Z_i] = 0$. If this assumption is satisfied, the asymptotic covariance operator in (24) simplifies into

$$\Omega_\alpha = \mathrm{E}\Big[\Big(U\otimes\tilde{Z}\Big)\tilde{\otimes}\Big(U\otimes\tilde{Z}\Big)\Big] + \alpha^2\mathrm{E}\Big[\Big(\Pi\tilde{Z}\otimes\tilde{Z}\Big)\tilde{\otimes}\Big(\Pi\tilde{Z}\otimes\tilde{Z}\Big)\Big] - \alpha^2 C_{\tilde{Z}\Pi\tilde{Z}}\tilde{\otimes}C_{\tilde{Z}\Pi\tilde{Z}}.$$

In econometrics, we are often interested in testing the significance of estimates and produce confidence bands. However, there is no obvious meaningful way to perform standard significance tests using the derived asymptotic covariance. Indeed, for fixed α the estimated residuals will be biased and one must specify $\Pi^\star$. On the other hand, if we assume $\alpha \to 0$, an estimator of (25) may be uninformative since V_Z^{-1} does not necessarily exist. A more practical approach would be to keep α fixed to obtain an estimate of $(\alpha I + V_Z)^{-1}$ and use it to derive an estimator of (25). Other statistical tests may involve applying a test operator to Ω_α.

We want to test the null hypothesis: $H_0 : \Pi = \Pi_0$ where Π_0 is known. A simple way to test this hypothesis is to look at $\hat{C}_{ZY} - \hat{V}_Z\Pi_0^*$. Under H_0, this operator equals $\hat{C}_{ZU}$ and should be close to zero. Moreover, under H_0,

$$\sqrt{n}\left(\hat{C}_{ZY} - \hat{V}_Z\Pi_0^*\right) \xrightarrow{d} \mathcal{N}(0, K_{ZU})$$

where

$$K_{ZU} = E\left[(u\otimes Z)\,\tilde{\otimes}\,(u\otimes Z)\right]$$

(see Dauxois et al., 1982).

Let $\{f_j, g_j : j = 1, 2, \ldots, q\}$ be a set of test functions, then

$$\begin{bmatrix} \sqrt{n}\left\langle\left(\hat{C}_{ZY} - \hat{V}_Z\Pi_0^*\right)f_1, g_1\right\rangle \\ \vdots \\ \sqrt{n}\left\langle\left(\hat{C}_{ZY} - \hat{V}_Z\Pi_0^*\right)f_q, g_q\right\rangle \end{bmatrix}$$

converges to a multivariate normal distribution with mean 0_q and covariance matrix the $q\times q$ matrix Σ with (j, l) element:

$$\Sigma_{jl} = E\left[\left\langle\sqrt{n}\hat{C}_{ZU}f_j, g_j\right\rangle\left\langle\sqrt{n}\hat{C}_{ZU}f_l, g_l\right\rangle\right] = \langle g_j, V_Z g_l\rangle\langle f_j, V_U f_l\rangle$$

where the second equality follows from (U_i, Z_i) iid and the homoskedasticity assumption given in Assumption 1. This covariance matrix can be easily estimated by replacing V_Z and V_U by their sample counterpart. The appropriately rescaled quadratic form converges to a chi-square distribution with q degrees of freedom which can be used to test H_0. The test functions could be cumulative normals as in Conley et al. (1997) or could be normal densities with same small variance but centered at different means. Note that here we assumed Π_0 completely specified under H_0. If instead, Π_0 depends on some unknown parameters θ which need to be estimated, the asymptotic variance Σ needs to be adjusted to take into account the estimation error.

5. Data-driven selection of α

The estimator involves a tuning parameter, α, which needs to be selected. It can be chosen as the solution to

$$\min_\alpha \frac{1}{\alpha}\left\|\hat{V}_Z\hat{\Pi}^*_\alpha - \hat{C}_{ZY}\right\|^2_{HS},$$

see Engl et al. (2000, p.102) or

$$\min_\alpha tr\left(\hat{\Pi}_\alpha\hat{\Pi}^*_\alpha\right)\left\|\hat{V}_Z\hat{\Pi}^*_\alpha - \hat{C}_{ZY}\right\|^2_{HS},$$

[4] See, for instance, Vilenkin (1968, p. 59–65).

[5] The kernel of Ω_α can be written as

$$\omega_\alpha(s,t,r,\tau) = \mathrm{E}\Big[\big(U(s) + \alpha\Pi\tilde{Z}(s)\big)Z(t)\big(U(r) + \alpha\Pi\tilde{Z}(r)\big)Z(\tau)\Big] - \alpha^2\mathrm{E}\Big[\Pi\tilde{Z}(s)\tilde{Z}(t)\Big]\mathrm{E}\Big[\Pi\tilde{Z}(r)\tilde{Z}(\tau)\Big].$$

see Engl et al. (2000, Proposition 4.37) where $\hat{\Pi}_\alpha^*$ in $\left\|\hat{V}_Z\hat{\Pi}_\alpha^* - \hat{C}_{ZY}\right\|_{HS}^2$ is obtained by iterated Tikhonov.

Another possibility is to use leave-one-out cross-validation

$$\min_\alpha \frac{1}{n}\sum_j \left\|y_i - \hat{\Pi}_\alpha^{(-i)} z_i\right\|^2$$

where $\hat{\Pi}_\alpha^{(-i)}$ has been computed using all observations except for the ith one. Centorrino (2016) studies the properties of the leave-one-out cross-validation for nonparametric IV regression and shows that this criterion is rate optimal in mean squared error. This method is also used in a binary response model by Centorrino and Florens (2015). Various data-driven selection techniques are compared via simulations in Centorrino et al. (2017).

An alternative approach would be to use a penalized minimum contrast criterion as in Goldenshluger and Lepski (2011). This could lead to a minimax-optimal estimator (Comte and Johannes, 2012).

6. Discrete observations

6.1. Effect of discretization

In this section, to simplify the exposition, we will refer to the arguments of (y_i, z_i), t, as time even though it could refer to a location or other characteristic. Suppose that the data (y_i, z_i) are not observed in continuous time but at discrete (not necessarily equally spaced) times within a fixed time span $\mathcal{T} \subset \mathbb{R}$ (for example $\mathcal{T} = [0, 1]$). It is necessary to construct pairs of curves (y_i^m, z_i^m), $i = 1, 2, \ldots, n$ such that $y_i^m \in \mathcal{E}$ and $z_i^m \in \mathcal{F}$, that can be evaluated at any desired t in the domain $\mathcal{T}$. These approximate curves can be obtained with some interpolation or smoothing procedure using step functions, kernel smoothing or splines for instance (see Ramsay and Silverman, 2005).[6]

Let the subscript m correspond to the smallest number of discrete measurements across $i = 1, 2, \ldots, n$ which defines the sampled domain $\mathcal{T}_m$ of the smoothed data $(y_i^m, z_i^m)_{i=1}^n$. We consider infill asymptotics where m grows with the sample size n (m stands for $m(n)$). That is, we assume that for all $t \in \mathcal{T}$, $Pr(t \in \mathcal{T}_m) \to 1$ as $n \to \infty$ hence $\mathcal{T}_m \to \mathcal{T}$ as $n \to \infty$. Intuitively, it means that new discrete measurements become available at values of t between the existing ones, as the sample size increases, in such a way that the sampled domain becomes dense and coincides with the true domain at the limit.

Using the smoothed observations, we compute the corresponding estimators of V_Z and C_{ZY} denoted $\hat{V}_Z^m$, $\hat{C}_{ZY}^m$ and the estimator of Π^* denoted $\hat{\Pi}_\alpha^{m*}$:

$$\hat{\Pi}_\alpha^{m*} = \left(\alpha I + \hat{V}_Z^m\right)^{-1}\hat{C}_{ZY}^m.$$

To assess the rate of convergence of $\hat{\Pi}_\alpha^{m*}$, we add the following conditions which guarantee that the discretization error is negligible with respect to the estimation error.

Assumption 5. $\left\|z_i^m - z_i\right\| = O_p\left(f\left(m\right)\right)$ and $\left\|y_i^m - y_i\right\| = O_p\left(f\left(m\right)\right)$.

Assumption 6. We have

$$\frac{f\left(m\right)}{\alpha n} = o\left(\alpha^{\beta \wedge 2}\right)$$

as n and $m = m\left(n\right)$ go to infinity.

Proposition 6. *Under Assumptions 1–6, the MSE of $\hat{\Pi}_\alpha^{m*} - \Pi^*$ has the same rate of convergence as that of the MSE of $\hat{\Pi}_\alpha^* - \Pi^*$ in Proposition 2 as n and $m = m(n)$ go to infinity.*

This proposition shows that using discrete observations does not affect the rate of convergence of the estimators as long as the discretization error is negligible with respect to the estimation error.

6.2. Relation with the Ridge estimator

In order to clarify some of the merits of using our functional estimator, we show how it relates to the standard ridge estimator. Suppose we have at hands n pairs of discretized $L^2[0, 1]$ functions (y_i^m, z_i^m), for $i \in \{1, 2, \ldots, n\}$, each evaluated at m equispaced points $t = \frac{1}{m}, \frac{2}{m}, \ldots, 1$. Let $\underline{z}$ and $\underline{y}$ be the $n \times m$ matrices with (i, j) element $\left(z_i\left(\frac{j}{m}\right)\right)$ and $\left(y_i\left(\frac{j}{m}\right)\right)$, $i = 1, \ldots, n, j = 1, \ldots, m$, respectively. From (15), the kernel of $\hat{\Pi}_\alpha$ can be written as

$$\hat{\pi}_\alpha = \underline{z}'(\alpha n I_n + nM)^{-1}\underline{y}, \tag{26}$$

If the $n \times n$ matrix M with (l, i) element $\frac{\langle z_l, z_i\rangle}{n} = \frac{1}{n}\int_0^1 z_l(t)z_i(t)dt$ is approached by the estimator $\hat{M} = \frac{\underline{z}\underline{z}'}{n}$ with (l, i) element $\frac{1}{nm}\sum_{j=1}^m z_l(\frac{j}{m})z_i(\frac{j}{m})$ $\forall l, i \in \{1, 2, \ldots, n\}$, we obtain the approximation defined by

$$\tilde{\hat{\pi}}_\alpha = \underline{z}'\left(\alpha n I_n + \underline{z}\underline{z}'\right)^{-1}\underline{y}. \tag{27}$$

Let ϕ_j denote the eigenvectors of $\underline{z}'\underline{z}$ associated with eigenvalues λ_j^2 and ψ_j denote the eigenvectors of $\underline{z}\underline{z}'$ associated with eigenvalues λ_j^2. For any vector v of dimension n, we have $v = \sum_j\langle v, \psi_j\rangle\psi_j$ and hence $\underline{z}'v = \sum_j\lambda_j\langle v, \psi_j\rangle\phi_j$ since $\underline{z}'\psi_j = \lambda_j\phi_j$. Therefore, we can write

$$(\alpha n I_m + \underline{z}'\underline{z})^{-1}\underline{z}'v = \sum_j \frac{\lambda_j}{\lambda_j^2 + \alpha n}\langle v, \psi_j\rangle\phi_j, \tag{28}$$

and

$$\begin{aligned} \underline{z}'(\alpha n I_n + \underline{z}\underline{z}')^{-1}v &= \sum_j \frac{1}{\lambda_j^2 + \alpha n}\langle v, \psi_j\rangle\underline{z}'\psi_j \\ &= \sum_j \frac{\lambda_j}{\lambda_j^2 + \alpha n}\langle v, \psi_j\rangle\phi_j. \end{aligned} \tag{29}$$

Combining (28) and (29) yields the identity

$$\underline{z}'\left(\alpha n I_n + \underline{z}\underline{z}'\right)^{-1} = (\alpha n I_m + \underline{z}'\underline{z})^{-1}\underline{z}', \tag{30}$$

which allows to rewrite (26) as the standard ridge estimator defined as

$$\hat{\pi}_\alpha^r = (\alpha n I_m + \underline{z}'\underline{z})^{-1}\underline{z}'\underline{y}. \tag{31}$$

Therefore, for L^2 functions, the ridge estimator (31) corresponds to an approximation of the functional estimator (26) based on $\hat{M} = \frac{\underline{z}\underline{z}'}{n}$. The key difference is the inversion of an $m \times m$ matrix, rather than an $n \times n$ matrix for the functional estimator. As the number of discrete measurements m increases, the required inversion of the $m \times m$ matrix $(\alpha n I_m + \underline{z}'\underline{z})$ in (31) becomes increasingly more computationally demanding, whereas the computation of the functional estimator remains unaltered. In other words, our estimator enables a more precise and flexible approximation of matrix M,[7] and is faster to compute as soon as $m > n$.

[6] Interpolation is used when the discrete measurements along the curves are errorless, whereas smoothing may be used to remove some observational errors.

[7] In the MATLAB programs, we implement a fast numerical method to construct M based on the trapezoidal rule.

7. Case where Z is endogenous

It is frequent in economics that the regressor Z is endogenous. In that case, the least squares estimator studied in Section 2 is not consistent. In this section, we propose an estimator based on instrumental variables (IV) W that satisfy $cov(U, W) = 0$. For illustration, assume we want to estimate the price elasticity of electricity. To do so, we use as dependent variable Y the log consumption of electricity and as independent variable Z the log price of electricity. For big consumers (i.e. big firms), the electricity price might be endogenous. A reliable instrument would be given by W, the wind speed. Indeed wind influences the price because a stronger wind increases the electricity production of wind turbines but can be considered as exogenous because one does not control wind.

Under the assumption, $cov(U, W) = 0$, it follows that

$$C_{YW} = \Pi C_{ZW} \tag{32}$$

where $C_{YW} = E(Y\langle W, .\rangle)$ and $C_{ZW} = E(Z\langle W, .\rangle)$. Similarly, we have

$$C_{WY} = C_{WZ}\Pi^* \tag{33}$$

where $C_{WZ} = E(W\langle Z, .\rangle)$.

We need the following identification condition:

Assumption 1′. The observations $(U_i, W_i, Y_i, Z_i), i = 1, 2, \ldots, n$ are independent, identically distributed as (U, W, Y, Z) where U, W, and Y are random processes of a separable Hilbert space $\mathcal{E}$ and Z is a random process of a separable Hilbert space $\mathcal{F}$. Moreover, $E(U_i) = 0$, $cov(U_i, W_i) = 0$, $cov(U_i, U_j|Z_1, \ldots, Z_n, W_1, \ldots, W_n) = 0$ for all $i \neq j$ and $= V_U$ for $i = j$ where V_U is a nonrandom trace-class operator.

Assumption 7. C_{WZ} is injective.

Assumption 1′ imposes that the instruments W_i are exogenous.

Under Assumption 7, Π is uniquely defined from (32). To see this, assume that there are two solutions Π_1 and Π_2 to (32). It follows that $(\Pi_1 - \Pi_2)C_{ZW} = 0$ or equivalently $C_{WZ}\left(\Pi_1^* - \Pi_2^*\right) = 0$. Hence the range of $\left(\Pi_1^* - \Pi_2^*\right)$ belongs to the null space of C_{WZ}. However, under Assumption 7, the null space of C_{WZ} is reduced to zero and thus the range of $\left(\Pi_1^* - \Pi_2^*\right)$ is equal to zero. It follows that $\Pi_1^*\varphi - \Pi_2^*\varphi = 0$ for all φ, hence $\Pi_1^* = \Pi_2^*$.

To construct an estimator of Π^*, we first apply the operator C_{ZW} on the l.h.s and r.h.s of Eq. (33) to obtain

$$C_{ZW}C_{WY} = C_{ZW}C_{WZ}\Pi^*.$$

Note that $C_{ZW} = C_{WZ}^*$ and therefore the operator $C_{ZW}C_{WZ}$ is self-adjoint. The operators C_{ZW}, C_{WZ}, and C_{WY} can be estimated by their sample counterparts. The estimator of Π^* is defined by

$$\hat{\Pi}_\alpha^* = \left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW}\hat{C}_{WY}. \tag{34}$$

Similarly, the estimator of Π is given by

$$\hat{\Pi}_\alpha = \hat{C}_{YW}\hat{C}_{WZ}\left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}.$$

Now, we explain how to compute $\hat{\Pi}_\alpha^*$ in practice. From (34), we have

$$\left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)\hat{\Pi}_\alpha^*\psi = \hat{C}_{ZW}\hat{C}_{WY}\psi.$$

Note that

$$\hat{C}_{ZW}\hat{C}_{WY}\psi = \frac{1}{n^2}\sum_{i,j}\langle y_j, \psi\rangle\langle w_i, w_j\rangle z_i,$$

$$\hat{C}_{ZW}\hat{C}_{WZ}\hat{\Pi}_\alpha^*\psi = \frac{1}{n^2}\sum_{i,j}\left\langle z_j, \hat{\Pi}_\alpha^*\psi\right\rangle\langle w_i, w_j\rangle z_i.$$

Taking the inner product with z_l yields n equations

$$\alpha\left\langle z_l, \hat{\Pi}_\alpha^*\psi\right\rangle + \frac{1}{n^2}\sum_{i,j}\left\langle z_j, \hat{\Pi}_\alpha^*\psi\right\rangle\langle w_i, w_j\rangle\langle z_l, z_i\rangle$$
$$= \frac{1}{n^2}\sum_{i,j}\langle y_j, \psi\rangle\langle w_i, w_j\rangle\langle z_l, z_i\rangle, l = 1, 2, \ldots, n$$

with n unknowns $\left\langle z_j, \hat{\Pi}_\alpha^*\psi\right\rangle, j = 1, 2, \ldots, n$. Let υ be the $n-$ vector of $\left\langle z_j, \hat{\Pi}_\alpha^*\psi\right\rangle$, y_ψ be the $n-$ vector of $\langle y_j, \psi\rangle$ and Q be the $n \times n$ matrix with (l, j) element equal to $\frac{1}{n^2}\sum_i\langle w_i, w_j\rangle\langle z_l, z_i\rangle$. We obtain $\upsilon = (\alpha I + Q)^{-1}Qy_\psi = Q(\alpha I + Q)^{-1}y_\psi$.

Hence, for each ψ, $\hat{\Pi}_\alpha^*\psi$ can be computed from

$$\hat{\Pi}_\alpha^*\psi = \frac{1}{\alpha}\left[\hat{C}_{ZW}\hat{C}_{WY}\psi - \hat{C}_{ZW}\hat{C}_{WZ}\hat{\Pi}_\alpha^*\psi\right]$$
$$= y'_\psi(\alpha I + Q)^{-1}\xi$$

where ξ is the $n-$ vector with ith element $\frac{1}{n^2}\sum_j\langle w_i, w_j\rangle z_i$. The computation of $\hat{\Pi}_\alpha\varphi$ can be done using the same approach as in Section 2.

Assumption 8. $C_{ZW}C_{WZ}$ is a trace-class operator and $\|\hat{C}_{ZW}\hat{C}_{WZ} - C_{ZW}C_{WZ}\|_{HS}^2 = O_p(1/n)$.

Assumption 9. There is a Hilbert–Schmidt operator R from $\mathcal{E}$ to $\mathcal{F}$ and a constant $\beta > 0$ such that $\Pi^* = (C_{ZW}C_{WZ})^{\beta/2}R$.

Like Assumption 4, Assumption 9 is a source condition which permits to characterize the rate of convergence of $\hat{\Pi}_\alpha^*$. It also guarantees that Π^* belongs to the orthogonal of the null space of C_{WZ} and hence, under Assumption 7, Assumption 9 is not needed to identify Π^*.

We decompose $\hat{\Pi}_\alpha^* - \Pi^*$ in the following manner:

$$\hat{\Pi}_\alpha^* - \Pi^*$$
$$= \left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW}\hat{C}_{WY} - \Pi^*$$
$$= \left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW}\hat{C}_{WU} \tag{35}$$
$$+ \left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW}\hat{C}_{WZ}\Pi^*$$
$$- (\alpha I + C_{ZW}C_{WZ})^{-1}C_{ZW}C_{WZ}\Pi^* \tag{36}$$
$$+ (\alpha I + C_{ZW}C_{WZ})^{-1}C_{ZW}C_{WZ}\Pi^* - \Pi^*. \tag{37}$$

Proposition 7. *Under Assumption 1′, 2, 8, and 9, the MSE of $\hat{\Pi}_\alpha^* - \Pi^*$ has the same rate of convergence as in Proposition 2.*

Remark 7.1. Proposition 7 shows that our estimator based on instrumental variables is consistent and its rate is the same as in the exogenous case.

Remark 7.2. To the best of our knowledge, Model (1) with endogeneity has never been studied before, therefore we are the first to provide a consistent estimator in that setting.

Remark 7.3. In the case where C_{WZ} is not injective, one could replace Assumption 9 by Assumption 9′.

Assumption 9′. Let $\Pi = \Pi_N + \Pi_{N^\perp}$ where $\Pi_N \in \mathcal{N}(C_{WZ})$ and $\Pi_{N^\perp} \in \mathcal{N}^\perp(C_{WZ})$. There is a Hilbert–Schmidt operator R from $\mathcal{E}$ to $\mathcal{F}$ and a constant $\beta > 0$ such that $\Pi_{N^\perp}^* = (C_{ZW}C_{WZ})^{\beta/2}R$.

Under Assumption 9′, we could establish a result similar to that of Proposition 3, namely that $\hat{\Pi}_\alpha^*$ converges to $\Pi_{N^\perp}^*$ with the same rate of convergence as before.

Table 1
Simulation results: Mean-square errors over 100 replications.

Std errors	Empirical MSE				Squared bias	Coef. of d.	
	$n=50$	$n=100$	$n=500$	$n=1000$	$\|\Pi-\Pi_\alpha\|_{HS}^2$	R^2	$\tilde{R}^2$
$\sigma_u=0.1$	.0154	.0135	.0126	.0124	.0095	.995	.995
	(.0027)	(.0017)	(.0008)	(.0005)			
$\sigma_u=0.25$	.0291	.0205	.0138	.0130	.0095	.976	.976
	(.0098)	(.0063)	(.0022)	(.0013)			
$\sigma_u=0.5$	.0773	.0438	.0194	.0156	.0095	.910	.911
	(.0363)	(.0193)	(.0057)	(.0028)			
$\sigma_u=1$	.2909	.1354	.0371	.0257	.0095	.712	.724
	(.1789)	(.0659)	(.0161)	(.0089)			
$\sigma_u=2$	.9128	.4755	.1245	.0668	.0095	.383	.423
	(.5495)	(.2607)	(.0660)	(.0378)			

Note: Standard deviations are reported in parentheses.

8. Simulations

This section consists of a simulation study of the estimators presented earlier. Let $\mathcal{E}=\mathcal{F}=L^2[0,1]$ and $\mathcal{S}=\mathcal{T}=[0,1]$. Π is an integral operator from to $L^2[0,1]$ to $L^2[0,1]$ with kernel $\pi(s,t)=1-|s-t|^2$.[8] We consider an Ornstein–Uhlenbeck process with zero mean and mean reversion rate equal to one to represent the error function. It is described by the differential equation $dU(s)=-U(s)ds+\sigma_u dG_u(s)$, for $s\in[0,1]$ and where G_u is a Wiener process and σ_u denotes the standard deviation of its increments dG_u. Note that this error function is stationary.

We study the model

$$Y_i=\Pi Z_i+U_i,\qquad i=1,\ldots,n$$

in two different settings. First, we consider design functions uncorrelated to the error functions ($cov(U,Z)=0$), then investigate the case where Z is endogenous ($cov(U,Z)\neq 0$).

8.1. Exogenous predictor functions

We consider the design function

$$Z_i(t)=\frac{\Gamma(\alpha_i+\beta_i)}{\Gamma(\alpha_i)+\Gamma(\beta_i)}t^{\alpha_i-1}(1-t)^{\beta_i-1}+\eta_i$$

for $t\in[0,1]$, with $\alpha_i,\beta_i\sim iid\ U[2,5]$ and $\eta_i\sim iid\ N(0,1)$, for all $i=1,\ldots,n$. These predictor functions are probability density functions of some random beta distributions over the interval $[0,1]$, with an additive Gaussian term.

The numerical simulation is performed as follows:

1. Construct both a pseudo-continuous interval of $[0,1]$, denoted $\mathcal{T}$, consisting of 1000 equally-spaced discrete steps, and a discretized interval of $[0,1]$, denoted $\tilde{\mathcal{T}}$, consisting of only 100 equally-spaced discrete steps.
2. Generate n predictor functions $z_i(t)$ and error functions $u_i(s)$, where $t,s\in\mathcal{T}$ so as to obtain pseudo-continuous functions.
3. Generate the n response functions $y_i(s)$ using the specified model where $s\in\mathcal{T}$.
4. Generate the sample of n discretized pairs of functions $(\tilde{z}_i,\tilde{y}_i)$ by extracting the corresponding values of the pairs (z_i,y_i) for all $t,s\in\tilde{\mathcal{T}}$.
5. Estimate Π using the regularization method on the sample of n pairs of functions $(\tilde{z}_i,\tilde{y}_i)$ (discrete points are connected by linear interpolation to obtain curves) and a fixed smoothing parameter $\alpha=.01$.
6. Repeat steps 2–5 100 times and calculate the MSE by averaging the quantities $\|\hat{\Pi}_\alpha-\Pi\|_{HS}^2=\int_{\tilde{\mathcal{T}}}\int_{\tilde{\mathcal{T}}}(\hat{\pi}_\alpha(s,t)-\pi(s,t))^2dtds$ over all repetitions.

All numerical integrations are performed using the trapezoidal rule (i.e. piecewise linear interpolation) although it is possible to use other quadrature rules (such as another Newton–Cotes rule or adaptive quadrature). In addition, the simulations of the stochastic processes for the error terms are constructed using the Euler–Maruyama method for approximating numerical solutions to stochastic differential equations.

Fig. 1 shows 10 discretized predictor functions (z_i), Ornstein–Uhlenbeck error functions for $\sigma_u=1$ (u_i), response functions (y_i) and an example of a response function for various values of σ_u.

Table 1 reports the MSE for 4 different sample sizes ($n=50,100,500,1000$) and 5 values of the standard deviation parameter ($\sigma_u=0.1,0.25,0.5,1,2$). Naturally, the use of a fixed smoothing parameter $\alpha=.01$ that is independent of the sample size prevents the MSE from converging toward zero. In fact, the MSE converges to $\|\Pi-\Pi_\alpha\|_{HS}^2$, which is a measure of the squared bias introduced by the regularization method.[9] The last two columns of Table 1 report the true global (R^2) and extended local ($\tilde{R}^2$) functional coefficients of determination, defined as

$$R^2=\frac{\int_{\mathcal{S}}\mathrm{var}(E[Y(s)|Z])ds}{\int_{\mathcal{S}}\mathrm{var}(Y(s))ds}=\frac{\int_{\mathcal{S}}\mathrm{var}(\Pi Z(s))ds}{\int_{\mathcal{S}}\mathrm{var}(Y(s))ds}$$

$$\tilde{R}^2=\int_{\mathcal{S}}\frac{\mathrm{var}(E[Y(s)|Z])ds}{\mathrm{var}(Y(s))ds}=\int_{\mathcal{S}}\frac{\mathrm{var}(\Pi Z(s))ds}{\mathrm{var}(Y(s))ds},$$

which are directly related to those proposed in Yao et al. (2005).[10]

Simulations results are in line with the theoretical results. We observe that, for a fixed α, the MSE decreases as the sample size grows. Further, the coefficients of determination decrease as the standard deviation of the error increases, since the estimation is made more difficult. As a result, the MSE grows with σ_u.

For illustration purposes, we provide two sets of surface plots. Fig. 2 shows 3D-plots of the actual kernel (top-left), the regularized kernel (top-right), their superposition (bottom-left) and the bias computed as their difference (bottom-right). The Tikhonov regularization appears to introduce most of the bias on the edges of the kernel.

Fig. 3 shows the mean estimated kernel for $n=500$ and Ornstein–Uhlenbeck errors with $\sigma_u=1$ (top-left), against the true

[8] Simulations have also been performed using different kernels. In particular, we have considered multiple kernels, allowing to include multiple functional predictors in a single functional model. Results suggest that the performance of the estimator is analogous in "multivariate" functional linear regression.

[9] The magnitude of this bias depends on both the design functions and the value of α since $\Pi_\alpha=(\alpha I+V_Z)^{-1}V_z\Pi$. We perform Monte-Carlo simulations to approximate the regularized operator Π_α using 100 random samples of 1000 z_i's.

[10] These true coefficients are approximated by their mean values using 1000 random functions over 100 simulations. In practice (when the true Π is unknown) it is possible to use a consistent estimators of those coefficients by using $\hat{\Pi}_\alpha$ and the sample counterpart of variance operators.

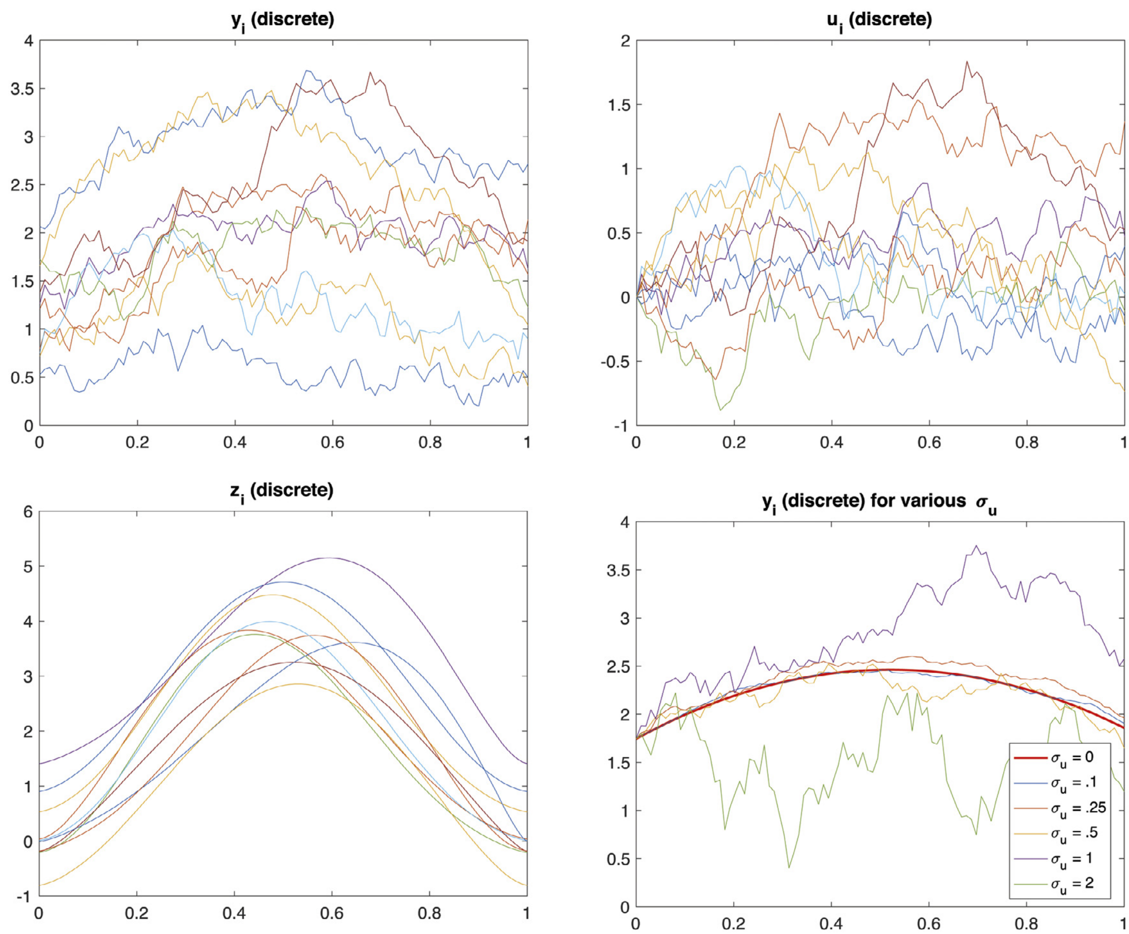

Fig. 1. Examples of simulated functions (top left: discretized y_i; top right: discretized u_i for $\sigma_U = 1$, bottom left: discretized z_i, bottom right: a single y_i for various σ_u).

kernel (bottom-left), against the regularized kernel (top-right), and its mean errors with respect to the true kernel (bottom-right). One may observe that the mean estimate is relatively close to the regularized kernel. However it does not perform well on the edges when compared to the true kernel.[11]

Let us now turn to the case where Z is endogenous.

8.2. Endogenous predictor functions

We consider the design function

$$Z_i(t) = bW_i(t) + \xi_i(t),$$

where $\xi_i(t) = aU_i(t) + c\epsilon_i(t)$ and the instrument w_i is defined as

$$W_i(t) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i) + \Gamma(\beta_i)} t^{\alpha_i - 1}(1-t)^{\beta_i - 1} + \eta_i$$

for $t \in [0, 1]$, $\alpha_i, \beta_i \sim iidU[2, 5]$ and $\eta_i \sim iid\mathcal{N}(0, 1)$, for all $i = 1, \dots, n$. Moreover, U_i and ε_i are Ornstein–Uhlenbeck processes with standard deviation parameters $\sigma_u = \sigma_\varepsilon = 1$. It is easily shown that ξ_i is also an Ornstein–Uhlenbeck process with unit mean-reversion rate described by the differential equation $d\xi(t) = -\xi(t) + \sqrt{a^2\sigma_u^2 + c^2\sigma_\varepsilon^2}dG_\xi(t)$. We further assume $a = 1, b \in [0, 1]$ and c such that $\int_{\mathcal{S}} \mathrm{var}(Y(s))ds$ is unchanged as b varies.[12] Hence, the choice of b amounts to that of the instrument's strength.

[11] A possible solution to the boundary effect would be to estimate the kernel on a larger support than necessary and then truncate.

[12] This assumption allows to keep the variance of Y stable when varying instrument strength. It implies $c = \sqrt{1 + (1 - b^2)\frac{\int_{\mathcal{S}} \mathrm{var}(\Pi W(s))ds}{\int_{\mathcal{S}} \mathrm{var}(\Pi \varepsilon(s))ds}}$.

The numerical simulation design is slightly modified so as to incorporate the generation of the instruments W and the dependence between Z and U:

1. Construct both a pseudo-continuous interval of [0, 1], denoted $\mathcal{T}$, consisting of 1000 equally-spaced discrete steps, and a discretized interval of [0, 1], denoted $\tilde{\mathcal{T}}$, consisting of only 100 equally-spaced discrete steps.
2. Generate n instrument functions $w_i(t)$ and error functions $u_i(s)$ and $\varepsilon_i(s)$, where $t, s \in \mathcal{T}$ so as to obtain pseudo-continuous functions.
3. Generate n predictor functions $z_i(t)$ using the design specified above, where $t, s \in \mathcal{T}$ so as to obtain pseudo-continuous functions.
4. Generate the n response functions $y_i(s)$ using the specified model where $s \in \mathcal{T}$.
5. Generate the sample of n discretized pairs of functions $(\tilde{w}_i, \tilde{z}_i, \tilde{y}_i)$ by extracting the corresponding values of the pairs (w_i, z_i, y_i) for all $t, s \in \tilde{\mathcal{T}}$.
6. Estimate Π using the regularization method on the sample of n triplets of functions $(\tilde{w}_i, \tilde{z}_i, \tilde{y}_i)$ and a fixed smoothing parameter $\alpha = .01$.
7. Repeat steps 2–5 100 times and calculate the *MSE* by averaging the quantities $\|\hat{\Pi}_\alpha - \Pi\|_{HS}^2 = \int_{\tilde{\mathcal{T}}}\int_{\tilde{\mathcal{T}}}(\hat{\pi}_\alpha(s, t) - \pi(s, t))^2 dtds$ over all repetitions.

Table 2 reports the *MSE* for 4 different sample sizes ($n = 50, 100, 500, 1000$) and 4 values of b when estimating the model without accounting for the endogeneity of Z using the estimator described in Section 2.2. Unsurprisingly, the estimation errors are

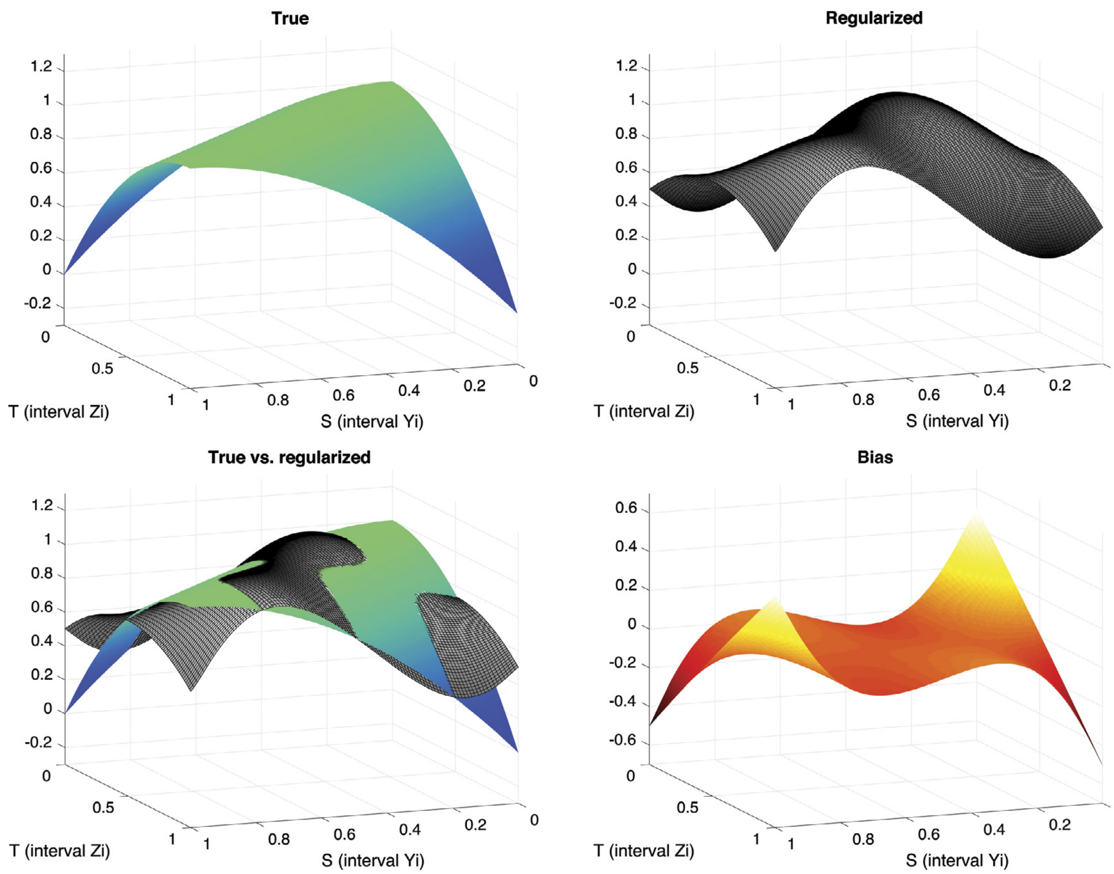

Fig. 2. True kernel vs. regularized kernel (top left: True; top right: Regularized, bottom left: True vs. regularized, bottom right: Bias).

Table 2
Non-IV estimator: Mean-square errors over 100 replications.

Instr. strength	Empirical MSE				Squared bias	Coef. of deter.	
	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$\lVert \Pi - \Pi_\alpha \rVert_{HS}^2$	R^2	$\tilde{R}^2$
$b = 0.25$	2.4834	1.4642	.4690	.3214	.0060	.5144	.5461
$(c = 2.3)$	(.4678)	(.2011)	(.0435)	(.0317)			
$b = 0.5$	2.3346	1.4504	.5826	.4541	.0027	.5140	.5450
$(c = 1.96)$	(.4014)	(.2416)	(.0679)	(.0460)			
$b = 0.75$	2.1858	1.5363	.8535	.7529	.0011	.5294	.5591
$(c = 1.55)$	(.4825)	(.2974)	(.1027)	(.0640)			
$b = 1$	2.4219	2.0547	1.6583	1.6310	.0006	.5633	.5919

Note: Standard deviations are reported in parentheses.

important. The squared bias is smaller to that of the previous design and decreases with b. The last two columns report R^2 and $\tilde{R}^2$ for the full model. They are relatively stable since $\int_S \mathrm{var}(Y(s))ds$ is fixed.

We now turn to the simulations results for the instrumental variable estimator described in Section 7. Table 3 reports the MSE's along with R^2 and the squared regularization biases. Squared biases are fairly small in this setup. This is related to the covariance operator of the predictor functions. R^2_{FS} denotes the first-stage regression's coefficient of determination. It shows how b relates to the instrument's strength. Naturally, weaker instruments are associated with larger MSE's, although the spread seems to vanish rather quickly in this setup.

For comparisons with the exogenous case, we provide a final set of surface plots. Fig. 4 shows 3D-plots of the mean IV estimated kernel (top-left), the mean non-IV (top-right), the superposition of the mean IV and the true kernels (bottom-left) and the mean estimation errors computed as the difference between the true kernel and the mean IV estimate (bottom-right). Note that the mean IV estimate is relatively close to the actual kernel, whereas the estimate when neglecting endogeneity exhibits a large bias.

9. Application

9.1. Introduction

This section presents an empirical application of the functional linear regression model with functional response to the study of the dynamics between daily electricity consumption and temperature patterns. There is a tradition of applications to electricity in functional data analysis (Ferraty and Vieu, 2006; Antoch et al., 2010; Andersson and Lillestol, 2010; Liebl, 2013), although mostly focused on forecasting. Rather, we propose an application illustrating the usefulness of our estimator for inference.

If electricity users were to optimize their consumption with respect to indoor air temperature on a real-time basis, the dynamics of aggregate electricity demand would depend on both past

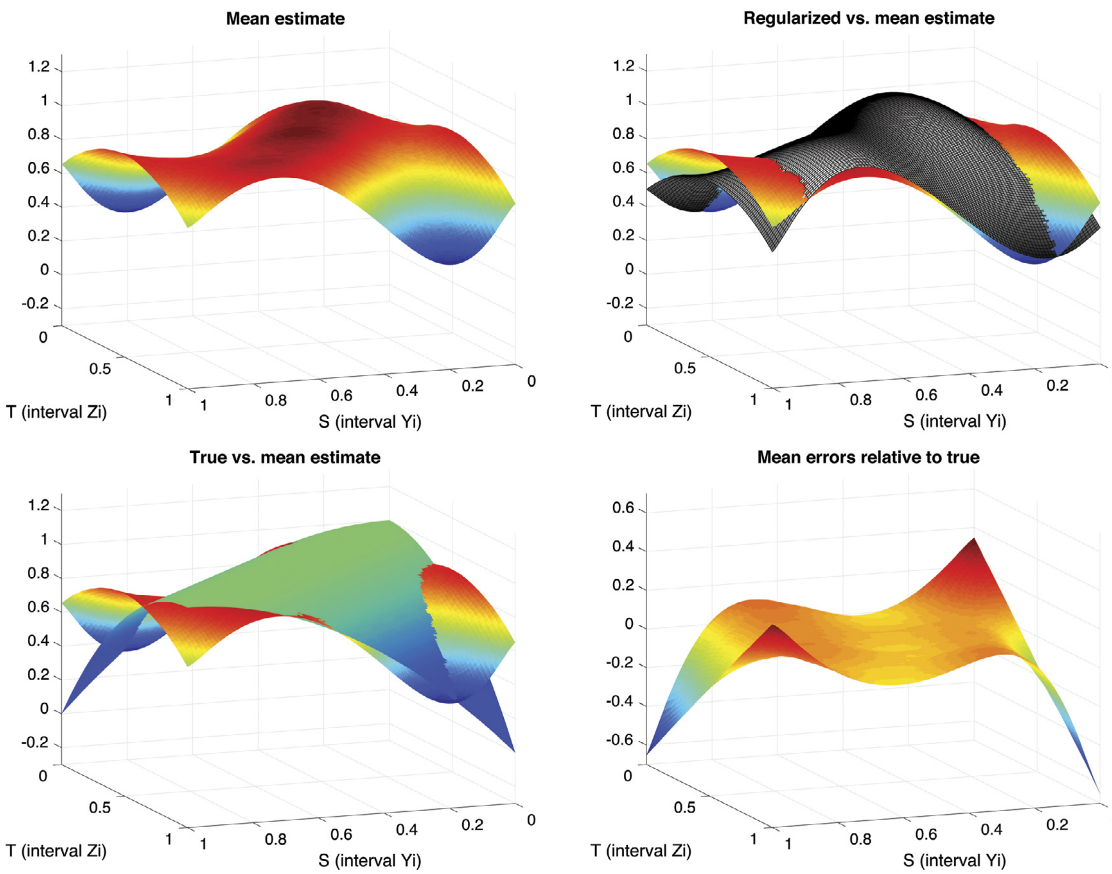

Fig. 3. True kernel vs. mean estimate (100 runs with $n = 500$, $\sigma_u = 1$) (top left: Mean estimate, top right: Regularized vs. mean estimate, bottom left: True vs. mean estimate, bottom right: Mean errors).

Table 3
IV estimator: Mean-square errors over 100 replications.

Instr. str.	Empirical MSE				Squared bias	Coef. of d.	
	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$\Vert\Pi - \Pi_\alpha\Vert^2_{HS}$	R^2	R^2_{FS}
$b = 0.25$	.2383	.1710	.0752	.0542	.0060	.0175	.0246
$(c = 2.3)$	(.2019)	(.1422)	(.0779)	(.0209)			
$b = 0.5$	.1040	.0619	.0315	.0276	.0027	.0737	.1092
$(c = 1.96)$	(.0859)	(.0349)	(.0099)	(.0053)			
$b = 0.75$	.0682	.0444	.0242	.0216	.0011	.1767	.2683
$(c = 1.55)$	(.0364)	(.0203)	(.0044)	(.0028)			
$b = 1$	.0466	.0330	.0211	.0199	.0006	.3287	.5048

Note: Standard deviations are reported in parentheses.

and future weather realizations. Let us consider the reduced-form model for the daily aggregate electricity demand trajectory Y as a function of the outside air temperature pattern Z, given by

$$Y(s) = \pi_0(s) + \int_{\mathcal{T}} \pi_1(s,t)Z(t)dt + U(s), \tag{38}$$

where Y, π_0, Z and U are L^2 functions of time indices s and t, although defined over possibly different time intervals $\mathcal{S}$ and $\mathcal{T}$. We will consider the case where $\mathcal{S} \subset \mathcal{T}$. That is, aggregate electricity consumption at time s, $Y(s)$, may possibly depend on current ($t = s$), past ($t < s$) and future ($t > s$) temperature levels $Z(t)$.

Abstracting from the uncertainty surrounding future realizations,[13] this reduced-form model can be related to the Euler equation of a continuous time model with partial adjustment dynamics based on the literature on dynamic linear rational expectations models (Muth, 1961; Kennan, 1979; Hansen and Sargent, 1980).[14]

This model can be estimated using the procedure developed in this paper if one has an i.i.d. sample of functional observations $\{y_i, z_i\}_{i=1,\ldots,n}$, and if u_i can be presumed a mean-zero i.i.d. random functional error process. We propose to estimate the model using daily patterns extracted from aggregate electricity consumption and temperature trajectories at the level of the Canadian province of Ontario. Albeit using successive days may induce autocorrelation in the error which is not taken into account in our theory, we believe that our estimator is still consistent in this setting.

13 A model like $Y(s) = \pi_0(s) + \int_0^s \pi_1(s,t)Z(t)dt + \int_s^1 \pi_2(s,t)E[Z(t)|I_s]dt$, which explicitly accounts for expectations (where I_s denotes the information set at time s), could be estimated with our method, although with significant departures from the original model.

14 The electrical engineering literature on home energy management develops rational expectations models for smart thermostats that optimize real-time energy consumption with respect to past weather measurements and available forecasts, given individual preferences settings (McLaughlin et al., 2011; Yu et al., 2013).

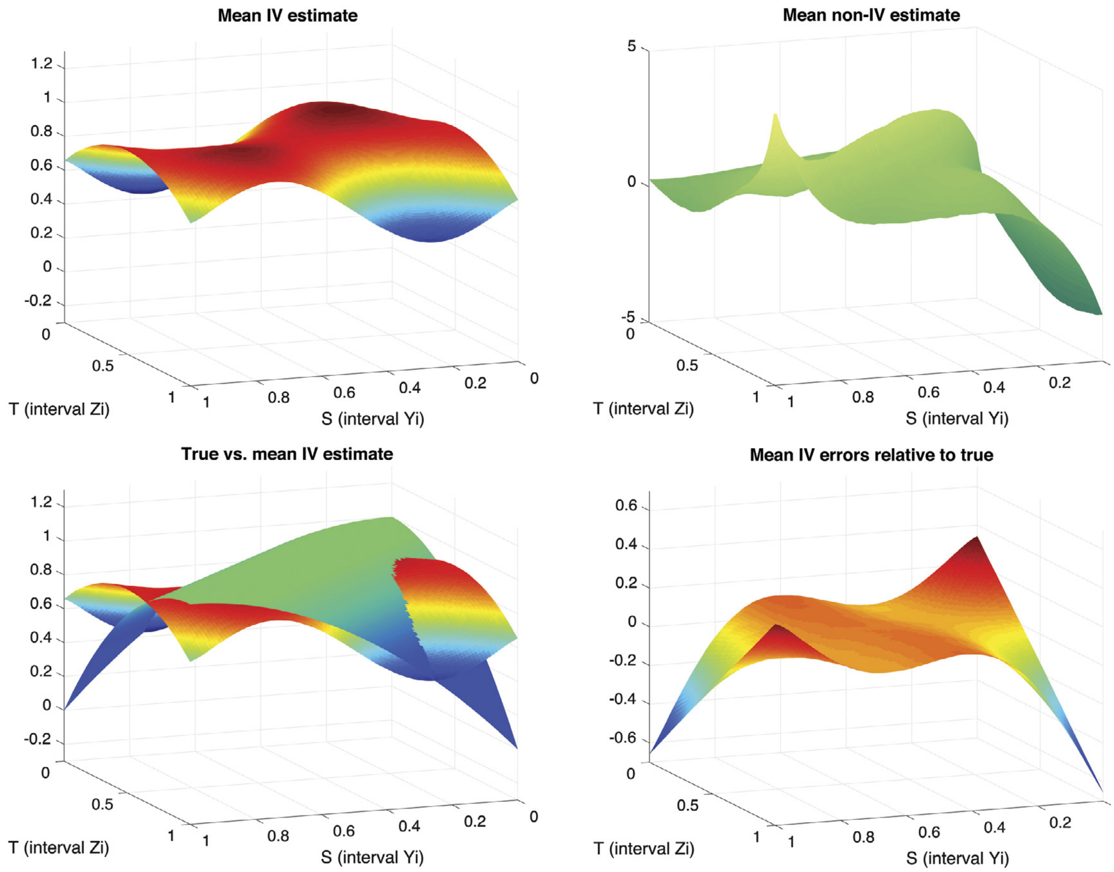

Fig. 4. True kernel vs. mean IV estimate (100 runs with $n = 500$, $\sigma_u = 1$ and $b = 0.75$) (top left: Mean estimated IV; top right: Mean estimated non-IV, bottom left: True vs. mean IV estimate, bottom right: Mean IV errors).

The remainder of the section describes institutional features of Ontario's electricity market, then provides details on data construction before performing a preliminary data analysis. Finally, estimation results are interpreted using contour plots and compared to OLS estimates.

9.2. *Electricity demand in Ontario*

Prior to presenting the data set construction, let us present some facts about the electricity market in Ontario. The market defines two categories of consumers. First, small consumers (residential end-users and small businesses) are billed for electricity usage by their local distribution company. The vast majority pays fixed time-of-use rates which are updated from season to season. The Ontario's energy markets regulator (Ontario Energy Board), defines the winter period from November 1 to April 30 and has two daily peak periods: one in the morning (7 am–11 am) and the other after worktime (5 pm–7 pm), a mid-peak period (11 am–5 pm) and an off-peak period (7 pm–7 am). The remaining part of the year is defined as the summer period and mid-peak and peak periods are reversed with respect to winter.

Second, large consumers[15] (large businesses and the public sector) are subject to the wholesale market price, determined on an hourly basis in a uniform-price multi-unit auction subject to operational constraints. The wholesale market price being quite volatile, some large consumers choose to go with retail contractors to avoid market risk exposure, although the bulk of electricity trade goes through the wholesale market. Furthermore, all large consumers must also pay the monthly *Global Adjustment* which represents other charges related to market, transport and regulatory operations.

Consequently, it is difficult to evaluate the extent to which aggregate electricity consumption depends on the wholesale price and the time-of-use price. Furthermore, short-term demand for electricity is typically perceived as being very price inelastic (see for instance Faruqui et al., 2013). For these reasons, we decide to abstract from the price effects.

9.3. *Preliminary data analysis*

The original data set consists of hourly observations of real-time aggregate electricity consumption and weighted average temperature in Ontario from January 1, 2010 to September 30, 2014. Hourly power data for Ontario are publicly available on the system operator's website, http://www.ieso.ca. Hourly province-wide temperature values have been constructed from hourly measurements at 77 weather stations in Ontario,[16] publicly available on Environment Canada's website, http://climat.meteo.gc.ca/.

Let $Z(t)$ denote our measure of temperature at time t for the entire province. It is constructed in four steps. First, we match a set of 41 Ontarian cities (of above 10,000 inhabitants)[17] to their three nearest weather stations. Second, we compute a weighted average using a distance metric. Third, we obtain $Z(t)$ as a weighted average of cities' temperatures, where weights are defined by each

15 Roughly speaking, businesses are considered large when their electricity bills exceed \$2000 per month.

16 The complete data set contained 139 weather stations although once matched to neighboring cities, only 77 were found relevant.

17 Those cities represent 85.3% of the province's population as of 2011.

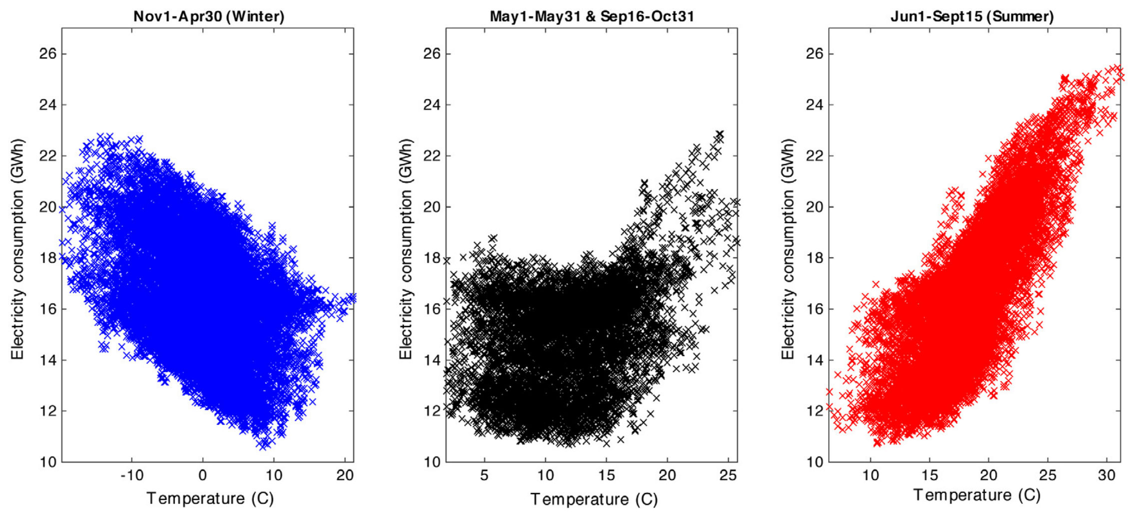

Fig. 5. Electricity consumption (GWh) vs. temperature (C).

Table 4
Descriptive statistics.

Year	Hourly temperature (C)				Hourly consumption (GW)				
	Mean	SD	Min	Max	Mean	SD	Min	Max	Obs
2010	8.27	9.17	−17.12	29.13	16.23	2.59	10.62	25.07	8,760
2011	8.02	9.37	−17.96	31.15	16.15	2.45	10.76	25.43	8,760
2012	9.32	8.67	−14.32	30.57	16.09	2.40	10.99	24.60	8,784
2013	7.55	9.36	−17.40	28.88	16.07	2.39	10.77	24.49	8,760
2014	7.58	11.02	−19.84	26.52	16.08	2.36	10.71	22.77	6,552
All	8.18	9.49	−19.84	31.15	16.12	2.45	10.62	25.43	41,616

city's relative population. The constructed province-wide hourly temperature variable is formally defined by

$$Z(t) = \sum_c \gamma_c \left\{ \sum_{w(c)} \rho_{w(c)} Z_{w(c)}(t) \right\}, \qquad \forall i, h$$

where $\gamma_c = \frac{Pop_c}{(\sum_j Pop_j)}$ is city c's weight, $\rho_{w(c)} = \frac{((lat_c - lat_{w(c)})^6 + (lon_c - lon_{w(c)})^6)^{-1}}{\sum_{l(c)} ((lat_c - lat_{l(c)})^6 + (lon_c - lon_{l(c)})^6)^{-1}}$ is station w's weight for city c's temperature average, lat denotes latitude, lon longitude, and $Z_w(t)$ is the temperature measurement at station w in hour t.[18] Finally, we use robust locally weighted polynomial regression on the constructed temperature series in order to smooth implausible jumps, which are most likely due to measurement errors. Table 4 reports descriptive statistics for hourly electricity consumption and our constructed measure of temperature. Unsurprisingly, we observe some correlation between annual consumption peaks and maximum temperatures.

Fig. 5 displays the relationships between electricity consumption and our constructed temperature measure for the periods spanning from November 1 to April 30 (blue), May 1 to May 31 (black), June 1 to September 15 (red) and September 16 to October 31 (black). The plots show evidence of a relatively linear relationship between electricity consumption and temperature for winter and summer months. On the other hand, the relation is much flatter in October and May.

The original data series are presented in Fig. 6. The plots in Figs. 5 and 6 suggest that power usage is more sensitive to warm weather than cold weather. Probably because cooling can be more energy-intensive than heating, but more importantly because electricity represents a small share of heating fuel for residential users in Ontario. For this latter reason, we will focus the analysis on summer months, i.e. the period from June 1 to September 15.

Hence, we extract the summer periods for each year from 2010 to 2014 and proceed to construct the functional data sample. It consists of 368 daily trajectories of 25 discrete observations (from midnight to midnight) for the dependent variable and a three-day window of 73 observations for the predictor variable. This window is chosen so that the dependent variable will always be regressed on at least 24 lagged hours and 24 future hours. One can expect significant correlation between successive hourly temperature measurements.

We discard weekends as well as statutory holidays in Ontario in summer months. For ease of interpretation, the temperature variable is transformed into cooling degrees defined as $Z_c(t) = Z(t) - \min_{Z \in \text{Summer}} (Z(t))$, with $\min_{Z \in \text{Summer}} (Z(t)) = 6°C$. The estimation sample (years 2010–2013) consists of 295 functional observations, and is presented in Fig. 7. The data used for out-of-sample prediction (year 2014) contains 73 functional observations, and is shown in Fig. 8. The bold lines show the sample mean trajectories. Peak mean summer temperatures occurs around 3 pm, whereas peak mean consumption occurs between 4 pm to 5 pm.

9.4. The functional model and estimation results

Following the preceding discussion, we specify $\mathcal{E} = L^2[0, 24]$ and $\mathcal{F} = L^2[-24, 48]$ with $\mathcal{S} = [0, 24]$ and $\mathcal{T} = [-24, 48]$. The functional linear regression model of interest is given by

$$Y_i(s) = \pi_0(s) + \int_{-24}^{48} \pi_1(s, t) Z_i(t) dt + \sum_{j \in J} \beta_j d_{ij} + U_i(s), \tag{39}$$

[18] The distance metric is the sum of differences in geographic coordinates to the sixth power. The exponent is chosen so as to put arbitrarily more weight on nearby weather stations with respect to those located further away from the cities.

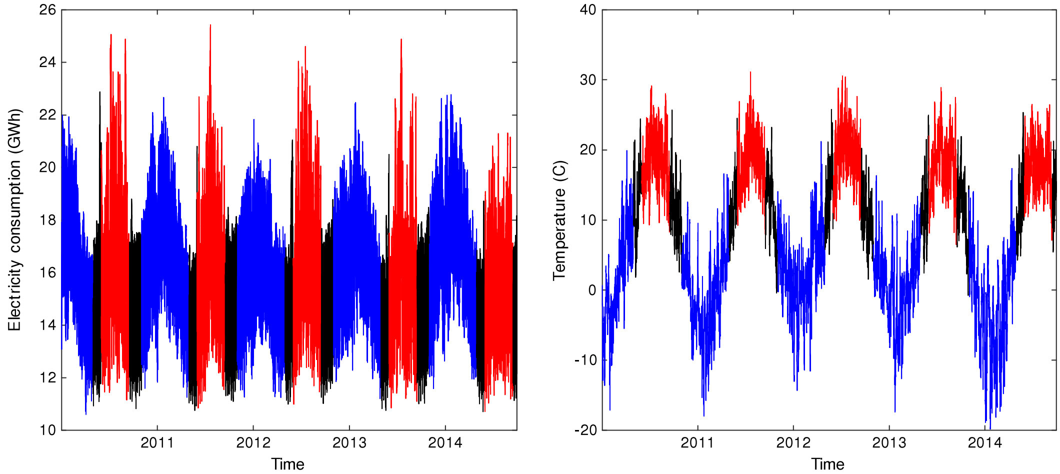

Fig. 6. Entire data series.

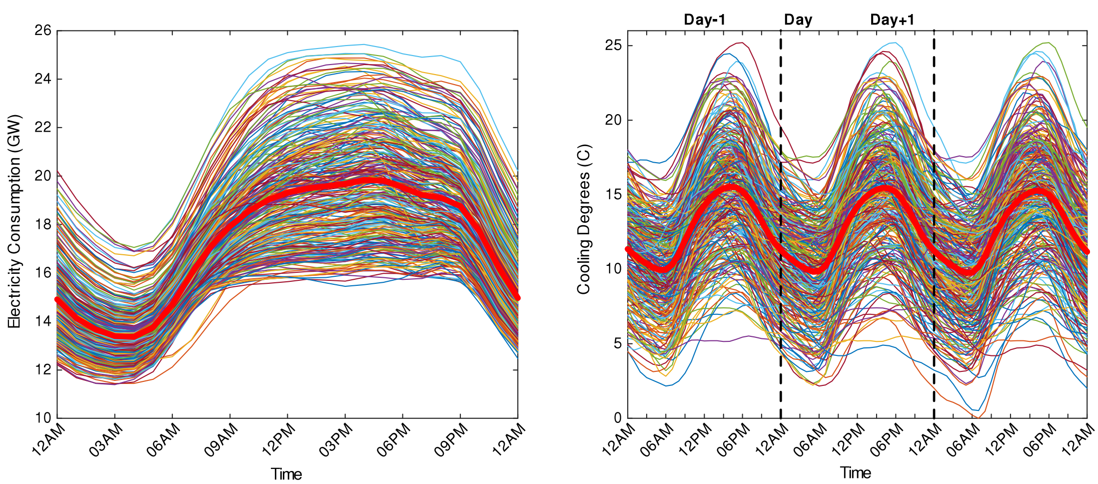

Fig. 7. Estimation sample (295 functional observations): Y_i's (left) and Z_i's (right).

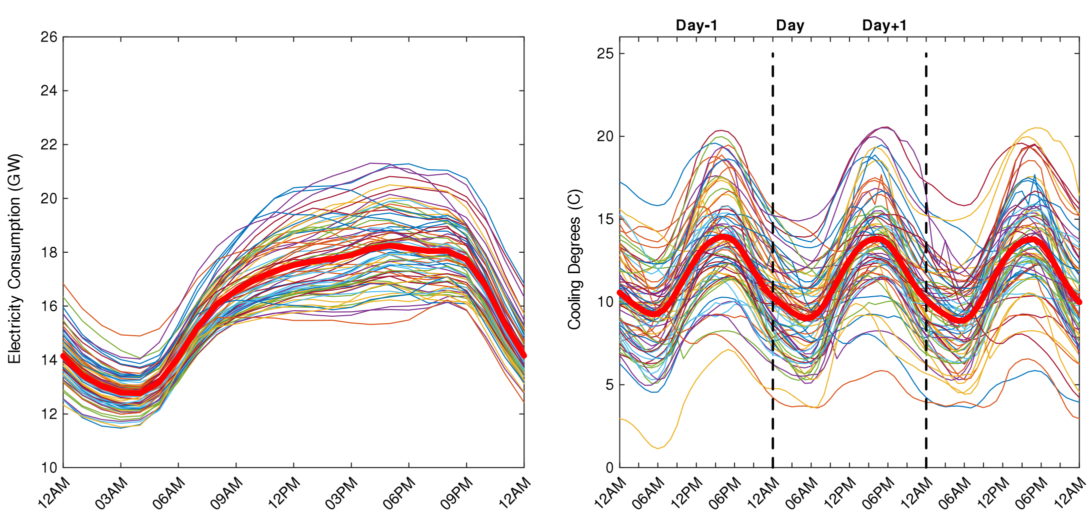

Fig. 8. Prediction sample (73 functional observations): Y_i's (left) and Z_i's (right).

where $Y_i(s)$ is the aggregate electricity consumption at time $s \in \mathcal{S}$ for day i, $\pi_0(s)$ is a constant function, $Z_i(t)$ is the temperature at time $t \in \mathcal{T}$, and $U_i(s)$ is a zero-mean error term. We also include a set of J binary variables d_{ij} aimed at capturing unobserved seasonalities, in particular for years, months of the year, weekdays and hours of the day. The object of interest is the kernel π_1, which

Table 5
Mean squared prediction errors for summer 2014.

	Func Reg		OLS	
	α_{CV}	α_{BP}	Hourly	3 hourly
α	0.15	3.40		
MSPE	16.14	15.85	28.39	17.20
Cond(Z'Z/n)			1.40e+06	3.71e+03

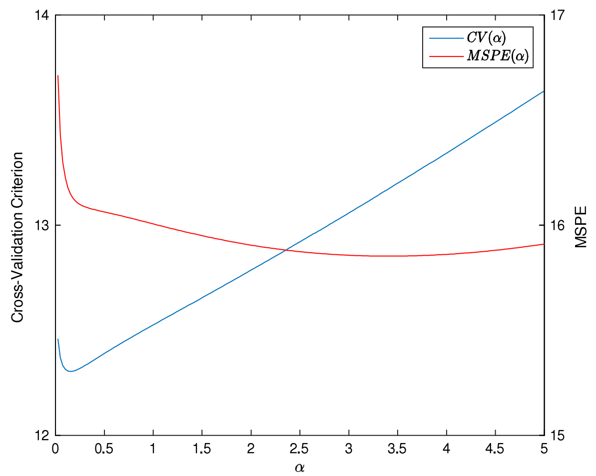

Fig. 9. CV criterion (blue) and MSPE (red) as functions of α. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

characterizes the dynamic relation between electricity consumption for air conditioning needs and temperature patterns. Since we are interested in daily electricity consumption patterns, the cross-sectional unit i denotes a daily functional observation.

We compare the performance of our estimator with that of the OLS estimator of the discrete analogous model given by

$$Y_i(s) = \pi_0(s) + \sum_{t=-24}^{48} \pi_1(s,t)Z_i(t) + \sum_{j\in\mathcal{J}} \beta_j d_{ij} + U_i(s). \tag{40}$$

To assess the performance of our estimator, we do an out-of-sample prediction exercise. First, data from 2010 to 2013 are used to estimate Π, let $\hat{\Pi}_\alpha$ be the resulting estimator. Then, the prediction of Y_i over 2014 is given by $\hat{Y}_i = \hat{\Pi}_\alpha Z_i$ where Z_i is the actual temperature observed in 2014. The Mean Squared Prediction Error (MSPE) for 2014 is computed as $\sum_i \int \left(\hat{Y}_i(s) - Y_i(s)\right)^2 ds$ where Y_i is the actual electricity consumption. Table 5 reports MSPE associated with four alternative estimators: (a) our functional estimator computed with α_{CV}, obtained with the leave-one-out CV method based on 2010–2013 data and (b) with α_{BP} which minimizes the MSPE for 2014, (c) the OLS estimator based on hourly predictors and (d) the OLS estimator based on only 24 predictors (3-hourly data). We find that the functional estimator performs better in terms of MSPE. The condition number[19] reported in Table 5 shows that the matrix $Z'Z/n$ is severely ill-conditioned. This explains the poor performance of OLS. The cross-validation criterion and the MSPE are plotted as functions of the tuning parameter in Fig. 9.

Fig. 10 displays contour plots of the 3-hourly OLS kernel estimate (left) and the regularized kernel estimate for α_{CV} (right). In order to ease interpretation of the results, we plot indicative dashed lines to separate out the daily windows and add a diagonal so as to emphasize the contemporaneous relation between the functions of interest. Estimates may be read both horizontally and vertically. The effects of the entire temperature pattern on electricity consumption at a given hour of the day are observed horizontally, whereas the effect of temperature at a specific time upon the daily electricity consumption pattern is read vertically. It is important to emphasize the ceteris paribus nature of those point estimates. Each corresponds to the additional effect of a slight increase in temperature at a specific time, e.g. 12 pm, on the electricity demand at a given time, e.g. 13 pm, *holding everything else (i.e. the temperature at any other point in time) constant*. The magnitudes of the correlation are indicated using colors from dark red to white with corresponding values given in the legend (lighter color corresponds to stronger correlation).

In this simple application to real-world data, OLS estimates are clearly too unstable to allow reliable interpretation of the results, even when reducing the set of predictors by a factor of 3. On the other hand, the α_{CV} -regularized kernel estimate appears somewhat undersmoothed. That is, in this application, the cross-validation method delivers too little smoothing with respect to the optimal smoothing for out-of-sample prediction. In comparison, the α_{BP}-regularized kernel estimate shown in Fig. 11 is smoother, and allows to uncover interesting insights about the relationship under study.

The kernel estimate suggests that the contemporaneous correlation between temperature and electricity consumption (i.e. coefficients along the indicative diagonal) is the largest from 12 pm to 4 pm (light yellow area), which almost coincides with the Ontario Energy Board's definition of summer peak period (11 am–5 pm). The relation is found of smaller magnitudes for periods 9 am–11 am and 4 pm–10 pm (yellow area), and even more so for 10 pm–9 am (orange area).

The estimated correlations with past and future temperatures (i.e. off-diagonal coefficients) beyond a 24-hour window are found relatively small (in darker areas). Therefore, the dynamic relationship between the variables is mainly characterized by the estimates in the lighter areas spanning around the indicative diagonal in Fig. 11. For instance, the outside air temperature at noon positively correlates with electricity consumption from 9 am to 9 pm that day. This possibly indicates the presence of both lag and anticipatory effects of outdoor temperature.

We observe that outdoor temperature has lasting effects on electricity consumption. A plausible explanation relates to the law of motion of indoor air temperature, which mainly depends on buildings insulation characteristics, the outdoor air temperature trajectory and indoor heating and cooling use. In the absence of sufficient air conditioning along the day, the indoor temperature level will eventually converge to the outdoor level. Since most people do not consume much air conditioning at home while being away, they eventually have to increase their consumption later on during the day in order to get back to their preferred level. Hence an increase in temperature at 9 am may positively affect consumption at 12 pm through its dynamic effect on indoor temperature.

We also find correlation with future temperature values, which are more difficult to interpret. For instance, consumption at 9 pm positively correlates with nighttime and next day temperature levels until 12 pm (yellow, orange and red areas). These estimates suggest that end-users may increase evening air conditioning in anticipation of higher nighttime temperature levels. A plausible underlying motivation may be to accumulate coolness so as to ensure a desired indoor temperature level at night and in the morning.

If agents were real-time optimizers, like smart thermostats, with rational expectations based on weather forecasts, the presence of partial dynamic adjustments of indoor temperature would create both lag effects and forward-looking dependence. The reality may however be different and one should keep in mind

[19] The condition number is the ratio of the largest eigenvalue on the smallest eigenvalue.

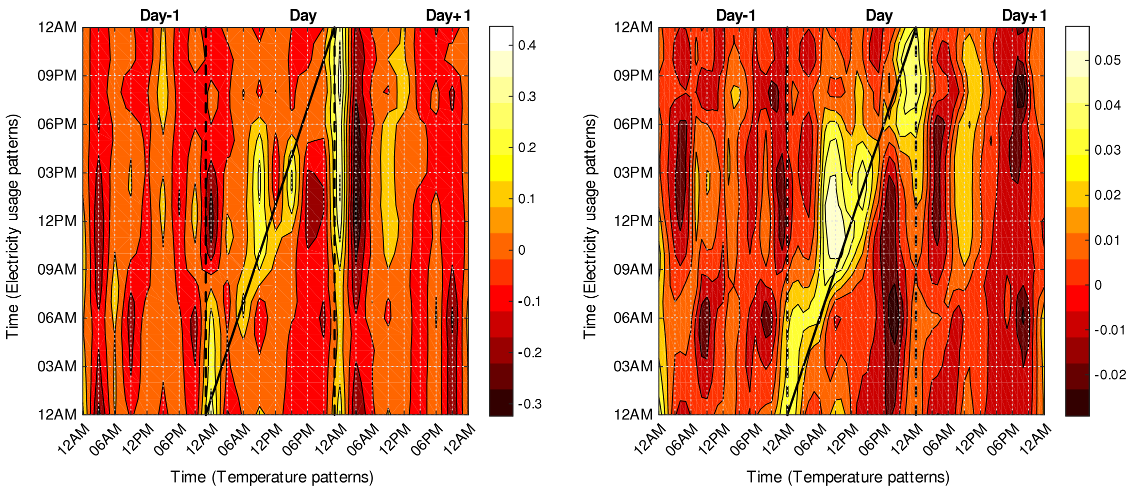

Fig. 10. 3 hourly OLS kernel estimate (left) and α_{CV}-regularized kernel estimate (right).

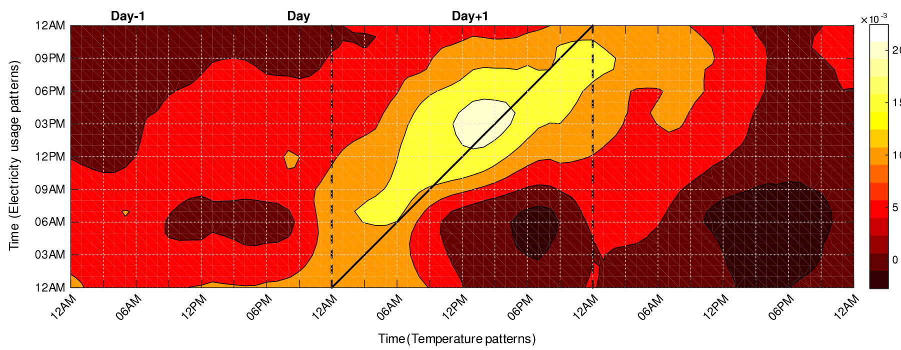

Fig. 11. α_{BP}-regularized kernel estimate.

that these estimates are subject to a regularization bias. Also, we acknowledge that these results may be subject to attenuation bias due to the ad-hoc construction of our temperature variable. Finally, the model considers aggregate measures of temperature and electricity consumption over the province of Ontario, the estimated parameters may therefore capture spatial correlation along with the temporal effects.

In order to assess the predictive power of the model, we plot $|\hat{u}_i(s)|/Y_i(s)$, the ratio of the residuals (in absolute terms) to the dependent variable, and the out-of-sample predictions using α_{BP} against actual trajectories for a workweek (Monday 16/6 to Friday 20/6) in summer 2014 in Fig. 12. We find that the model provides a reasonable prediction (4% mean absolute prediction error).

9.5. Conclusion of the application

In conclusion, the estimation results indicate positive correlations between electricity consumption and past, current and future temperature levels. The correlation with future values may be due to anticipatory behavior of agents, who look at the temperature forecast before deciding on whether to leave the air conditioning on that day, for instance. It may also be due to a measurement error in the temperature variable, or even spatial correlations captured through the temporal dimension from the use of aggregate variables. In general, the estimated coefficients vary across hours of the day, and the effect of temperature appears to persist for about 12 h. The model allows to obtain a good out-of-sample prediction of electricity consumption in the summer of 2014 using data for summers of 2010 to 2013.

The growing deployment of smart-metering technologies in electricity systems creates a need for micro-econometric models able to capture the dynamic behaviors of end-users with respect to market fundamentals. Such models allow to take advantage of the large amount of data so as to provide new insights to practitioners and policymakers about the behavior of consumers. In particular, information with regards to the end-users' behavior with respect to changes in weather or prices is valuable to local distribution companies since it can contribute to improve their demand-side bidding strategies in day-ahead markets, as suggested by Patrick and Wolak (1997). Access to smart meters data being limited, we have proposed instead a model of aggregate electricity consumption as a function of past and future outside air temperature. Nonetheless, we expect the functional regression framework to provide a valuable avenue for taking advantage of the large data sets from individual households' smart meters.

10. Conclusion

In this paper, we considered two functional regression models with functional response. In the first model, the regressor was an exogenous function. In the second model, the regressor was an endogenous function and identification relied on instrumental variables. For both models, we have proposed an estimator which is simple to implement, depends on only one smoothing parameter, and is robust to situations where eigenvalues are multiple. Moreover, the indexes t and s could be vectors and our framework allows for multiple regressors.

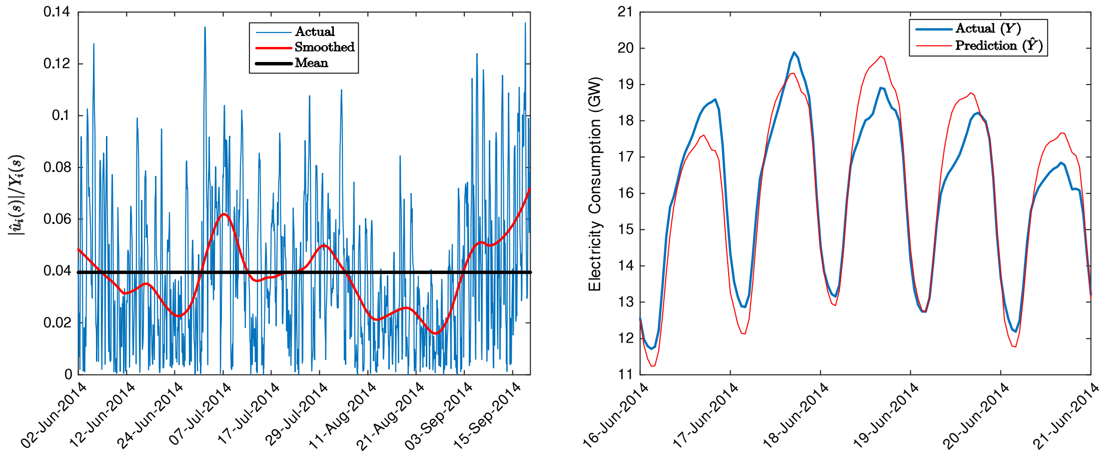

Fig. 12. $|\hat{u}_i(s)|/Y_i(s)$ (left) and out-of-sample prediction (right).

Various extensions would be interesting to investigate.

- We assumed that we observed a random sample, hence the data were independent. Many relevant applications involve observations which are cross-sectionally correlated.
- When the regressor Z is not observed, Z needs to be estimated first, which induces an error-in-variable. This case was considered in Park and Qian (2012).
- We discussed only the case where we penalized the HS norm of Π, the extension to other norms including norms of the derivatives of Π would be useful in the applications where the interest lies in the derivatives of Π.
- In our framework, we could relatively easily impose restrictions on Π (estimation under constraint).
- Instead of Tikhonov, one could consider the iterative regularization scheme called Landweber–Fridman (see Kress, 1999).
- The derivation of an oracle inequality for the choice of α is left for future research.

Appendix. Proofs

Proof of Proposition 1.

$$\left\|\hat{\Pi}_\alpha\right\|_{HS}^2 = \left\|\hat{C}_{YZ}\left(\alpha I + \hat{V}_Z\right)^{-1}\right\|_{HS}^2$$
$$\leq \left\|\hat{C}_{YZ}\right\|_{HS}^2 \left\|\left(\alpha I + \hat{V}_Z\right)^{-1}\right\|_{op}$$

using the fact that, if A is a HS operator and B is a bounded operator, $\|AB\|_{HS} \leq \|A\|_{HS}\|B\|_{op}$ where $\|B\|_{op} \equiv \sup_{\|\phi\|\leq 1}\|B\phi\|$ is the operator norm. Then, we have

$$\left\|\hat{\Pi}_\alpha\right\|_{HS}^2 \leq \frac{1}{\alpha}\left\|\hat{C}_{YZ}\right\|_{HS}^2.$$

It remains to show that $\hat{C}_{YZ}$ is a HS operator. $\hat{C}_{YZ}$ is an integral operator with degenerate kernel $\frac{1}{n}\sum_{i=1}^{n} y_i(s) z_i(t)$. A sufficient condition for $\hat{C}_{YZ}$ to be HS is that its kernel is square integrable which is true because Y_i and Z_i are elements of Hilbert spaces. The result of Proposition 1 follows.

Proof of Proposition 2. To prove Proposition 2, we need two preliminary lemmas.

Lemma 8. *Let $A = B + C$ where B is a zero mean random operator and C is a nonrandom operator. Then,*

$$E\left(\|A\|_{HS}^2\right) = E\left(\|B\|_{HS}^2\right) + \|C\|_{HS}^2.$$

Proof of Lemma 8.

$$\begin{aligned}
E\left(\|A\|_{HS}^2\right) &= E\left(\sum_j \langle A\phi_j, A\phi_j\rangle\right) \\
&= E\left(\sum_j \langle A^*A\phi_j, \phi_j\rangle\right) \\
&= E\left(\sum_j \langle (B+C)^*(B+C)\phi_j, \phi_j\rangle\right) \\
&= E\left(\sum_j \langle B^*B\phi_j, \phi_j\rangle\right) \\
&\quad + E\left(\sum_j \langle C^*B\phi_j, \phi_j\rangle\right) \\
&\quad + E\left(\sum_j \langle B^*C\phi_j, \phi_j\rangle\right) \\
&\quad + E\left(\sum_j \langle C^*C\phi_j, \phi_j\rangle\right).
\end{aligned}$$

The second and third terms on the r.h.s are equal to zero because $E(B) = 0$ and C is deterministic. We obtain $E\left(\|A\|_{HS}^2\right) = E\left(\|B\|_{HS}^2\right) + \|C\|_{HS}^2$.

Lemma 9. *Let A be a random Hilbert–Schmidt operator from $\mathcal{E}$ to $\mathcal{F}$.*

$$E\left(\|A\|_{HS}^2\right) = trE\left(A^*A\right) = trE\left(AA^*\right).$$

Proof of Lemma 9. We have

$$\begin{aligned}
E\left(\|A\|_{HS}^2\right) &= E\left(tr(A^*)A\right) \\
&= E\left(tr(AA^*)\right) \\
&= E\sum_j \langle AA^*\phi_j, \phi_j\rangle \\
&= trE\left(AA^*\right)
\end{aligned}$$

where the second equality follows from the fact that $tr(AB) = tr(BA)$ when A and B are Hilbert–Schmidt operators (see Pedersen, 1989).

Lemma 10. $\left\|\alpha(\alpha I+V_Z)^{-1}V_Z^{\beta/2}R\right\|_{HS}^2=O\left(\alpha^{\beta\wedge 2}\right)$.

Proof of Lemma 10. Let $\left\{\lambda_j,\varphi_j\right\}$ be the eigenvalues and orthonormal eigenfunctions of V_Z.

$$\begin{aligned}&\left\|\alpha(\alpha I+V_Z)^{-1}V_Z^{\beta/2}R\right\|_{HS}^2\\&=\alpha^2\sum_j\left\langle(\alpha I+V_Z)^{-1}V_Z^{\beta/2}R\varphi_j,(\alpha I+V_Z)^{-1}V_Z^{\beta/2}R\varphi_j\right\rangle\\&=\alpha^2\sum_j\frac{\lambda_j^\beta}{(\lambda_j+\alpha)^2}\left\langle R\varphi_j,R\varphi_j\right\rangle^2\\&\le\alpha^2\sup_\lambda\frac{\lambda^\beta}{(\lambda+\alpha)^2}\sum_j\left\langle R\varphi_j,R\varphi_j\right\rangle^2\\&=O\left(\alpha^{\beta\wedge 2}\right).\end{aligned}$$

The last equality follows from the fact that $\sum_j\left\langle R\varphi_j,R\varphi_j\right\rangle^2=\|R\|_{HS}^2<\infty$ and, using the notation $\lambda=\mu^2$, we have

$$\sup_\lambda\frac{\alpha^2\lambda^\beta}{(\lambda+\alpha)^2}=\sup_\mu\frac{\alpha^2\mu^{2\beta}}{\left(\mu^2+\alpha\right)^2}=O\left(\alpha^{\beta\wedge 2}\right)$$

by Carrasco et al., (2007, Proposition 3.11). Consequently,

$$\left\|\alpha(\alpha I+V_Z)^{-1}V_Z^{\beta/2}R\right\|_{HS}^2=O\left(\alpha^{\beta\wedge 2}\right).$$

This ends the proof of Lemma 10.

We turn to the proof of Proposition 2.

Replacing y_i by Πz_i+u_i in the expression of $\hat{C}_{ZY}$, we obtain

$$\begin{aligned}\hat{C}_{ZY}&=\frac{1}{n}\sum_i z_i\langle y_i,.\rangle\\&=\frac{1}{n}\sum_i z_i\langle u_i,.\rangle+\frac{1}{n}\sum_i z_i\langle\Pi z_i,.\rangle\\&=\hat{C}_{ZU}+\hat{V}_Z\Pi^*.\end{aligned}$$

We decompose $\hat{\Pi}_\alpha^*-\Pi^*$ in the following manner:

$$\begin{aligned}\hat{\Pi}_\alpha^*-\Pi^*&=\left(\alpha I+\hat{V}_Z\right)^{-1}\hat{C}_{ZY}-\Pi^*\\&=\left(\alpha I+\hat{V}_Z\right)^{-1}\hat{C}_{ZU}&(41)\\&\quad+\left(\alpha I+\hat{V}_Z\right)^{-1}\hat{V}_Z\Pi^*-(\alpha I+V_Z)^{-1}V_Z\Pi^*&(42)\\&\quad+(\alpha I+V_Z)^{-1}V_Z\Pi^*-\Pi^*.&(43)\end{aligned}$$

To study the rate of convergence of the MSE, we will study the rates of the three terms (41)–(43).

Applying Lemma 8 on the decomposition (41)–(43), we have

$$\begin{aligned}&E\left(\left\|\hat{\Pi}_\alpha-\Pi\right\|_{HS}^2|Z_1,Z_2,\ldots,Z_n\right)\\&=E\left(\|(41)\|_{HS}^2|Z_1,Z_2,\ldots,Z_n\right)+\|(42)+(43)\|_{HS}^2\\&\le E\left(\|(41)\|_{HS}^2|Z_1,Z_2,\ldots,Z_n\right)+2\|(42)\|_{HS}^2+2\|(43)\|_{HS}^2.\end{aligned}$$

We study the first term of the r.h.s. By Lemma 9,

$$\begin{aligned}&E\left(\|(41)\|_{HS}^2|Z_1,Z_2,\ldots,Z_n\right)\\&=E\left(\left\|\left(\alpha I+\hat{V}_Z\right)^{-1}\hat{C}_{ZU}\right\|_{HS}^2|Z_1,Z_2,\ldots,Z_n\right)\\&=trE\left(\left(\alpha I+\hat{V}_Z\right)^{-1}\hat{C}_{ZU}\hat{C}_{ZU}^*\left(\alpha I+\hat{V}_Z\right)^{-1}|Z_1,Z_2,\ldots,Z_n\right)\\&=tr\left\{\left(\alpha I+\hat{V}_Z\right)^{-1}E\left(\hat{C}_{ZU}\hat{C}_{ZU}^*|Z_1,Z_2,\ldots,Z_n\right)\left(\alpha I+\hat{V}_Z\right)^{-1}\right\}.\end{aligned}$$

Note that

$$\hat{C}_{ZU}\hat{C}_{ZU}^*\varphi=\frac{1}{n^2}\sum_{i,j}z_i\left\langle z_j,\varphi\right\rangle\left\langle u_i,u_j\right\rangle,$$

$$\begin{aligned}E\left(\hat{C}_{ZU}\hat{C}_{ZU}^*\varphi|Z_1,Z_2,\ldots,Z_n\right)&=\frac{1}{n}\sum_i z_i\langle z_i,\varphi\rangle E\\&\quad\times\left[\langle u_i,u_i\rangle|Z_1,Z_2,\ldots,Z_n\right]\\&=\frac{1}{n}\sum_i z_i\langle z_i,\varphi\rangle\,tr\,(V_U)\\&=\frac{1}{n}tr\,(V_U)\,\hat{V}_Z\varphi\end{aligned}$$

because the u_i are uncorrelated. To see that $E\left[\langle u,u\rangle\right]=trV_U$, decompose u on the basis formed by the eigenfunctions ψ_j of V_U so that $u=\sum_j\left\langle u,\psi_j\right\rangle\psi_j$. It follows that $\langle u,u\rangle=\sum_j\left\langle u,\psi_j\right\rangle^2$ and $E\langle u,u\rangle=\sum_j\left\langle V_U\psi_j,\psi_j\right\rangle=tr\,(V_U)$. Hence,

$$\begin{aligned}&E\left(\|(41)\|_{HS}^2|Z_1,Z_2,\ldots,Z_n\right)\\&\quad=\frac{1}{n}tr\,(V_U)\,tr\left(\left(\alpha I+\hat{V}_Z\right)^{-1}\hat{V}_Z\left(\alpha I+\hat{V}_Z\right)^{-1}\right)\\&\quad\le\frac{C}{n\alpha}\end{aligned}$$

where C is a generic constant. It follows that $E\left(\|(41)\|_{HS}^2\right)\le\frac{C}{n\alpha}$.

Now, we turn toward the term (42). We have

$$\begin{aligned}&\left(\alpha I+\hat{V}_Z\right)^{-1}\hat{V}_Z\Pi^*-(\alpha I+V_Z)^{-1}V_Z\Pi^*\\&=\left[-\left(I-\left(\alpha I+\hat{V}_Z\right)^{-1}\hat{V}_Z\right)+\left(I-(\alpha I+V_Z)^{-1}V_Z\right)\right]\Pi^*.\end{aligned}$$

Using $I=\left(\alpha I+\hat{V}_Z\right)^{-1}\left(\alpha I+\hat{V}_Z\right)$, we obtain

$$I-\left(\alpha I+\hat{V}_Z\right)^{-1}\hat{V}_Z=\alpha\left(\alpha I+\hat{V}_Z\right)^{-1}.$$

Hence,

$$\begin{aligned}(42)&=\left[-\alpha\left(\alpha I+\hat{V}_Z\right)^{-1}+\alpha(\alpha I+V_Z)^{-1}\right]\Pi^*\\&=-\left(\alpha I+\hat{V}_Z\right)^{-1}\left(V_Z-\hat{V}_Z\right)\alpha(\alpha I+V_Z)^{-1}\Pi^*\end{aligned}$$

where the last equality follows from $A^{-1}-B^{-1}=A^{-1}(B-A)B^{-1}$.

Now, we have

$$\begin{aligned}&\left\|\left(\alpha I+\hat{V}_Z\right)^{-1}\left(V_Z-\hat{V}_Z\right)\alpha(\alpha I+V_Z)^{-1}\Pi^*\right\|_{HS}^2\\&\le\left\|\left(\alpha I+\hat{V}_Z\right)^{-1}\right\|_{op}^2\left\|\left(V_Z-\hat{V}_Z\right)\right\|_{op}^2\left\|\alpha(\alpha I+V_Z)^{-1}\Pi^*\right\|_{HS}^2\end{aligned}$$

where $\left\|\left(\alpha I+\hat{V}_Z\right)^{-1}\right\|_{op}^2\le 1/\alpha^2$, $\left\|\left(V_Z-\hat{V}_Z\right)\right\|_{op}^2=O_p(1/n)$ by Assumption 2 and $\left\|\alpha(\alpha I+V_Z)^{-1}\Pi^*\right\|_{HS}^2=O\left(\alpha^{\beta\wedge 2}\right)$ by Lemma 10.

If $\beta > 1$ then the term corresponding to (42) is negligible with respect to (41). If $\beta < 1$, then (41) is negligible with respect to (42).

Now, we turn our attention toward the term (43). We have

$$\begin{aligned}
&(\alpha I + V_Z)^{-1} V_Z \Pi^* - \Pi^* \\
&= (\alpha I + V_Z)^{-1} (V_Z - \alpha I - V_Z) \Pi^* \\
&= \alpha(\alpha I + V_Z)^{-1} \Pi^* \\
&= \alpha(\alpha I + V_Z)^{-1} V_Z^{\beta/2} R
\end{aligned}$$

by Assumption 4. The rate of this term follows from Lemma 10.

This concludes the proof of Proposition 2.

Proof of Proposition 3. We decompose $\hat{\Pi}^*_\alpha - \Pi^*_{N\perp}$ in the following manner:

$$\begin{aligned}
\hat{\Pi}^*_\alpha - \Pi^*_{N\perp} &= \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZY} - \Pi^*_{N\perp} \\
&= \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZU} && (44) \\
&\quad + \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{V}_Z \Pi^* - (\alpha I + V_Z)^{-1} V_Z \Pi^* && (45) \\
&\quad + (\alpha I + V_Z)^{-1} V_Z \Pi^*_{N\perp} - \Pi^*_{N\perp} && (46)
\end{aligned}$$

where (46) comes from the fact that $V_Z \Pi^* = V_Z \Pi^*_{N\perp}$. Following the proof of Proposition 2, we can establish that the rates of (44) and (46) are the same as those of (41) and (43). The rate of (45) will be however different from that of (42). The reason is that whereas $V_Z \Pi^* = V_Z \Pi^*_{N\perp}$, $\hat{V}_Z \Pi^* \neq \hat{V}_Z \Pi^*_{N\perp}$. Using the steps of the proof of Proposition 2, we have

$$\begin{aligned}
&\|(45)\|^2_{HS} \\
&= \left\| \left(\alpha I + \hat{V}_Z\right)^{-1} \left(V_Z - \hat{V}_Z\right) \alpha (\alpha I + V_Z)^{-1} \Pi^* \right\|^2_{HS} \\
&\le \left\| \left(\alpha I + \hat{V}_Z\right)^{-1} \right\|^2_{op} \left\| \left(V_Z - \hat{V}_Z\right) \right\|^2_{op} \left\| \alpha (\alpha I + V_Z)^{-1} \Pi^* \right\|^2_{HS}
\end{aligned}$$

where $\left\| \left(\alpha I + \hat{V}_Z\right)^{-1} \right\|^2_{op} \le 1/\alpha^2$, $\left\| \left(V_Z - \hat{V}_Z\right) \right\|^2_{op} = O_p(1/n)$ by Assumption 2 and $\left\| \alpha(\alpha I + V_Z)^{-1} \Pi^* \right\|^2_{HS} = O(1)$. We do not get the rate $\left\| \alpha(\alpha I + V_Z)^{-1} \Pi^* \right\|^2_{HS} = O\left(\alpha^{\beta \wedge 2}\right)$ because Lemma 10 does not apply here (it applies with Π replaced by $\Pi_{N\perp}$). Hence the rate of $\|(45)\|^2_{HS}$ is $O_p\left(1/(n\alpha^2)\right)$.

Proof of Proposition 5. Under the assumptions (U_i, Z_i) i.i.d., $E\|Z_i\|^4 < \infty$, and $E\|U_i\|^2 \|Z_i\|^2 < \infty$, Theorem 4 ensures the root-n asymptotic normality of $\frac{1}{n} \sum_i \begin{pmatrix} U_i \otimes Z_i \\ Z_i \otimes Z_i \end{pmatrix}$. By the continuous mapping theorem, one has (23) using the continuous transformation $\begin{pmatrix} A \\ B \end{pmatrix} \mapsto (\alpha I + V_Z)^{-1} A + \alpha(\alpha I + V_Z)^{-1} B (\alpha I + V_Z)^{-1} \Pi^\star$. The covariance operator of $\sqrt{n}(\hat{\Pi}^\star_\alpha - \Pi^\star_\alpha)$ may be written using (22) as

$$\begin{aligned}
\Omega_\alpha &= E\Bigg[\left(\frac{1}{\sqrt{n}} \sum_i \left((u_i + \alpha \Pi \tilde{z}_i) \otimes \tilde{z}_i - \alpha C_{\tilde{z}\Pi\tilde{z}} \right) \right) \\
&\quad \tilde{\otimes} \left(\frac{1}{\sqrt{n}} \sum_i \left((u_i + \alpha \Pi \tilde{z}_i) \otimes \tilde{z}_i - \alpha C_{\tilde{z}\Pi\tilde{z}} \right) \right) \Bigg] \\
&= E\Bigg[\left((U + \alpha \Pi \tilde{Z}) \otimes \tilde{Z} - \alpha C_{\tilde{Z}\Pi\tilde{Z}} \right) \\
&\quad \tilde{\otimes} \left((U + \alpha \Pi \tilde{Z}) \otimes \tilde{Z} - \alpha C_{\tilde{Z}\Pi\tilde{Z}} \right) \Bigg],
\end{aligned}$$

where the second line is obtained from the i.i.d. assumption. Straightforward developments yield (24). Now letting $\alpha \to 0$ gives (25).

Proof of Proposition 6. We have

$\hat{\Pi}^{m*}_\alpha - \Pi^* = \hat{\Pi}^{m*}_\alpha - \hat{\Pi}^*_\alpha + \hat{\Pi}^*_\alpha - \Pi^*$.

We focus on the term $\hat{\Pi}^{m*}_\alpha - \hat{\Pi}^*_\alpha$.

$$\begin{aligned}
\hat{\Pi}^{m*}_\alpha - \hat{\Pi}^*_\alpha &= \left(\alpha I + \hat{V}^m_Z\right)^{-1} \hat{C}^m_{ZY} - \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZY} \\
&= \left(\alpha I + \hat{V}^m_Z\right)^{-1} \left(\hat{C}^m_{ZY} - \hat{C}_{ZY}\right) \\
&\quad + \left[\left(\alpha I + \hat{V}^m_Z\right)^{-1} - \left(\alpha I + \hat{V}_Z\right)^{-1} \right] \hat{C}_{ZY}.
\end{aligned}$$

$$\begin{aligned}
\left\| \hat{C}^m_{ZY} - \hat{C}_{ZY} \right\|^2_{HS} &= \left\| \frac{1}{n} \sum_{i=1}^n z^m_i \langle y^m_i, . \rangle - \frac{1}{n} \sum_{i=1}^n z_i \langle y_i, . \rangle \right\|^2_{HS} \\
&= \left\| \frac{1}{n} \sum_{i=1}^n \left\{ \left(z^m_i - z_i\right) \langle y_i, . \rangle + z^m_i \langle y^m_i - y_i, . \rangle \right\} \right\|^2_{HS} \\
&\le \frac{2}{n^2} \sum_{i=1}^n \Big\{ \left\| \left(z^m_i - z_i\right) \langle y_i, . \rangle \right\|^2_{HS} \\
&\quad + \left\| z^m_i \langle y^m_i - y_i, . \rangle \right\|^2_{HS} \Big\}.
\end{aligned}$$

$$\begin{aligned}
\left\| \left(z^m_i - z_i\right) \langle y_i, . \rangle \right\|^2_{HS} &= \sum_j \left\langle \left(z^m_i - z_i\right) \langle y_i, \phi_j \rangle, \left(z^m_i - z_i\right) \langle y_i, \phi_j \rangle \right\rangle \\
&= \left\| z^m_i - z_i \right\|^2 \sum_j \langle y_i, \phi_j \rangle^2 \\
&= O_p\left(f(m)^2\right).
\end{aligned}$$

$$\begin{aligned}
\left\| z^m_i \langle y^m_i - y_i, . \rangle \right\|^2_{HS} &= \sum_j \left\langle z^m_i \langle y^m_i - y_i, \phi_j \rangle, z^m_i \langle y^m_i - y_i, \phi_j \rangle \right\rangle \\
&= \left\| z^m_i \right\|^2 \sum_j \langle y^m_i - y_i, \phi_j \rangle^2 \\
&= \left\| z^m_i \right\|^2 \left\| y^m_i - y_i \right\|^2 \\
&= O_p\left(f(m)^2\right).
\end{aligned}$$

Hence,

$$\left\| \left(\alpha I + \hat{V}^m_Z\right)^{-1} \left(\hat{C}^m_{ZY} - \hat{C}_{ZY}\right) \right\|^2_{HS} = O_p\left(\frac{f(m)^2}{\alpha^2 n^2} \right).$$

$$\begin{aligned}
&\left[\left(\alpha I + \hat{V}^m_Z\right)^{-1} - \left(\alpha I + \hat{V}_Z\right)^{-1} \right] \hat{C}_{ZY} \\
&= \left(\alpha I + \hat{V}^m_Z\right)^{-1} \left(\hat{V}_Z - \hat{V}^m_Z\right) \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZY} \\
&= \left(\alpha I + \hat{V}^m_Z\right)^{-1} \left(\hat{V}_Z - \hat{V}^m_Z\right) \hat{\Pi}^*_\alpha.
\end{aligned}$$

Hence,

$$\left\| \left(\alpha I + \hat{V}^m_Z\right)^{-1} \left(\hat{V}_Z - \hat{V}^m_Z\right) \hat{\Pi}^*_\alpha \right\|^2_{HS} = O_p\left(\frac{f(m)^2}{\alpha^2 n^2} \right).$$

This concludes the proof of Proposition 6.

Proof of Proposition 7. Using the fact that $\|a + b + c\|^2_{HS} \le 3\left(\|a\|^2_{HS} + \|b\|^2_{HS} + \|c\|^2_{HS}\right)$, we can evaluate the terms (35)–(37) separately. The proof follows closely that of Proposition 2. Let $\mathbf{Z}$ and $\mathbf{W}$ be the sets $(Z_1, Z_2, \dots, Z_n)$ and $(W_1, W_2, \dots, W_n)$.

$$\begin{aligned}
&E\left(\|(35)\|^2_{HS} | \mathbf{Z}, \mathbf{W} \right) \\
&= tr\Bigg\{ \left(\alpha I + \hat{C}_{ZW} \hat{C}_{WZ}\right)^{-1} \hat{C}_{ZW} E\left(\hat{C}_{WU} \hat{C}_{UW} | \mathbf{Z}, \mathbf{W} \right) \hat{C}_{WZ} (\alpha I \\
&\quad + \hat{C}_{ZW} \hat{C}_{WZ} \Big)^{-1} \Bigg\}.
\end{aligned}$$

Using

$$E\left(\hat{C}_{WU}\hat{C}_{UW}|\mathbf{Z},\mathbf{W}\right)=\frac{1}{n}tr\left(V_U\right)\hat{V}_W,$$

we obtain

$$E\left(\|(35)\|_{HS}^2\right)\leq\frac{C}{n\alpha}$$

for some constant C.

The proof regarding the rates of convergence of (36) and (37) is similar to that of Proposition 2 and is not repeated here.

References

Ait-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. J. Finance LIII, 499–547.
Ait-Sahalia, Y., Wang, Y., Yared, F., 2001. Do option markets correctly price the probabilities of movement of the underlying asset. J. Econometrics 102, 67–110.
Andersson, J., Lillestol, J., 2010. Modeling and forecasting electricity consumption by functional data analysis. J. Energy Mark. 3, 3–15.
Antoch, J., Prchal, L., De Rosa, M., Sarda, P., 2010. Electricity consumption prediction with functional linear regression using spline estimators. J. Appl. Stat. 37, 2027–2041.
Aue, A., Norinho, D., Hormann, S., 2015. On the prediction of stationary functional time series. J. Amer. Statist. Assoc. 110 (509), 378–392.
Bondarenko, O., 2003. Estimation of risk-neutral densities using positive convolution approximation. J. Econometrics 116, 85–112.
Bosq, D., 2000. Linear Processes in Function Spaces: Theory and Applications. Springer-Verlag, New York.
Cardot, H., Ferraty, F., Sarda, P., 2003. Spline estimators for the functional linear model. Statist. Sinica 13, 571–591.
Carrasco, M., Florens, J.P., Renault, E., 2007. In: Heckman, J.J., Leamer, E.E. (Eds.), Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization. In: Handbook of Econometrics, vol. 6B.
Centorrino, S., 2016. Data-Driven Selection of the Regularization Parameter in Additive Nonparametric Instrumental Regressions. Mimeo, Stony Brook University.
Centorrino, S., Fève, F., Florens, J.P., 2017. Additive nonparametric instrumental regressions: A guide to implementation. J. Econ. Methods 6 (1).
Centorrino, S., Florens, J.P., 2015. Nonparametric Instrumental Variable Estimation of Binary Response Models. Mimeo, Stony Brook University.
Comte, F., Johannes, J., 2012. Adaptive functional linear regression. Ann. Statist. 40 (6), 2765–2797.
Conley, T., Hansen, E., Luttmer, J., Scheinkman, L., 1997. Short-term interest rates as subordinated diffusions. Rev. Financ. Stud. 10 (3).
Crambes, C., Kneip, A., Sarda, P., 2009. Smoothing spline estimators for functional linear regression. Ann. Statist. 37, 35–72.
Crambes, C., Mas, A., 2013. Asymptotics of prediction in functional linear regression with functional outputs. Bernoulli 19, 2627–2651.
Cuevas, A., Febrero, M., Fraiman, R., 2002. Linear functional regression: the case of fixed design and functional response. Canad. J. Statist. 30 (2), 285–300.
Darolles, S., Fan, Y., Florens, J.P., Renault, E., 2011. Nonparametric instrumental regression. Econometrica 79, 1541–1565.
Dauxois, J., Pousse, A., Romain, Y., 1982. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. J. Multivariate Anal. 12, 136–154.
Engl, H., Hanke, M., Neubauer, A., 2000. Regularization of Inverse Problems. Kluwer Academic Publishers, Dordrecht.
Faruqui, A., Sergici, N., Lessem, F., Denton, B., Spencer, S., King, C., 2013. Report prepared for Ontario Power Authority, Impact Evaluation of Ontario's Time-of-Use Rates: First Year Analysis.
Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations. Springer-Verlag, London.
Florens, J.-P., Johannes, J., Van Bellegem, S., 2011. Identification and estimation by penalization in nonparametric instrumental regression. Econometric Theory 27, 472–496.
Florens, J.-P., Van Bellegem, S., 2015. Instrumental variable estimation in functional linear models. J. Econometrics 186, 465–476.
Garcia, R., Gençay, R., 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. J. Econometrics 94, 93–115.
Garcia, R., Ghysels, E., Renault, E., 2010. In: Hansen, L.P., Aït-Sahalia, Y. (Eds.), The Econometrics of Option Pricing. In: Handbook of Financial Econometrics, Elseviers, North-Holland.
Goldenshluger, A., Lepski, O., 2011. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. Ann. Statist. 39, 1608–1632.
Gouriéroux, C., Jasiak, J., 2001. Financial Econometrics. Princeton University Press, Princeton.
Hall, P., Horowitz, J., 2007. Methodology and convergence rates for functional linear regression. Ann. Statist. 35, 70–91.
Hansen, L.P., Sargent, T., 1980. Formulating and estimating dynamic linear rational expectations models. J. Econom. Dynam. Control 2 (1), 7–46.
Horowitz, J., Lee, S., 2007. Nonparametric instrumental variables estimation of a quantile regression model. Econometrica 75 (4), 1191–1208.
Horvath, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. Springer, New-York.
Jackwerth, J., Rubinstein, M., 1996. Recovering probability distributions from option prices. J. Finance 51, 1611–1631.
Kargin, V., Onatski, A., 2008. Curve forecasting by functional autoregression. J. Multivariate Anal. 99, 2508–2526.
Kennan, J., 1979. The estimation of partial adjustment models with rational expectations. Econometrica 47 (6), 1441–1455.
Kress, R., 1999. Linear Integral Equations. Springer.
Liebl, D., 2013. Modeling and forecasting electricity spot prices: A functional data perspective. Ann. Appl. Stat. 7 (3), 1562–1592.
Manresa, E., 2015. Estimating the Structure of Social Interactions using Panel Data. mimeo, MIT Sloan.
McLaughlin, L., Yu, L., Jia, M., Murphy-Hoye, E., Piccioli, A., Pratt, Z., Tong, L., 2011. Thermal Dynamic for Home Energy Management: A Case Study. Tech. Rep. ACSP-TR-10-11-01. Cornell Univ., Ithaca, NY, USA.
Muth, J., 1961. Rational expectations and the theory of price movements. Econometrica 29, 315–335.
Park, J., Qian, J., 2012. Functional regression of continuous state distribution. J. Econometrics 167, 397–412.
Patrick, R., Wolak, F., 1997. Estimating Customer-Level Demand for Electricity under Real Time Pricing. Mimeo, Stanford University.
Pedersen, G., 1989. Analysis Now. Springer-Verlag, New York.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed.. Springer, New York.
Rudin, W., 1991. Functional Analysis, second ed.. McGraw Hill.
Vilenkin, N.I.A., 1968. Special Functions and the Theory of Group Representations. American Mathematical Society.
Yao, F., Müller, H.G., Wang, J.L., 2005. Functional linear regression analysis for longitudinal data. Ann. Statist. 33 (6), 2873–2903.
Yu, Z., McLaughlin, L., Jia, L., Murphy-Hoye, M.C., Pratt, A., Tong, L., 2013. Modeling and stochastic control for home energy management. IEEE Trans. Smart Grid 4 (4), 2244–2255.