

Functional linear regression with functional response*

David Benatia

Marine Carrasco

Université de Montréal

Université de Montréal

Jean-Pierre Florens

Toulouse School of Economics

February 2015

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 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 standard 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 . We derive the rate of convergence of the mean-square error, the asymptotic distribution of the estimator, and develop tests on Π . Often, the full trajectories are not observed but only a discretized version is available. We address this issue in the scenario where the data become more and more frequent (in-fill asymptotics). We also consider the case where Z is endogenous and instrumental variables are used to estimate Π .

Key Words: Functional regression, instrumental variables, linear operator, Tikhonov regularization

*The authors thank the participants of the 6th French Econometrics Conférence and especially their discussant Jan Johannes for helpful comments.

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 standard 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 (in-fill asymptotics).

We also consider the case where Z is endogenous and instrumental variables are used to estimate Π .

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 Cardo, Ferraty, and Sarda (2003), Hall and Horowitz (2007), Horowitz and Lee (2007), Darolles, Fan, Florens and Renault (2011), and Crambes, Kneib, and Sarda (2009). 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) and Ferraty and Vieu (2006). Cuevas, Febrero, and Fraiman (2002) consider a fixed design setting and propose an estimator of Π based on interpolation. Yao, Müller, and Wang (2005) consider the case where both predictor and response trajectories are observed at discrete and irregularly spaced times. Their estimator is based on spectral cut-off regularized inverse using nonparametric estimators of the principal components. Crambes and Mas (2013) consider again 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. Antoch, Prchal, Rosa, and Sarda (2010) use a functional linear regression with functional response 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.

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. The proofs are collected in Appendix.

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) $i = 1, 2, \dots, n$ 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 Hilbert space \mathcal{E} equipped with an inner product \langle, \rangle and Z belongs to a 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 operator from \mathcal{F} to \mathcal{E} . Moreover, Z is exogenous so that $\text{cov}(Z, U) = 0$. This assumption will be relaxed in Section 7.

For illustration, consider the following example

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

where \mathcal{S} and \mathcal{T} are some intervals of \mathbb{R} . Then, Π can be represented as an integral operator such that

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

for any $\varphi \in \mathcal{F}$. π is referred to as the kernel of the operator Π . Model (1) means that $Y(t)$ depends not only on $Z(t)$ but also on all the $Z(s)$, for $s \neq t$. The object of interest is the estimation of the operator Π .

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

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

Hence, we have the following relationships:

$$C_{YZ} = \Pi V_Z, \tag{2}$$

$$C_{ZY} = V_Z \Pi^* \tag{3}$$

where Π^* is the adjoint of Π . 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 (3). 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^* can not be obtained directly by solving $\hat{C}_{ZY} = \hat{V}_Z \Pi^*$ 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 could be used but it would not be continuous. To stabilize the inverse, we need to use some regularization scheme. We adopt Tikhonov regularization (see Kress, 1999 and Carrasco, Florens, and Renault, 2007).

The estimator of Π^* is defined as

$$\hat{\Pi}_\alpha^* = \left(\alpha I + \hat{V}_Z \right)^{-1} \hat{C}_{ZY} \quad (4)$$

and that of Π is defined by

$$\hat{\Pi}_\alpha = \hat{C}_{YZ} \left(\alpha I + \hat{V}_Z \right)^{-1} \quad (5)$$

where α is some positive regularization parameter which will be allowed to converge to zero as n goes to infinity. The estimators (4) and (5) can be viewed as generalization of ordinary least-squares estimators. They also have an interpretation as the solution to an inverse problem.

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} is

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

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

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

where $\hat{r} = (y_1, \dots, y_n)'$ and K is the operator from \mathcal{H} to \mathcal{E}^n such that $K\Pi = (\Pi z_1, \dots, \Pi z_n)'$. 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 $f = (f_1, \dots, f_n)'$ and $g = (g_1, \dots, g_n)'$. Let us check that $\hat{\Pi}_\alpha$ is a classical Tikhonov regularized inverse of the operator K :

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

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

$$\langle K\Pi, f \rangle_{\mathcal{E}^n} = \langle \Pi, B \rangle_{\mathcal{H}}. \quad (6)$$

Note that

$$\begin{aligned}
\langle \Pi, B \rangle_{\mathcal{H}} &= \text{tr}(\Pi B^*) \\
&= \sum_j \langle \Pi B^* \varphi_j, \varphi_j \rangle \\
&= \sum_j \langle B^* \varphi_j, \Pi^* \varphi_j \rangle
\end{aligned}$$

where φ_j is a basis of \mathcal{E} . On the other hand,

$$\begin{aligned}
\langle K\Pi, f \rangle_{\mathcal{E}^n} &= \frac{1}{n} \sum_i \langle \Pi z_i, f_i \rangle_{\mathcal{E}} \\
&= \frac{1}{n} \sum_i \langle z_i, \Pi^* f_i \rangle_{\mathcal{F}}.
\end{aligned}$$

Using $f_i = \sum_j \langle f_i, \varphi_j \rangle \varphi_j$, we obtain

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

It follows from (6) that $B^* \varphi_j = \frac{1}{n} \sum_i \langle f_i, \varphi_j \rangle z_i$ for all j and hence

$$B^* \varphi = \frac{1}{n} \sum_i \langle f_i, \varphi \rangle z_i$$

for all φ in \mathcal{E} . Now, we look for B the adjoint of B^* . B is the solution of

$$\langle B^* \varphi_1, \varphi_2 \rangle_{\mathcal{F}} = \langle \varphi_1, B \varphi_2 \rangle_{\mathcal{E}}.$$

We have

$$\begin{aligned}
\langle B^* \varphi_1, \varphi_2 \rangle_{\mathcal{F}} &= \frac{1}{n} \sum_i \langle f_i, \varphi_1 \rangle \langle z_i, \varphi_2 \rangle_{\mathcal{F}} \\
&= \left\langle \varphi_1, \frac{1}{n} \sum_i \langle z_i, \varphi_2 \rangle f_i \right\rangle_{\mathcal{E}}.
\end{aligned}$$

Hence,

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

We have

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

and

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

It follows that

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

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

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

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

2.3 Identification

It is easier to study the identification from the viewpoint of Equation (3). 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 (3), 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

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

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

$$\begin{aligned}
\langle TH, K \rangle &= \text{tr}(THK^*) \\
&= \text{tr}(V_Z HK^*) \\
&= \text{tr}(HK^*V_Z)
\end{aligned}$$

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

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

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

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.

2.4 Computation of the estimator

To show how to compute $\hat{\Pi}_\alpha^*$ explicitly, we multiply the left and right of (4) by $(\alpha I + \hat{V}_Z)$ to obtain

$$\begin{aligned}\hat{C}_{ZY}\psi &= (\alpha I + \hat{V}_Z) \hat{\Pi}_\alpha^* \psi \Leftrightarrow \\ \frac{1}{n} \sum_{i=1}^n z_i \langle y_i, \psi \rangle &= \alpha \hat{\Pi}_\alpha^* \psi + \frac{1}{n} \sum_{i=1}^n z_i \langle z_i, \hat{\Pi}_\alpha^* \psi \rangle.\end{aligned}\quad (7)$$

Then, we take the inner product with z_l , $l = 1, 2, \dots, n$ on the left and right hand side of (7), to obtain n equations:

$$\frac{1}{n} \sum_{i=1}^n \langle z_l, z_i \rangle \langle y_i, \psi \rangle = \alpha \langle z_l, \hat{\Pi}_\alpha^* \psi \rangle + \frac{1}{n} \sum_{i=1}^n \langle z_l, z_i \rangle \langle z_i, \hat{\Pi}_\alpha^* \psi \rangle, \quad l = 1, 2, \dots, n, \quad (8)$$

with n unknowns $\langle z_i, \hat{\Pi}_\alpha^* \psi \rangle$, $i = 1, 2, \dots, n$. Let M be the $n \times n$ matrix with (l, i) element $\langle z_l, z_i \rangle / n$, v the n -vector of $\langle z_i, \hat{\Pi}_\alpha^* \psi \rangle$ and w the n -vector of $\langle y_i, \psi \rangle$. (8) 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:

$$\begin{aligned}\hat{\Pi}_\alpha^* \psi &= \frac{1}{\alpha n} \sum_{i=1}^n z_i \left(\langle y_i, \psi \rangle - \langle z_i, \hat{\Pi}_\alpha^* \psi \rangle \right) \\ &= \frac{1}{\alpha n} \underline{z}' (I - M(\alpha I + M)^{-1}) w \\ &= \frac{1}{n} \underline{z}' (\alpha I + M)^{-1} w\end{aligned}\quad (9)$$

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 (9), we obtain

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

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

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

Hence, to compute $\hat{\Pi}_\alpha \varphi$, we need to know $\hat{\Pi}_\alpha z_i$. From (5), 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 , $i = 1, 2, \dots, n$, we obtain

$$\begin{aligned}\alpha \hat{\Pi}_\alpha z_i + \hat{\Pi}_\alpha \hat{V}_Z z_i &= \hat{C}_{YZ} z_i \Leftrightarrow \\ \alpha \left(\hat{\Pi}_\alpha z_i \right) (t) + \frac{1}{n} \sum_{j=1}^n \left(\hat{\Pi}_\alpha z_j \right) (t) \langle z_j, z_i \rangle &= \frac{1}{n} \sum_{j=1}^n y_j (t) \langle z_j, z_i \rangle, \quad i = 1, 2, \dots, n. \quad (11)\end{aligned}$$

For each t , we can solve the n equations with n unknowns $\left(\hat{\Pi}_\alpha z_j \right) (t)$ given by (11) and deduct $\hat{\Pi}_\alpha \varphi$ from (10).

The prediction of Y_i is given by

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

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. U_i is a random process of \mathcal{E} such that $E(U_i) = 0$, $cov(U_i, U_j | Z_1, Z_2, \dots, Z_n) = 0$ for all $i \neq j$ and $= V_U$ for $i = j$ where V_U is a 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 $\left\| \hat{V}_Z - V_Z \right\|_{HS}^2 = O_p(1/n)$.

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

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 $E(\|U_i\|^2) < \infty$.

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

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

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

Proof: See Appendix.

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

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

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

$$\begin{aligned}\hat{\Pi}_\alpha^* - \Pi^* &= (\alpha I + \hat{V}_Z)^{-1} \hat{C}_{ZY} - \Pi^* \\ &= (\alpha I + \hat{V}_Z)^{-1} \hat{C}_{ZU}\end{aligned}\tag{12}$$

$$+ (\alpha I + \hat{V}_Z)^{-1} \hat{V}_Z \Pi^* - (\alpha I + V_Z)^{-1} V_Z \Pi^*\tag{13}$$

$$+ (\alpha I + V_Z)^{-1} V_Z \Pi^* - \Pi^*.\tag{14}$$

To study the rate of convergence of the MSE, we will study the rates of the three terms (12), (13), and (14).

Proposition 2 *Assume Assumptions 1 to 4 hold.*

If $\beta > 1$, then $MSE = O_p\left(\frac{1}{n\alpha} + \alpha^{\beta\wedge 2}\right)$.

If $\beta < 1$, then $MSE = O_p\left(\frac{\alpha^{\beta\wedge 2}}{n\alpha^2} + \alpha^{\beta\wedge 2}\right)$.

4 Asymptotic normality for fixed α and tests

Assumption 5. (U_i, Z_i) are iid and $E(U_i|Z_i) = 0$.

Under Assumption 5 and some extra moment conditions (see Dauxois, Pousse, and Romain (1982) and Mas (2006)), we have

$$\sqrt{n} (\hat{V}_Z - V_Z) \xrightarrow{d} \mathcal{N}(0, K_Z),$$

$$\sqrt{n} \hat{C}_{ZU} \xrightarrow{d} \mathcal{N}(0, K_{ZU})$$

where K_Z and K_{ZU} are covariance operators and the convergence is either in the space of Hilbert-space operators (Dauxois et al. 1982) or in the space of trace-class operators (Mas, 2006). Moreover, $\sqrt{n} (\hat{V}_Z - V_Z)$ and $\sqrt{n} \hat{C}_{ZU}$ are asymptotically independent.

In this section, we consider the case where α is fixed. In that case, $\hat{\Pi}_\alpha^*$ 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

$$\begin{aligned}
\hat{\Pi}_\alpha^* - \Pi_\alpha^* &= (\alpha I + \hat{V}_Z)^{-1} \hat{C}_{ZU} \\
&\quad + (\alpha I + \hat{V}_Z)^{-1} \hat{V}_Z \Pi^* - (\alpha I + V_Z)^{-1} V_Z \Pi^* \\
&= (\alpha I + V_Z)^{-1} \hat{C}_{ZU} \\
&\quad + \left[(\alpha I + \hat{V}_Z)^{-1} - (\alpha I + V_Z)^{-1} \right] \hat{C}_{ZU} \\
&\quad + \alpha (\alpha I + \hat{V}_Z)^{-1} (V_Z - \hat{V}_Z) (\alpha I + V_Z)^{-1} \Pi^* \\
&= (\alpha I + V_Z)^{-1} \hat{C}_{ZU} \\
&\quad + \alpha (\alpha I + V_Z)^{-1} (\hat{V}_Z - V_Z) (\alpha I + V_Z)^{-1} \Pi^*
\end{aligned} \tag{15}$$

$$\begin{aligned}
&\quad + O_p\left(\frac{1}{n}\right).
\end{aligned} \tag{16}$$

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 $O_p(1/\sqrt{n})$ and will affect the asymptotic distribution. This distribution is not simple. We are going to characterize it below.

From Equations (15) and (16), neglecting the $O_p(1/n)$ term, we have

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

where 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 $(Y \otimes X)K = K^*Y \otimes X$ for $X \in \mathcal{F}, X \in \mathcal{E}$ and $K \in \mathcal{H}$. The third line uses the definition of V_Z and the fourth introduces the notation $\tilde{Z} \equiv (\alpha I + V_Z)^{-1}Z$. 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, denoted \otimes , distributes over addition.

The covariance operator of $\hat{\Pi}_\alpha^* - \Pi_\alpha^*$ is an operator which maps the space of Hilbert-Schmidt operators from \mathcal{E} to \mathcal{F} , denoted \mathcal{H} , 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,¹ and may be used like in Dauxois, Pousse and Romain (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 $(\Pi_1, \Pi_2) \in \mathcal{H}^2$ is a mapping from \mathcal{H} into itself, hence $\Pi_1 \tilde{\otimes} \Pi_2$ is an element of the Hilbert space of Hilbert-Schmidt operators from \mathcal{H} to \mathcal{H} equipped with the Hilbert-Schmidt inner product. For $T = \varphi \otimes \psi \in \mathcal{H}$, $\Pi_1 = X \otimes Z \in \mathcal{H}$ and $\Pi_2 = Y \otimes W \in \mathcal{H}$, this tensor product is equivalently defined as :

$$\begin{aligned} \text{(i)} \quad & (\Pi_1 \tilde{\otimes} \Pi_2)T = \langle T, \Pi_1 \rangle_{\mathcal{H}} \Pi_2 \in \mathcal{H} \\ \text{(ii)} \quad & \left((X \otimes Z) \tilde{\otimes} (Y \otimes W) \right) (\varphi \otimes \psi) = \left((X \otimes Y) \varphi \right) \otimes \left((Z \otimes W) \psi \right), \\ & \forall \varphi, X, Y \in \mathcal{F}, \psi, Z, W \in \mathcal{E}, \end{aligned}$$

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

$$E \left[\langle \cdot, \Pi_1 - E[\Pi_1] \rangle_{\mathcal{H}} (\Pi_2 - E[\Pi_2]) \right] = E \left[(\Pi_1 - E[\Pi_1]) \tilde{\otimes} (\Pi_2 - E[\Pi_2]) \right]. \quad (17)$$

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.

¹See, for instance, Vilenkin (1968, p.59-65).

Theorem 3 (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 4 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

$$\sqrt{n}(\hat{\Pi}_\alpha^* - \Pi_\alpha^*) \xrightarrow{d} \mathcal{N}(0, \Omega_\alpha), \quad (18)$$

where the asymptotic covariance operator² Ω_α for fixed α is given by

$$\Omega_\alpha = E \left[\left((U + \alpha \Pi \tilde{Z}) \otimes \tilde{Z} \right) \tilde{\otimes} \left((U + \alpha \Pi \tilde{Z}) \otimes \tilde{Z} \right) \right] - \alpha^2 C_{\tilde{Z} \Pi \tilde{Z}} \tilde{\otimes} C_{\tilde{Z} \Pi \tilde{Z}}, \quad (19)$$

which simplifies to

$$\Omega_0 = E \left[(U \otimes V_Z^{-1} Z) \tilde{\otimes} (U \otimes V_Z^{-1} Z) \right], \quad (20)$$

when $\alpha \rightarrow 0$.

Proof. 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 3 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 (18) using the continuous transformation $\begin{pmatrix} A \\ B \end{pmatrix} \mapsto (\alpha I + V_Z)^{-1} A + \alpha (\alpha I + V_Z)^{-1} B (\alpha I + V_Z)^{-1} \Pi^*$. The covariance operator of $\sqrt{n}(\hat{\Pi}_\alpha^* - \Pi_\alpha^*)$

²The kernel of Ω_α has four dimensions and its kernel may be written as

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

may be written using (17) as

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

where the second line is obtained from the i.i.d. assumption. Straightforward developments yield (19). Now letting $\alpha \rightarrow 0$ gives (20). ■

Furthermore, under the strict exogeneity assumption, $E[U_i|Z_i] = 0$, the asymptotic covariance operator in (19) is simplified into

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

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 Π^* . On the other hand, if we assume $\alpha \rightarrow 0$, an estimator of (20) 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 (20). 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

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

and $(x \otimes y)(f) = \langle x, f \rangle y$ and $(\Pi_1 \tilde{\otimes} \Pi_2) T = \langle T, \Pi_1^* \rangle_{\mathcal{H}} \Pi_2$ (see Dauxois, Pousse, and Romain, 1982)

Let $\{\phi_j : j = 1, 2, \dots, q\}$ be a set of test functions, then

$$\begin{bmatrix} \sqrt{n} \left\langle \left(\hat{C}_{ZY} - \hat{V}_Z \Pi_0^* \right) \phi_1, \phi_1 \right\rangle \\ \vdots \\ \sqrt{n} \left\langle \left(\hat{C}_{ZY} - \hat{V}_Z \Pi_0^* \right) \phi_q, \phi_q \right\rangle \end{bmatrix}$$

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

$$\begin{aligned} \Sigma_{jl} &= E \left[\left\langle \sqrt{n} \hat{C}_{ZY} \phi_j, \phi_j \right\rangle \left\langle \sqrt{n} \hat{C}_{ZY} \phi_l, \phi_l \right\rangle \right] \\ &= \left\langle \phi_j, V_Z \phi_l \right\rangle \left\langle \phi_j, V_U \phi_l \right\rangle. \end{aligned}$$

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, Hansen, Luttmer, and Scheinkman (1997) or could be normal densities with same small variance but centered at different means.

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

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

See Engl, Hanke, and Neubauer (2000, p.102).

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 i th one. Centorrino (2014) 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 (2014).

Various data-driven selection techniques are compared via simulations in Centorrino, Fève, and Florens (2013).

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

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. We use some smoothing to construct pairs of curves (y_i^m, z_i^m) , $i = 1, 2, \dots, n$ such that $y_i^m \in \mathcal{E}$ and $z_i^m \in \mathcal{F}$. This smoothing can be obtained by approximating the curves by step functions or kernel smoothing for instance. The subscript m corresponds to the smallest number of discrete observations across $i = 1, 2, \dots, n$. m grows with the sample size n .

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 6. $\|z_i^m - z_i\| = O_p(f(m))$ and $\|y_i^m - y_i\| = O_p(f(m))$.

Assumption 7.

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

Proposition 5 *Under Assumptions 1 to 4, 6, and 7, 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.*

7 Case where Z is endogenous

Now, assume Z is endogenous but we observe instrumental variables W such that $\text{cov}(U, W) = 0$. Hence, $E((Y - \Pi Z) \langle W, \cdot \rangle) = 0$. It follows that

$$C_{YW} = \Pi C_{ZW} \quad (21)$$

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

$$C_{WY} = C_{WZ} \Pi^* \quad (22)$$

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

We need the following identification conditions:

Assumption 8. C_{WZ} is injective.

Under this assumption, Π is uniquely defined from (21). To see this, assume that there are two solutions Π_1 and Π_2 to (21). It follows that $(\Pi_1 - \Pi_2) C_{ZW} = 0$ or equivalently $C_{WZ} (\Pi_1^* - \Pi_2^*) = 0$. Hence the range of $(\Pi_1^* - \Pi_2^*)$ belongs to the null space of C_{WZ} . However, under Assumption 6, the null space of C_{WZ} is reduced to zero and thus the range of $(\Pi_1^* - \Pi_2^*)$ 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 Equation (22) 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

$$\hat{\Pi}_\alpha^* = \left(\alpha I + \hat{C}_{ZW} \hat{C}_{WZ} \right)^{-1} \hat{C}_{ZW} \hat{C}_{WY}. \quad (23)$$

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 (23), 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

$$\begin{aligned} \hat{C}_{ZW}\hat{C}_{WY}\psi &= \frac{1}{n^2} \sum_{i,j} \langle y_j, \psi \rangle \langle w_i, w_j \rangle z_i, \\ \hat{C}_{ZW}\hat{C}_{WZ}\hat{\Pi}_\alpha^* \psi &= \frac{1}{n^2} \sum_{i,j} \langle z_j, \hat{\Pi}_\alpha^* \psi \rangle \langle w_i, w_j \rangle z_i. \end{aligned}$$

Taking the inner product with z_l yields n equations

$$\begin{aligned} &\alpha \langle z_l, \hat{\Pi}_\alpha^* \psi \rangle + \frac{1}{n^2} \sum_{i,j} \langle z_j, \hat{\Pi}_\alpha^* \psi \rangle \langle w_i, w_j \rangle \langle z_l, z_i \rangle \\ &= \frac{1}{n^2} \sum_{i,j} \langle z_j, \hat{\Pi}_\alpha^* \psi \rangle \langle w_i, w_j \rangle \langle z_l, z_i \rangle, \quad l = 1, 2, \dots, n \end{aligned}$$

with n unknowns $\langle z_j, \hat{\Pi}_\alpha^* \psi \rangle$, $j = 1, 2, \dots, n$. Then, for each ψ , $\hat{\Pi}_\alpha^* \psi$ can be computed from

$$\hat{\Pi}_\alpha^* \psi = \frac{1}{\alpha} \left[\hat{C}_{ZW}\hat{C}_{WY}\psi - \hat{C}_{ZW}\hat{C}_{WZ}\hat{\Pi}_\alpha^* \psi \right].$$

The computation of $\hat{\Pi}_\alpha \varphi$ can be done using the same approach as in Section 2.

Assumption 9. $C_{ZW}C_{WZ}$ is a trace-class operator and $\left\| \hat{C}_{ZW}\hat{C}_{WZ} - C_{ZW}C_{WZ} \right\|_{HS}^2 = O_p(1/n)$.

Assumption 10. There is a Hilbert-Schmidt operator R from \mathcal{E} to \mathcal{F} and a constant $\beta > 0$ such that $\Pi^* = (C_{ZW}C_{WZ})^{\beta/2} R$.

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

$$\begin{aligned} &\hat{\Pi}_\alpha^* - \Pi^* \\ &= \left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1} \hat{C}_{ZW}\hat{C}_{WY} - \Pi^* \end{aligned} \tag{24}$$

$$= \left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1} \hat{C}_{ZW}\hat{C}_{WU} \tag{25}$$

$$+ \left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1} \hat{C}_{ZW}\hat{C}_{WZ}\Pi^* - (\alpha I + C_{ZW}C_{WZ})^{-1} C_{ZW}C_{WZ}\Pi^* \tag{26}$$

$$+ (\alpha I + C_{ZW}C_{WZ})^{-1} C_{ZW}C_{WZ}\Pi^* - \Pi^*. \tag{27}$$

Proposition 6 *Under Assumptions 1, 2, 8, 9, and 10, the MSE of $\hat{\Pi}_\alpha^* - \Pi^*$ has the same rate of convergence as in Proposition 2.*

8 Simulations

This section consists of a simulation study of the estimator presented earlier. Let $\mathcal{E} = \mathcal{F} = L^2[0, 1]$ and $\mathcal{S} = \mathcal{T} = [0, 1]$. Π is an integral operator from $L^2[0, 1]$ to $L^2[0, 1]$ with kernel $\pi(s, t) = 1 - |s - t|^2$.³ 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

$$Y_i = \Pi Z_i + U_i, \quad i = 1, \dots, n$$

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

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 $i = 1, \dots, n$. 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.

³Simulations 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.

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 $(\tilde{z}_i, \tilde{y}_i)$ and a fixed smoothing parameter $\alpha = .01$.
6. Repeat steps 2-5 100 times and calculate the MSE by averaging the quantities
$$\|\hat{\Pi}_\alpha - \Pi\|_{HS}^2 = \int_{\tilde{\mathcal{T}}} \int_{\tilde{\mathcal{T}}} (\hat{\pi}_\alpha(s, t) - \pi(s, t))^2 dt ds$$
 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).⁴ 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.

Figure 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 towards 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.⁵ The last two columns of Table 1 report the true global (R^2) and extended local (\tilde{R}^2) functional

⁴In practice, the nature of the functions of interest should provide guidance for the researcher with regards to the selection of the appropriate integration method. As we study square integrable functions in this setup, the trapezoidal rule allows reducing the discretization bias with respect to the rectangular rule.

⁵The 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.

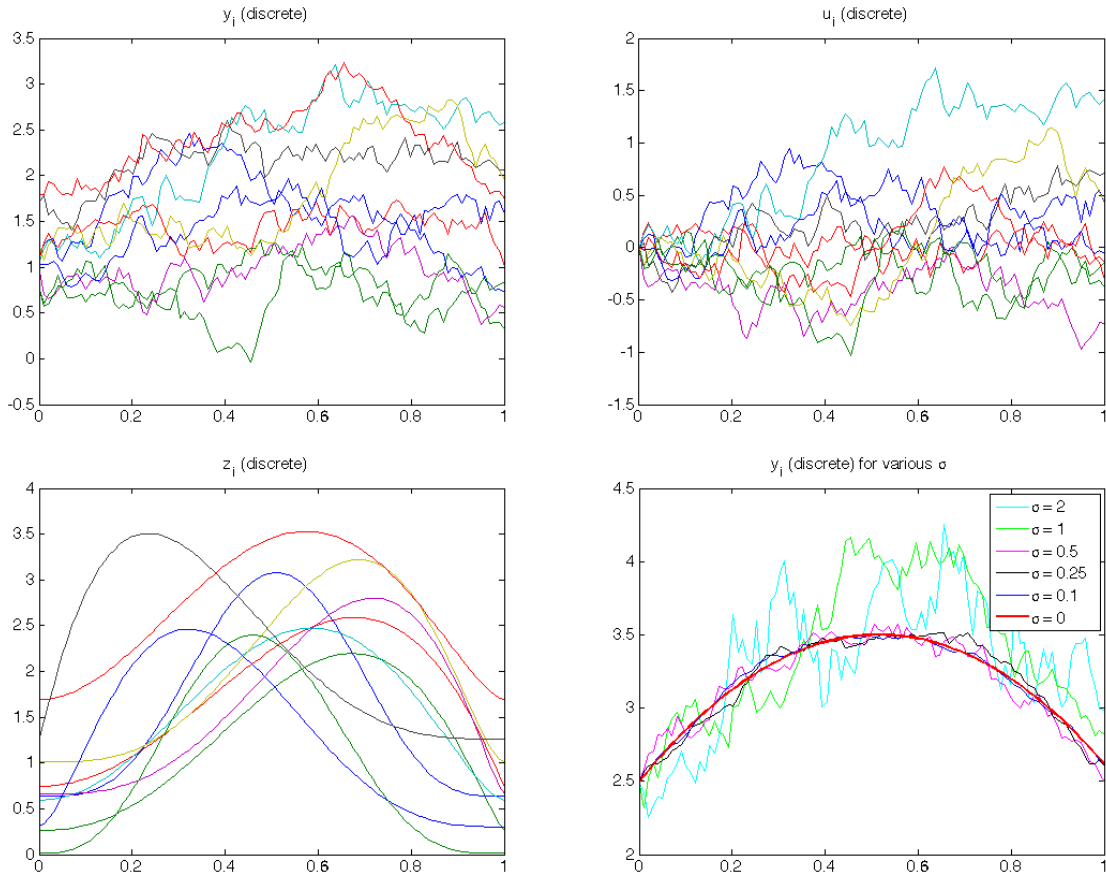


Figure 1: Examples of simulated functions (top left: discretized y_i ; top right: discretized u_i for $\sigma_U = 1$, bottom left: discretized z_i , bottom right: a single y_i for various σ_u).

coefficients of determination, defined as

$$R^2 = \frac{\int_S \text{var}(E[Y(s)|Z])ds}{\int_S \text{var}(Y(s))ds} = \frac{\int_S \text{var}(\Pi Z(s))ds}{\int_S \text{var}(Y(s))ds}$$

$$\tilde{R}^2 = \int_S \frac{\text{var}(E[Y(s)|Z])ds}{\text{var}(Y(s))ds} = \int_S \frac{\text{var}(\Pi Z(s))ds}{\text{var}(Y(s))ds},$$

which are directly related to those proposed in Yao, Muller and Wang (2005).⁶

Table 1: Simulation results: Mean-Square Errors over 100 replications

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

Note: Standard deviations are reported in parentheses.

Simulation 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 error function's standard deviation parameter 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. Figure 2 shows 3D-plots of the actual kernel (top-left), the regularized kernel (top-right), their superposition

⁶These 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 estimator of those coefficients by using $\hat{\Pi}_\alpha$ and the sample counterpart of variance operators.

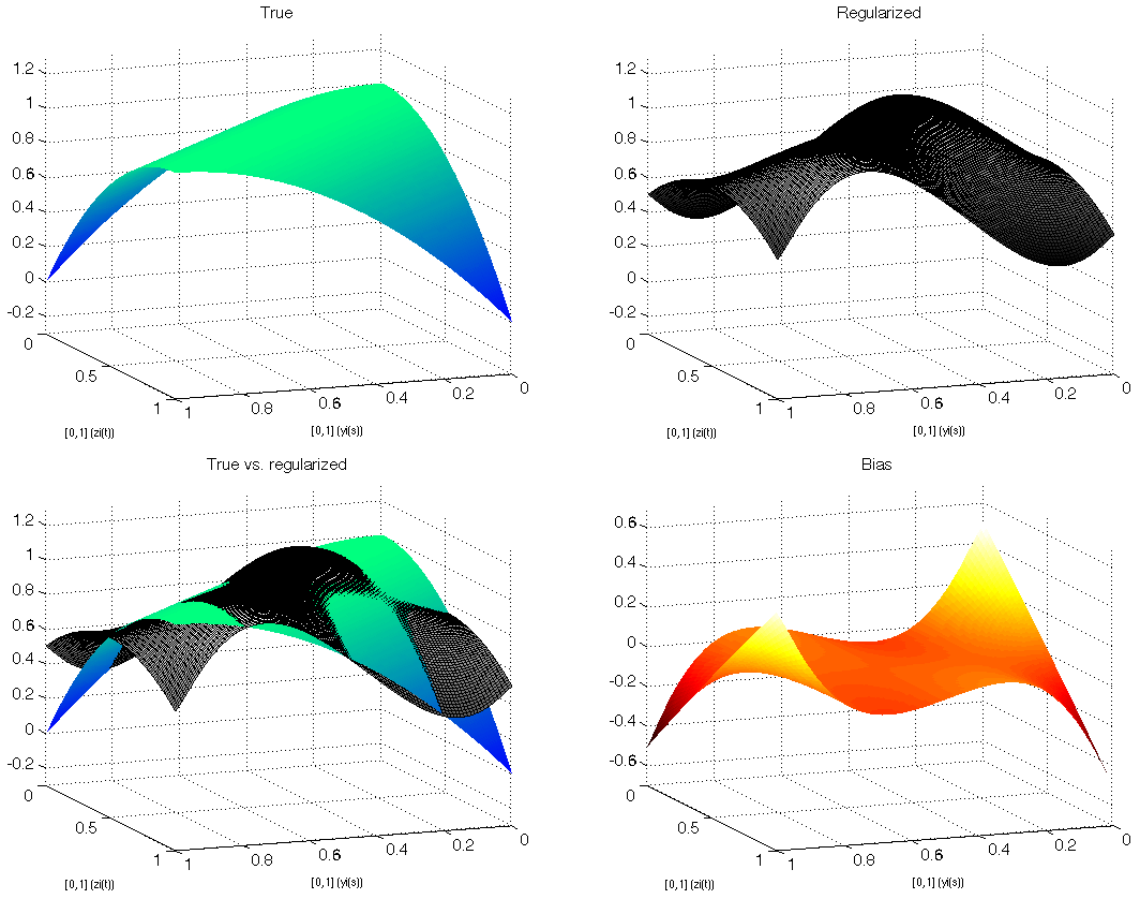


Figure 2: True kernel vs. regularized kernel (top left: True; top right: Regularized, bottom left: True vs. regularized, bottom right: Bias).

(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.

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

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

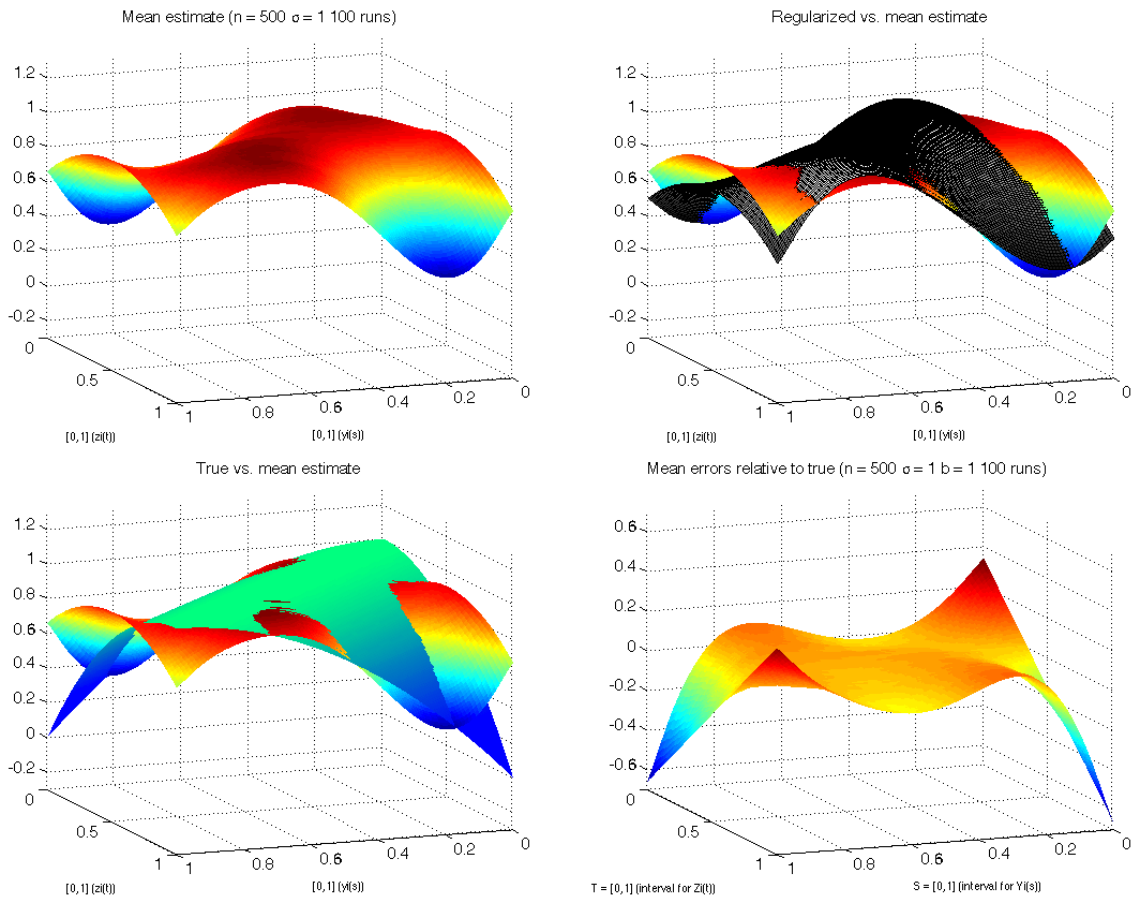


Figure 3: True kernel vs. mean estimate (100 runs with $n = 500$, $\sigma_u = 1$) (top left: Mean estimate, top right: Regularized vs. mean estimate, bottom left: True vs. mean estimate, bottom right: Mean errors)

8.2 Endogenous predictor functions

We consider the design function

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

where $\xi_i(t) = aU_i(t) + c\varepsilon_i(t)$ 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 $t \in [0, 1]$, $\alpha_i, \beta_i \sim iid U[2, 5]$ and $\eta_i \sim iid N(0, 1)$, 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 $\int_S \text{var}(Y(s)) ds$ is unchanged as b varies.⁷ Hence, the choice of b amounts to that of the instrument's strength.

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 $t, s \in \tilde{\mathcal{T}}$.

⁷This assumption allows to keep the variance of Y stable when varying instrument strength. It implies $c = \sqrt{1 + (1 - b^2) \frac{\int_S (\text{var}(\Pi W(s))) ds}{\int_S (\text{var}(\Pi \varepsilon(s))) ds}}$.

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 $\|\hat{\Pi}_\alpha - \Pi\|_{HS}^2 = \int_{\tilde{T}} \int_{\tilde{T}} (\hat{\pi}_\alpha(s, t) - \pi(s, t))^2 dt ds$ 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 . Unsurprisingly, the estimation errors are 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 $\int_S \text{var}(Y(s)) ds$ is fixed.

Table 2: Non-IV estimator: Mean-Square Errors over 100 replications

Instr. strength	Sample sizes				Squared bias	Coef. of deter.	
	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$\ \Pi - \Pi_\alpha\ _{HS}^2$	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
($c = 1$)	(.5305)	(.3525)	(.1581)	(.1121)			

Note: Standard deviations are reported in parentheses.

We now turn to the simulations results for the IV estimator. 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_{FS}^2 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.

Table 3: IV estimator: Mean-Square Errors over 100 replications

Instr. str.	Empirical MSE				Squared bias $\ \Pi - \Pi_\alpha\ _{HS}^2$	Coef. of d.	
	$n = 50$	$n = 100$	$n = 500$	$n = 1000$		R^2	R_{FS}^2
$b = 0.25$.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
($c = 1$)	(.0244)	(.0138)	(.0029)	(.0021)			

Note: Standard deviations are reported in parentheses.

For comparisons with the exogenous case, we provide a final set of surface plots. Figure 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 long tradition of applications to electricity data in functional data analysis (see among others Ferraty and Vieu, 2006, Liebl, 2013), although particularly focused on statistical predictions. Since the fully functional linear regression model has already been considered for electricity demand prediction purposes (Andersson and Lillestol, 2010, Antoch et al., 2008), we propose an application illustrating the usefulness of our estimator in a different domain. We believe that the growing deployment of smart-metering technologies in electricity systems creates a need

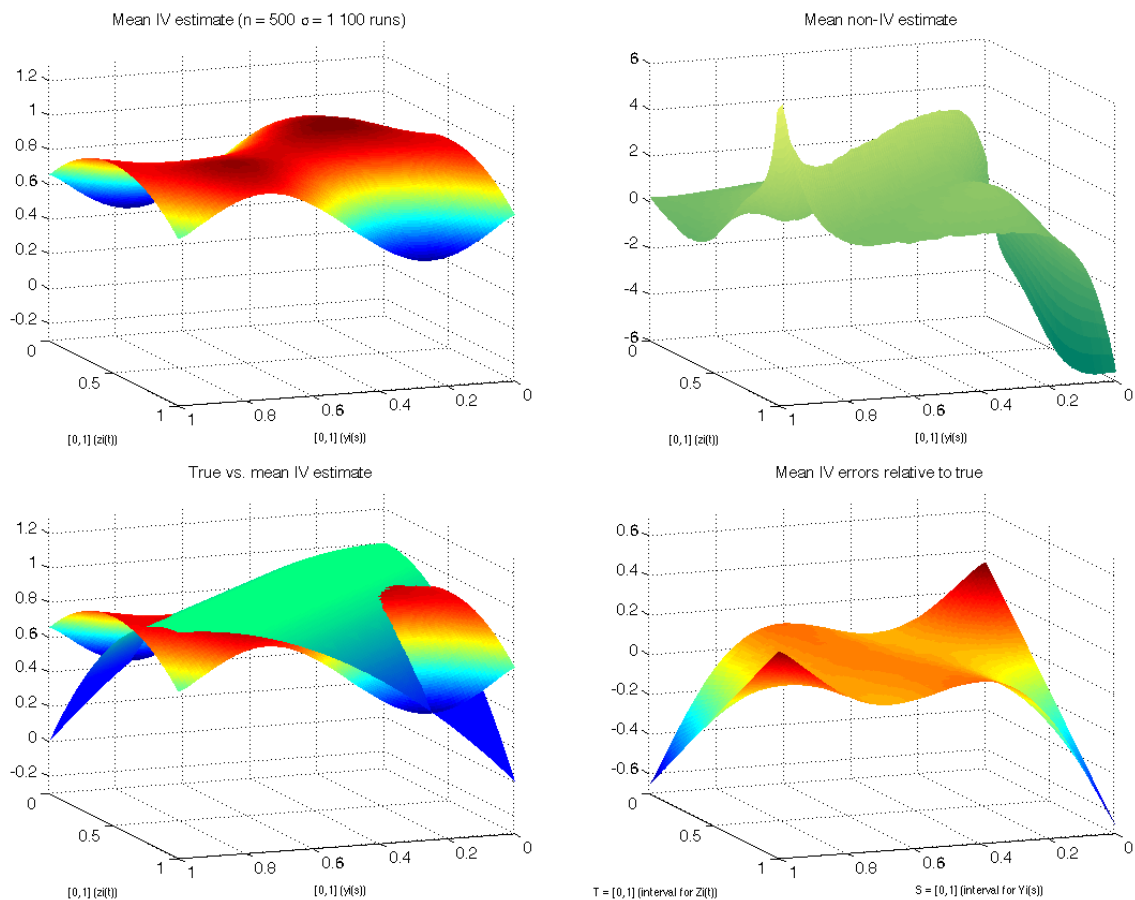


Figure 4: True kernel vs. mean IV estimate (100 runs with $n = 500$, $\sigma_u = 1$ and $b = 0.75$) (top left: Mean estimated IV; top right: Mean estimated non-IV, bottom left: True vs. mean IV estimate, bottom right: Mean IV errors)

for econometric models able to capture the dynamic behaviors of end-users with respect to market fundamentals. Those 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 the day-ahead markets, as proposed by Patrick and Wolak (1997).

We propose a discrete-time model based on the literature on dynamic linear rational expectations (Muth, 1961, Hansen and Sargent, 1980), in which agents have perfect foresight as in Kennan (1979).⁸ Since short-term weather forecasts are widely available and relatively accurate, this should not be too strong of an assumption when considering short time-windows. Electricity consumption is assumed to generate two forms of utility gains for consumers: a non-durable gain from the consumption that has immediate rewards (e.g. using a computer) and a durable gain from electrical heating and air conditioning. The model is easy to estimate and allows the researcher to test a variety of specifications. Like any nonparametric model, it is nevertheless data-intensive and as such may be more suited to microeconomic applications. An essential aspect of the Tikhonov regularization scheme used in the estimation procedure is the choice of the parameter α . Ultimately, we plan to use the leave-one-out CV criterion, but for the moment this is not implemented due to computational issues.

We first present a model that generates the functional regression of interest, then provide details on data construction before performing a preliminary data analysis. We find that some functional series exhibit non-stationarity issues and therefore choose to estimate the functional linear regression model on first-differenced series. Results are interpreted using contour plots. We also report goodness-of-fit statistics and show how to perform general statistical tests on the estimated kernel.

⁸It is also possible to estimate a model like $Y(s) = \int_0^s \pi_0(s, t)Z(t)dt + \int_s^1 \pi_1(s, t)E[Z(t)|I_s]dt + U(s)$, which explicitly accounts for expectations, although it requires a significant departure from the original model presented in this paper.

9.2 The Model

Let us focus on the electricity usage patterns of a region during the winter period, where the use of electrical heating represents a significant share of electricity consumption. The problem can be considered symmetric for the summer period with air conditioning.

At the aggregate level, a representative forward-looking electricity consumer with perfect foresight may be understood as solving a dynamic control problem when choosing its electricity consumption pattern $\{q_t, c_t\}_{-\infty}^{+\infty}$. The agent chooses between the consumption of electrical heating q_t which generates a durable effect through the heat stock denoted y_t , and the consumption for other electrical appliances c_t , which are assumed to generate only instantaneous rewards in utility terms. Furthermore, there exists a desired level of heat stock, y_t^* , that depends linearly on outdoor temperature, x_t , some function of time ξ_{1t} and a random component unobservable to the econometrician, ε_t .⁹ In particular, the effective impact of a given heat stock on utility is presumed to increase with outside air temperature, since heating needs are expected to rise at the aggregate level when the temperature drops. The problem is given by

$$\begin{aligned}
 & \max_{\{y_{t+1}, c_t\}_{-\infty}^{+\infty}} \sum_{t=-\infty}^{+\infty} \beta^t \left(u_t(c_t) - \frac{(y_t - y_t^*)^2}{2} - p_t(c_t + q_t) \right) \\
 & \text{s.t.} \quad y_{t+1} - y_t = \delta(x_t - y_t) + \eta q_t, \quad \forall t \\
 & \quad \sum_{t=-\infty}^{+\infty} p_t(c_t + q_t) \leq B \quad (\lambda) \\
 & \quad (y_{t+1} - y_t)^2 \leq R \quad \left(\frac{\gamma_t}{2}\right) \\
 & \quad y_0 = y_0^*, \text{ given} \\
 & \quad y_t^* = \xi_{1t} + \xi_2 x_t + \varepsilon_t \\
 & \quad \{\xi_{1,t}, x_t, \varepsilon_t\}_{-\infty}^{+\infty}, \text{ given}
 \end{aligned} \tag{28}$$

where p_t is the energy price, $q_t + c_t$ is total electricity consumed, $u_t(\cdot)$ is a strictly increasing and concave utility function associated with the consumption of c_t , and B is a given budget for electricity expenses. The first constraint corresponds to the law of motion of y_t , it explicitly assumes that the heat stock's depreciation rate depends on

⁹This is a standard assumption in dynamic partial adjustment models where the desired level of capital stock depends on exogenous *forcing* variables related to profitability (Kennan, 1979).

outdoor temperature.¹⁰ The second is a budget constraint. Finally, $(y_{t+1} - y_t)^2 \leq R$ is a ramping constraint on the heat stock, i.e. heat stock adjustments are limited through time.¹¹ We further impose the shadow cost associated with this constraint to be positive and constant over time, i.e. $\gamma_t = \gamma > 0, \forall t$.¹² The first order conditions are given by

$$p_t(1 + \lambda) = u'_t(c_t), \text{ and} \quad (29)$$

$$y_{t+1} - \frac{\gamma + \beta(1 + \gamma)}{\beta\gamma}y_t + \frac{1}{\beta}y_{t-1} = \frac{1 + \lambda}{\eta\beta\gamma}(p_t - (1 - \delta)p_{t+1}) - \frac{y_t^*}{\gamma}. \quad (30)$$

Equation (29) is a usual optimality condition stating that the marginal utility of consuming an additional unit of c_t must equalize the cost of doing so (which includes the shadow cost of the budget constraint). We will impose a linear form for $u'_t(\cdot)$ later on in the text. Equation (30) is a Euler equation that determines the optimal path for the heat stock. This second-order difference equation can be solved following the method in Hansen and Sargent (1980). First, one must solve for the roots of the characteristic second-order lag polynomial, denoted ρ_1 and ρ_2 , using the factorized form

$$1 - \frac{\gamma + \beta(1 + \gamma)}{\beta\gamma}z + \frac{1}{\beta}z^2 = (1 - \rho_1z)(1 - \rho_2z). \quad (31)$$

Assuming $\beta, \delta \in (0, 1)$ and $\gamma > 0$ is sufficient to have $\rho_1 \in (0, 1)$ and $\rho_2 = \frac{1}{\beta\rho_1}$, and obtain the unique solution of this problem given by

$$y_t = \rho_1 y_{t-1} - \rho_1 \sum_{j=0}^{\infty} (\beta\rho_1)^j \left(\frac{1 + \lambda}{\eta\beta\gamma} (p_t + j - (1 - \delta)p_{t+j+1}) - \frac{y_{t+j}^*}{\gamma} \right), \quad (32)$$

This expression is standard in dynamic models with quadratic adjustment costs.¹³ We

¹⁰This follows the electrical engineering literature, which generally assumes the motion of a building's inside air temperature to depend on outside air temperature (see McLaughlin et al., 2011 and Yu et al., 2012).

¹¹It is a convenient assumption since it allows to introduce quadratic adjustment costs in the model. It allows to bound power usage for heating purposes which is one explanation to the fact that electrical heaters and air conditioners take time to have a significant impact on the heat stock.

¹²This assumption allows to have a closed-form solution to the model following Hansen and Sargent (1980), since the second-order difference equation given by the Euler conditions will have constant parameters.

¹³see for instance equation (4) in Hansen and Sargent (1980) p.12.

can now make use of the lag operator L to get rid of the autoregressive term by writing

$$y_t(1 - \rho_1 L) = A + \frac{\rho_1}{\gamma} \sum_{j=0}^{\infty} (\beta \rho_1)^j y_{t+j}^*, \quad (33)$$

with $A = -(\frac{\rho_1(1+\lambda)}{\eta\beta\gamma}) \sum_{j=0}^{\infty} (\beta \rho_1)^j (p_{t+j} - (1-\delta)p_{t+j+1})$. Clearly, A depends on the whole price path and the tightness of the budget constraint. Assuming a constant retail electricity price, p_t , hence yields a constant A . Furthermore, since $|\rho_1| < 1$, the lag polynomial in (33) can be inverted in order to obtain

$$y_t = \frac{A}{1 - \rho_1} + \frac{\rho_1}{\gamma(1 - \rho_1)} \sum_{j=1}^{\infty} (\rho_1)^j y_{t-j}^* + \frac{\rho_1}{\gamma(1 - \beta\rho_1^2)} y_t^* + \frac{\rho_1}{\gamma(1 - \beta\rho_1^2)} \sum_{j=1}^{\infty} (\beta\rho_1)^j y_{t+j}^*, \quad (34)$$

which depends on outdoor temperature through its effect on the desired heat stock. Let us now define the desired consumption level for heating as $q_t^* = \frac{1}{\eta}(y_{t+1}^* - (1-\delta)y_t^* - \delta x_t)$. Multiplying both sides of (34) by $(1 - (1-\delta)L)L^{-1}$ yields

$$\begin{aligned} q_t = \frac{1}{\eta} & \left(\frac{\delta}{1 - \rho_1} A - \delta x_t + \frac{\rho_1}{\gamma(1 - \rho_1)} \sum_{j=1}^{\infty} (\rho_1)^j (\eta q_{t-j}^* + \delta x_{t-j}) + \frac{\rho_1}{\gamma(1 - \beta\rho_1^2)} (\eta q_t^* + \delta x_t) \right. \\ & \left. + \frac{\rho_1}{\gamma(1 - \beta\rho_1^2)} \sum_{j=1}^{\infty} (\beta\rho_1)^j (\eta q_{t+j}^* + \delta x_{t+j}) \right). \end{aligned} \quad (35)$$

From $y_t^* = \xi_{1t} + \xi_2 x_t + \varepsilon_t$, we obtain $\eta q_t^* = (\xi_{1,t+1} - (1-\delta)\xi_{1t}) + \xi_2(x_{t+1} - (1-\delta)x_t) + (\varepsilon_{t+1} - (1-\delta)\varepsilon_t)$. Substituting this expression in (35) gives us the decision rule we are after

$$q_t = \kappa_{0,t} + \sum_{j=-\infty}^{+\infty} \kappa_{1,t,j} x_{t+j} + \epsilon_t, \quad (36)$$

with

$$\begin{aligned} \kappa_{0,t} = & \frac{\eta^{-1}\delta}{(1 - \rho_1)} A - \eta^{-1}\delta x_t + \frac{\eta^{-1}\rho_1}{\gamma(1 - \rho_1)} \sum_{j=1}^{\infty} (\rho_1)^j (\xi_{1,t-j+1} - (1-\delta)\xi_{1,t-j}) \\ & + \frac{\eta^{-1}\rho_1}{\gamma(1 - \beta\rho_1^2)} \sum_{j=0}^{\infty} (\beta\rho_1)^j (\xi_{1,t+j+1} - (1-\delta)\xi_{1,t+j}), \end{aligned}$$

$$\begin{aligned} \sum_{j=-\infty}^{+\infty} \kappa_{1,t,j} x_{t+j} &= \frac{\eta^{-1} \rho_1}{\gamma(1-\rho_1)} \sum_{j=1}^{\infty} (\rho_1)^j (x_{t-j+1} - (1-\delta)x_{t-j}) \\ &\quad + \frac{\eta^{-1} \rho_1}{\gamma(1-\beta\rho_1^2)} \sum_{j=0}^{\infty} (\beta\rho_1)^j (x_{t+j+1} - (1-\delta)x_{t+j}), \text{ and} \end{aligned}$$

$$\epsilon_t = \frac{\eta^{-1} \rho_1}{\gamma(1-\rho_1)} \sum_{j=1}^{+\infty} (\rho_1)^j (\varepsilon_{1,t-j+1} - (1-\delta)\varepsilon_{1,t-j}) + \frac{\eta^{-1} \rho_1}{\gamma(1-\beta\rho_1^2)} \sum_{j=0}^{\infty} (\beta\rho_1)^j (\varepsilon_{t+j+1} - (1-\delta)\varepsilon_{t+j}).$$

We can solve for the coefficients of (36) to get a simpler expression (to be done). Changing variable $j = s - t$ gives

$$q_t = \kappa_{0,t} + \sum_{s=t-\infty}^{\infty} \kappa_{1,t,s-t} x_s + \epsilon_t. \quad (37)$$

which allows to write q_t as

$$q(t) = \kappa_0(t) + \sum_{s=-\infty}^{\infty} \kappa_1(t, s) x(s) + \epsilon(t), \quad (38)$$

for any t . This model can be estimated as is, only if one has information on the electricity consumed for electrical heating. This is hardly observable in practice. Instead we can proceed as follows. Given (29), one has $c_t = u_t'^{-1}(p(1+\lambda))$. Since $p(1+\lambda)$ is a constant, we assume $u_t'^{-1}(p) = G_t \Gamma + \nu_t \forall p$, where G_t is a vector of exogenous covariates, Γ is the vector of associated parameters and ν_t is an error term. Substituting into the FOC gives $c_t = G_t \Gamma + \nu_t$. Now, if one only has data on $Q(t) = q_t + c_t$, the model becomes

$$q(t) + c(t) = \kappa_0(t) + G(t)\Gamma + \sum_{s=-\infty}^{\infty} \kappa_1(t, s) x(s) + \epsilon(t) + \nu(t), \quad (39)$$

which may be simplified to

$$Q(t) = \kappa_0^*(t) + \sum_{s=-\infty}^{\infty} \kappa_1(t, s) x(s) + u(t), \quad (40)$$

making clear the fact that electricity consumption c_t with instantaneous utility gains

only has a immediate impact of aggregate consumption.¹⁴ Henceforth, appropriate de-seasonalization of the series should allow to single out the dynamic effects of temperature upon electricity consumption. The decision rule for the analogous continuous-time model is therefore given by

$$Q(t) = \pi_0(t) + \int_{-\infty}^{\infty} \pi_1(t, s)x(s)ds + u(t). \quad (41)$$

This model can be estimated using the procedure described above if one has an *i.i.d.* sample of observations $\{q_i, x_i\}_{i=1, \dots, n}$, and if u_i is presumed to be a mean-zero *i.i.d.* random functional error process. Remark that this assumption does not prevent $u_i(t)$ to be correlated with $u_i(t')$, $\forall t, t'$, which is clearly the case in the structural model developed above. A suited data set would hence consist of micro-data from smart-meters at different location within a given region. However, such data is not publicly available to the best of our knowledge, therefore we propose to estimate the model using daily patterns extracted from annual trajectories, which will exhibit serial correlation in the error terms by design. The model to be estimated is

$$Q_i(t) = \pi_0(t) + \int_I \pi_1(t, s)x_i(s)ds + u_i(t), \quad (42)$$

where I is an interval of $(-\infty, +\infty)$.

9.3 Preliminary Data Analysis

Prior to presenting the data set construction, let us present some facts about the electricity market in Ontario. There are two types of consumers in the province. First, small consumers (residential and small business) 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. Also, non-linear pricing schemes (tiered rates) apply to 10% of small consumers and an even smaller share has fixed-rate contracts with retailers.¹⁵ Second, large consumers (large business and the public sector) are subject to the

¹⁴This is of course a consequence of the chosen functional form for the utility function.

¹⁵Further details are available at <http://www.ontarioenergyboard.ca/OEB/Consumers/Electricity/Electricity+Prices>

wholesale market price.¹⁶ The wholesale market price is determined on an hourly basis in a uniform-price multi-unit auction subject to operational constraints. It is considered as being quite volatile. Therefore, 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 as bilateral contracts represent a small share of total exchanges in Ontario. 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. In the model presented before, we assumed the price to be constant since short-term demand for electricity is typically perceived as being very price inelastic.¹⁷ Note that it would be possible to relax this assumption by using the time-of-use prices and the wholesale price. The latter being endogenous, one would need to use the instrumental variable version of the Tikhonov estimator. A natural instrument for the wholesale electricity price in Ontario is wind power production, which depreciates prices significantly and is as good as randomly assigned. We do not pursue this idea here.

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 through the system operator’s website,¹⁸ whereas hourly province-wide temperature data have been constructed from hourly measurements at 77 weather stations in Ontario,¹⁹ publicly available on Environment Canada’s website.²⁰

Let $X(t)$ denote our measure of temperature in hour t for the entire province. It is constructed in three steps. First, we match a set of 41 Ontarian cities (of above 10,000 inhabitants)²¹ to their three nearest weather stations, then compute a weighted average using a distance metric. Finally, we obtain $X(t)$ as a weighted average of cities’ temperature, where weights are defined as each city’s relative population. The

¹⁶Roughly speaking, businesses are considered large when their electricity bills exceed \$2,000 per month.

¹⁷For example, the recent implementation of time-of-use pricing has been proved to have had limited effects on electricity consumption in Ontario (Faruqui et al., 2013)

¹⁸<http://www.ieso.ca>

¹⁹The complete data set contains 139 weather stations although once matched to neighboring cities, only 77 are found relevant.

²⁰<http://climat.meteo.gc.ca/>

²¹Those cities represent 85.3% of the province’s population as of 2011.

Table 4: Descriptive statistics

Year	Hourly temperature (celsius)				Hourly load (GW)				Obs
	Mean	SD	Min	Max	Mean	SD	Min	Max	
2010	8.27	9.17	-17.12	29.13	16.23	2.59	10.62	25.07	8760
2011	8.02	9.37	-17.96	31.15	16.15	2.45	10.76	25.43	8760
2012	9.32	8.67	-14.32	30.57	16.09	2.40	10.99	24.60	8784
2013	7.55	9.36	-17.40	28.88	16.07	2.39	10.77	24.49	8760
2014	7.58	11.02	-19.84	26.52	16.08	2.36	10.71	22.77	6552
All	8.18	9.49	-19.84	31.15	16.12	2.45	10.62	25.43	41616

constructed province-wide hourly temperature variable is formally defined by

$$X(t) = \sum_c \gamma_c \left\{ \sum_{w(c)} \rho_{w(c)} X_{w(c)}(t) \right\}, \quad \forall i, h$$

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 and $X_w(t)$ is temperature measurement at station w in hour t .²² 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 1 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.

This market operates over two distinct periods each year. The winter period spans from November 1 to April 30 and has two daily peak periods: one in the morning (7am-11am) and the other after worktime (5pm-7pm), a mid-peak period (11am-5pm) and an off-peak period (7pm-7am). The remaining part of the year is considered as the summer period and mid-peak and peak periods are reversed with respect to winter. Given this fact and the model developed above, dividing up the sample into two corresponding

²²Distance is calculated as the difference 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. This is an implicit assumption on the spatial distribution of electricity consumers around the cities.

periods appears natural. We define the periods consequently as shown in Figure 5 (blue for winter and red for summer) so decide to discard the spring, from May 1 to May 31, and autumn, from Sep 15 to Oct 31, periods (black line) given the limited heating or and air conditioning needs in these periods. Figure 6 displays the relationship between electricity consumption and temperature in winter, summer and the discarded months.

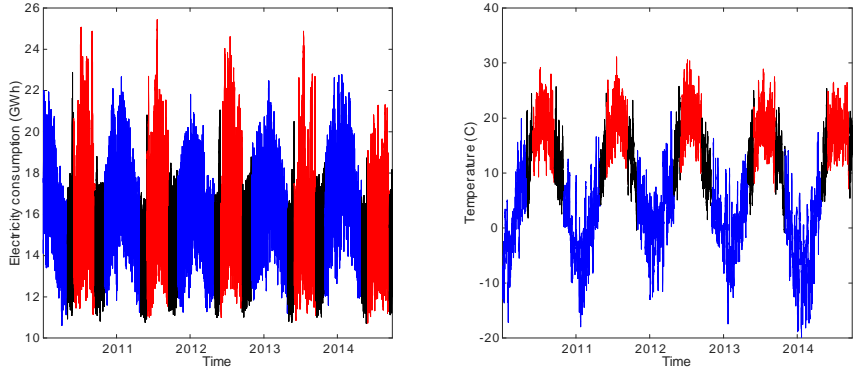


Figure 5: Entire data series

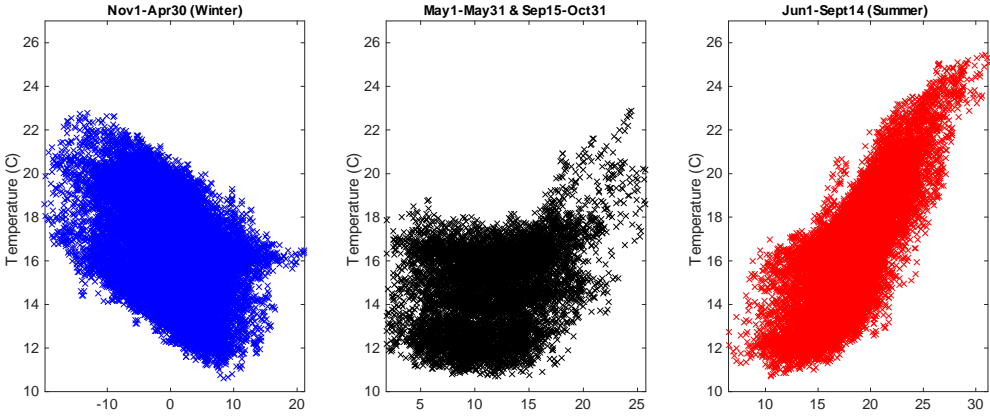


Figure 6: Electricity Consumption (GWh) vs. Temperature (celsius)

The plots show evidence of a relatively linear relationship between the load and temperature for winter and summer months. On the other hand, the relation is much

flatter in October and May. The figures also suggest that power usage is more sensitive to warm weather than cold weather. Most likely because cooling is more energy-intensive than heating.

For each of the four sample periods, holidays and weekends are dropped. The time-series are then linearly projected onto a set of binary variables for hours of the day, days of the week, weeks of the year and years, in order to control for unobserved seasonalities unrelated to temperature which are captured into $\pi_0(t)$. This procedure is motivated by the FWL theorem, and is analogous to adding fixed-effects into model (42). Next, for ease of interpretation, we transform the temperature variables so as to obtain upward-sloping versions of those in Figure 6 that are defined on the positive real line using

$$\begin{aligned} X_h(t) &= X(t) - \min_{X \in \text{Summer}} (X(t)), \text{ and} \\ X_c(t) &= \max_{X \in \text{Winter}} (X(t)) - X(t), \end{aligned} \tag{43}$$

which have strictly positive linear correlations with the load.

The estimation samples are constructed with daily trajectories of 25 discrete observations for the dependent variable and a three-day window of 73 observations for the predictor variable. This window is chosen so that the dependent will always be regressed on at least 24 lagged hours and 24 future hours. By design, the residual term in model (42) will exhibit serial correlation across daily trajectories. We focus on Thursdays with the purpose of alleviating serial correlation in the residuals from one day to the next, but still find considerable serial correlation within the samples and even diagnose non-stationarities in the residual functions. Thus, we focus on the trajectories in first-differences shown in Figure 7 and 8 for demand (left) and temperature (right).²³

At this point it seems important to survey the recent literature on temporal aspects of functional data analysis. Gabrys, Horvath, and Kokoszka (2010) state that error correlation in a fully functional linear regression model like the one analyzed bears the

²³Plots of the trajectories in levels and corresponding residuals are to be found in to Appendix.

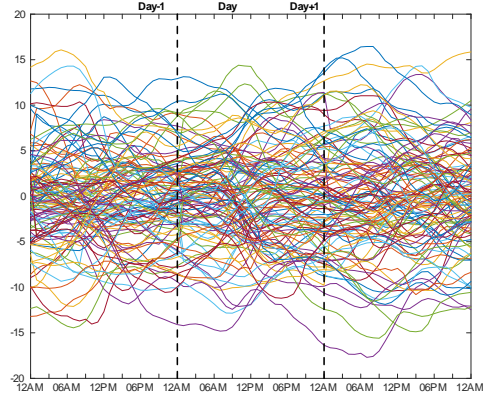
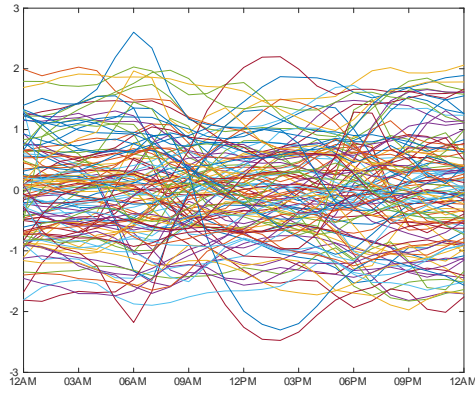


Figure 7: Sample of first-differences for Thursdays (Winter)

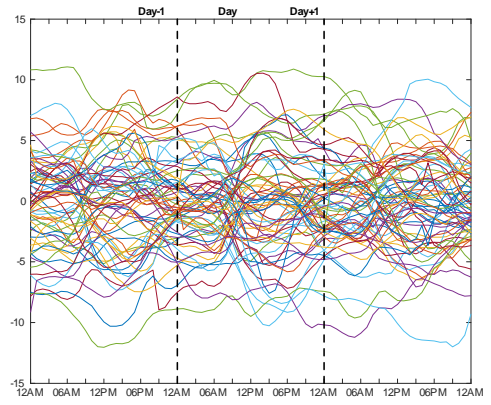
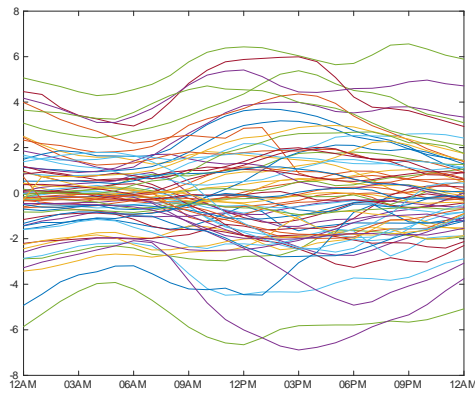


Figure 8: Sample of first-differences for Thursdays (Summer)

same consequence than in a multivariate regression, i.e. it affects variance estimates and consequently the distribution of test statistics. Horvath, Huskova, and Rice (2013) develop a test of independence for functional data. Hormann and Kokoszka (2010) show that the estimation procedure of Yao, Muller and Wang (2005), which is related to our setting, is robust to weak dependence. Indeed, even with some temporal dependence, the estimation of 44 should not be a problem as long as the functional variables are stationary. Fortunately, there exist stationarity tests related to the standard KPSS tests for stationarity of functional time series (Horvath, Kokoszka, and Rice, 2014). For the time being, we do not extend those testing procedures and results to our setting although this could be done in future research.

9.4 The Functional Model and Estimation Results

Based on the model developed above and the data considerations, let $\mathcal{E} = L^2[0, 24]$ and $\mathcal{F} = L^2[-24, 48]$ with $\mathcal{S} = [0, 24]$ and $\mathcal{T} = [-24, 48]$. We specify the functional linear regression model,

$$Q_i(t) = \pi_0(t) + \int_{[-24,48]} \pi_1(t, s)X_i(s)ds + U_i(t), \quad (44)$$

where $Q_i(t)$ is total electricity consumption in hour t , $\pi_0(s)$ is a constant function, $X_i(s)$ is temperature in hour $s \in [-24, 48]$, and $U_i(t)$ is a zero-mean error term. The object of interest is the kernel π_1 , which characterizes the dynamic relation between electricity consumption for heating/AC needs and temperature patterns. Since we are interested in daily electricity consumption patterns, the cross-sectional unit i denotes a daily observation.

Figure 9 and 10 display contour plots of the estimated kernels for a smoothing parameter $\alpha = 2$ using first-differenced trajectories, for the winter period and summer period respectively. The estimated kernel remains unchanged for higher difference orders. Plots of the residuals and goodness-of-fit statistics are included in the appendix.

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 functionals. Let us consider an illustrative example. We observe from Figure 9 that the contemporaneous effect of temperature in winter is larger early

morning and in the afternoon (red regions); which correspond to the on-peak periods. More generally, one can read the results both horizontally and vertically. The effect of the entire temperature pattern on electricity consumption at a given hour of the day is observed horizontally, whereas the effect of temperature at a specific time upon the daily electricity consumption pattern is read vertically. The magnitudes of the correlation are indicated using colors from dark blue to dark red with corresponding values given in the legend. The *seemingly S shape* of the kernel around the indicative diagonal is likely to be an artefact of the Tikhonov regularization, as observed in our simulations, and consequently interpretative attempts of the results should account of this feature.

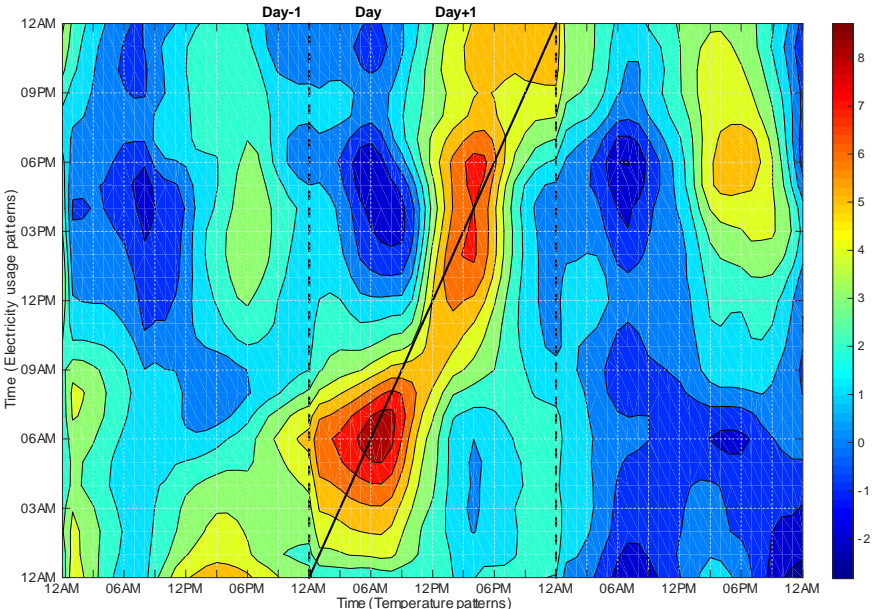


Figure 9: Contour plots of estimation results for winter months with $\alpha = 2$

In general, we observe that the effect of temperature on the consumption is of larger magnitude in the summer period. Furthermore, the lagged and anticipatory effects appear to span over a longer interval around the diagonal. The red areas, where the magnitude of the correlation is highest, show in particular that electricity consumption from 2 am to 12 am strongly depends on temperatures from 7 pm to 12 am, suggesting that consumers anticipate evening temperatures by turning up the A/C. Also, afternoon

consumption appears to depend on morning weather, suggesting the existence lag effects. Nevertheless, the kernel for summer is hard to interpret and statistical tests could provide further insights on the results.

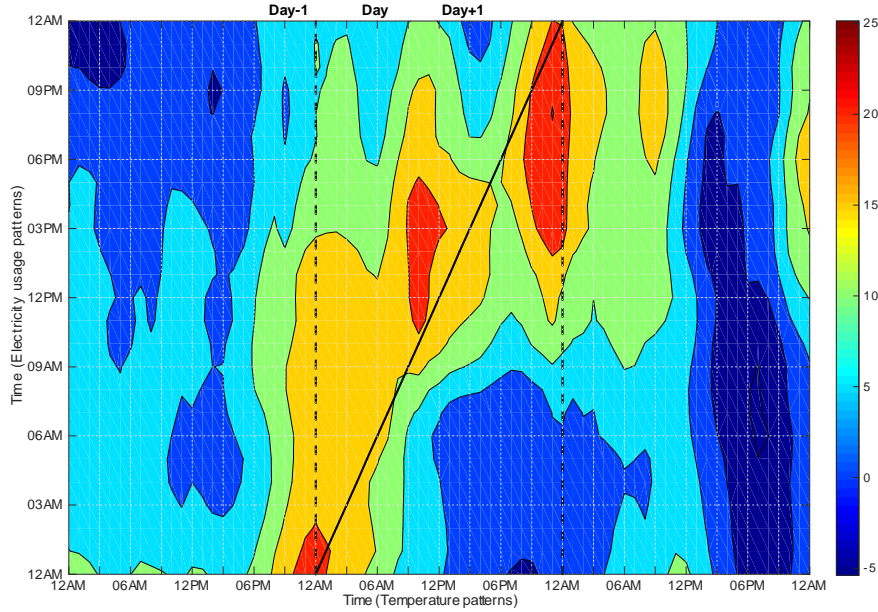


Figure 10: Contour plots of estimation results for summer months with $\alpha = 2$

Clearly, estimation results suggest the existence of both lag and anticipatory effects. Even for periods longer than a couple of hours. For instance in summer, the daily consumption pattern appear to depend negatively, though to a relatively low extent, on the day-after's evening temperatures. This suggest that during the hot season when temperatures are persistently high for days, consumers may choose to use less A/C. In general, the estimated coefficients of past, current and future values of temperature also seem to vary greatly across hours of the day.

9.5 Conclusion of the application

This simple application sought to shed light on the potential uses of our estimator in a different context than statistical prediction. The empirical analysis is based on a linear rational expectations model with perfect foresight. We use the functional model to

quantify the dynamic relationship between daily electricity consumption and temperature patterns in the province of Ontario. The data set consists of hourly consumption and hourly temperature data at the province level. Results show that electricity consumption decisions are subject to not only contemporaneous temperature but also past and future realizations. This relation seems to differ for winter and summer months and across hours of the day. When evaluating future temperatures, we use actual values rather than expectations since short-term weather forecasts may be considered reasonably accurate.

A Appendix

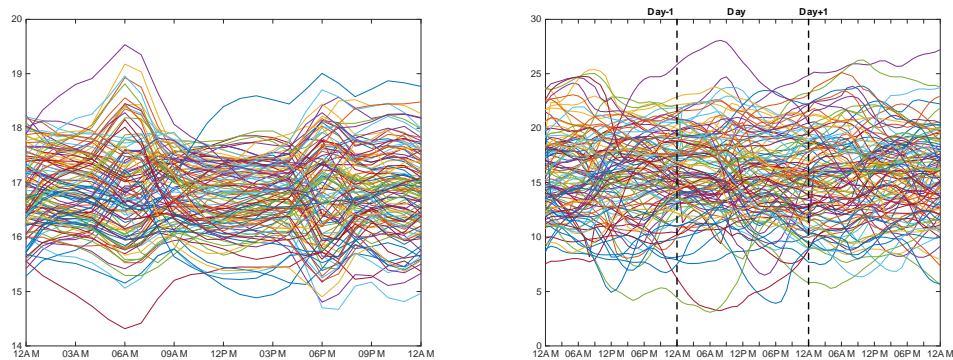


Figure 11: Sample of levels for Thursdays (Winter)

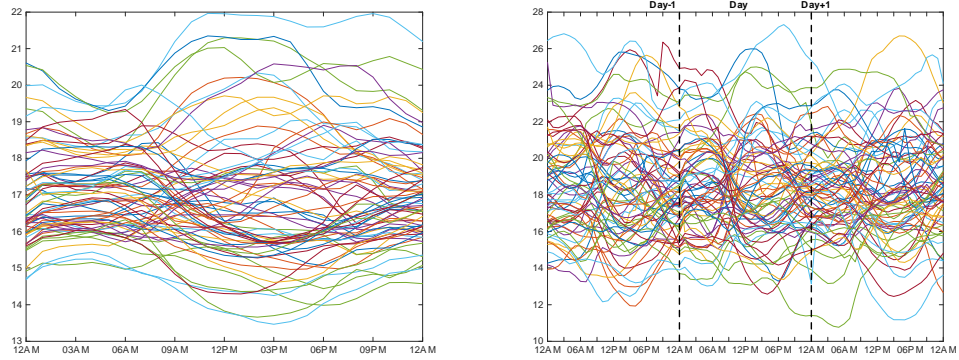


Figure 12: Sample of levels for Thursdays (Summer)

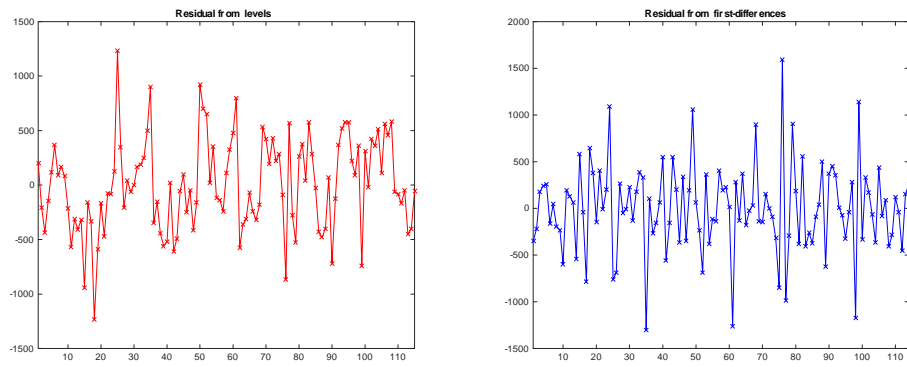


Figure 13: Residuals for series in levels and differences (Winter)

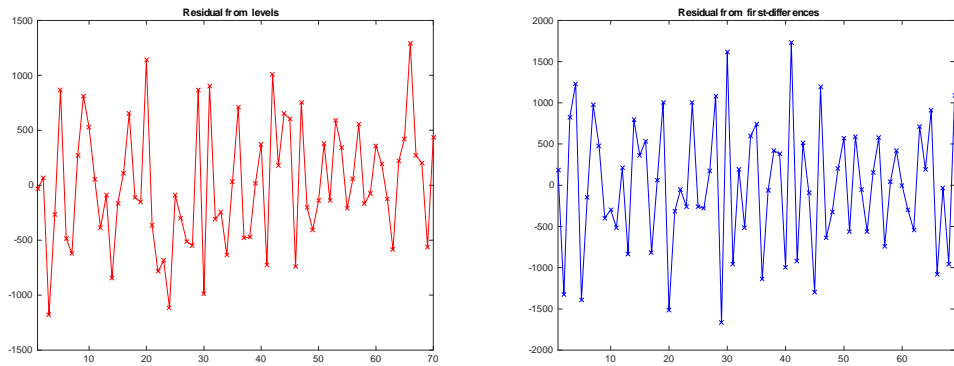


Figure 14: Residuals for series in levels and differences (Summer)

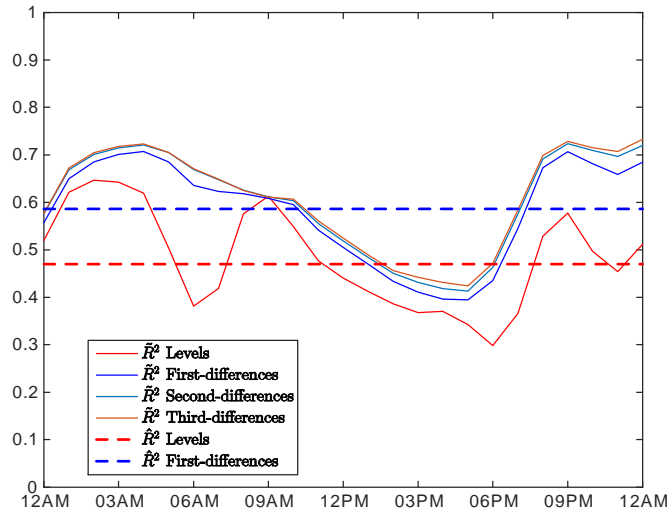


Figure 15: Coefficients of determination (local and global) for series in levels and differences (Winter)

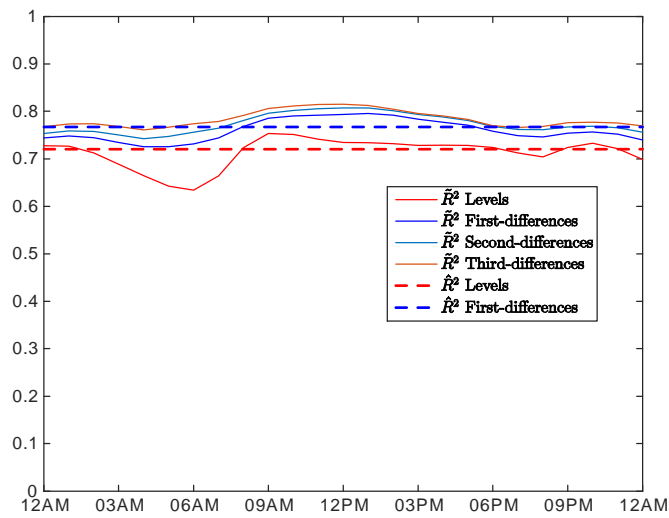


Figure 16: Coefficients of determination (local and global) for series in levels and differences (Summer)

B Proofs

Proof of Proposition 1.

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

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 7 *Let $A = B + C$ where B is a zero mean random operator and C is a non-random operator. Then,*

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

Proof of Lemma 7.

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

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

Lemma 8 *Let A be a random operator from \mathcal{E} to \mathcal{F} .*

$$E(\|A\|_{HS}^2) = \text{tr} E(A^*A).$$

Proof of Lemma 8. We have

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

We turn to the proof of Proposition 2. Applying Lemma 7 on the decomposition

(12), (13), and (14), we have

$$\begin{aligned}
& E \left(\left\| \hat{\Pi}_\alpha - \Pi \right\|_{HS}^2 \mid Z_1, Z_2, \dots, Z_n \right) \\
&= E \left(\|(12)\|_{HS}^2 \mid Z_1, Z_2, \dots, Z_n \right) + \|(13) + (14)\|_{HS}^2 \\
&\leq E \left(\|(12)\|_{HS}^2 \mid Z_1, Z_2, \dots, Z_n \right) + 2 \|(13)\|_{HS}^2 + 2 \|(14)\|_{HS}^2.
\end{aligned}$$

We study the first term of the r.h.s. By Lemma 8,

$$\begin{aligned}
& E \left(\|(12)\|_{HS}^2 \mid Z_1, Z_2, \dots, Z_n \right) \\
&= E \left(\left\| \left(\alpha I + \hat{V}_Z \right)^{-1} \hat{C}_{ZU} \right\|_{HS}^2 \mid Z_1, Z_2, \dots, Z_n \right) \\
&= \text{tr} E \left(\left(\alpha I + \hat{V}_Z \right)^{-1} \hat{C}_{ZU} \hat{C}_{ZU}^* \left(\alpha I + \hat{V}_Z \right)^{-1} \mid Z_1, Z_2, \dots, Z_n \right) \\
&= \text{tr} \left\{ \left(\alpha I + \hat{V}_Z \right)^{-1} E \left(\hat{C}_{ZU} \hat{C}_{ZU}^* \mid Z_1, Z_2, \dots, Z_n \right) \left(\alpha I + \hat{V}_Z \right)^{-1} \right\}.
\end{aligned}$$

Note that

$$\begin{aligned}
\hat{C}_{ZU} \hat{C}_{ZU}^* \varphi &= \frac{1}{n^2} \sum_{i,j} z_i \langle z_j, \varphi \rangle \langle u_i, u_j \rangle, \\
E \left(\hat{C}_{ZU} \hat{C}_{ZU}^* \varphi \mid Z_1, Z_2, \dots, Z_n \right) &= \frac{1}{n} \sum_i z_i \langle z_i, \varphi \rangle E [\langle u_i, u_i \rangle \mid Z_1, Z_2, \dots, Z_n] \\
&= \frac{1}{n} \sum_i z_i \langle z_i, \varphi \rangle \text{tr} (V_U) \\
&= \frac{1}{n} \text{tr} (V_U) \hat{V}_Z \varphi
\end{aligned}$$

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

$$\begin{aligned}
E \left(\|(12)\|_{HS}^2 \mid Z_1, Z_2, \dots, Z_n \right) &= \frac{1}{n} \text{tr} (V_U) \text{tr} \left(\left(\alpha I + \hat{V}_Z \right)^{-1} \hat{V}_Z \left(\alpha I + \hat{V}_Z \right)^{-1} \right) \\
&\leq \frac{C}{n\alpha}
\end{aligned}$$

where C is a generic constant. It follows that $E(\|(12)\|_{HS}^2) \leq \frac{C}{n\alpha}$.

Now, we turn toward the term (13). We have

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

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

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

Hence,

$$\begin{aligned} (13) &= \left[-\alpha (\alpha I + \hat{V}_Z)^{-1} + \alpha (\alpha I + V_Z)^{-1} \right] \Pi^* \\ &= (\alpha I + \hat{V}_Z)^{-1} (V_Z - \hat{V}_Z) \alpha (\alpha I + V_Z)^{-1} \Pi^* \end{aligned}$$

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

Now, we have

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

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

If $\beta > 1$ then the term corresponding to (13) is negligible with respect to (12). If $\beta < 1$, then (12) is negligible with respect to (13).

Now, we turn our attention toward the term (14). 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. Let $\{\lambda_j, \varphi_j\}$ be the eigenvalues and orthonormal eigenfunctions of V_Z .

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

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

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

by Carrasco, Florens, and Renault (2007, Proposition 3.11). Consequently,

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

This concludes the proof of Proposition 2.

Proof of Proposition 5.

We have

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

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

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

$$\begin{aligned}\|\hat{C}_{ZY}^m - \hat{C}_{ZY}\|_{HS}^2 &= \left\| \frac{1}{n} \sum_{i=1}^n z_i^m \langle y_i^m, \cdot \rangle - \frac{1}{n} \sum_{i=1}^n z_i \langle y_i, \cdot \rangle \right\|_{HS}^2 \\ &= \left\| \frac{1}{n} \sum_{i=1}^n \{ (z_i^m - z_i) \langle y_i, \cdot \rangle + z_i^m \langle y_i^m - y_i, \cdot \rangle \} \right\|_{HS}^2 \\ &\leq \frac{2}{n^2} \sum_{i=1}^n \{ \|(z_i^m - z_i) \langle y_i, \cdot \rangle\|_{HS}^2 + \|z_i^m \langle y_i^m - y_i, \cdot \rangle\|_{HS}^2 \}.\end{aligned}$$

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

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

Hence,

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

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

Hence,

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

This concludes the proof of Proposition 5.

Proof of Proposition 6. Using the fact that $\|a + b + c\|_{HS}^2 \leq 3 (\|a\|_{HS}^2 + \|b\|_{HS}^2 + \|c\|_{HS}^2)$, we can evaluate the terms (25), (26), and (27) 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) .

$$\begin{aligned}
& E \left(\|(25)\|_{HS}^2 \mid \mathbf{Z}, \mathbf{W} \right) \\
&= \text{tr} \left\{ \left(\alpha I + \hat{C}_{ZW} \hat{C}_{WZ} \right)^{-1} \hat{C}_{ZW} E \left(\hat{C}_{WU} \hat{C}_{UW} \mid \mathbf{Z}, \mathbf{W} \right) \hat{C}_{WZ} \left(\alpha I + \hat{C}_{ZW} \hat{C}_{WZ} \right)^{-1} \right\}.
\end{aligned}$$

Using

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

we obtain

$$E \left(\|(25)\|_{HS}^2 \right) \leq \frac{C}{n\alpha}$$

for some constant C .

The proof regarding the rates of convergence of (26) and (27) is similar to that of Proposition 2 and is not repeated here.

REFERENCES

- Andersson, J. and J. Lillestol, 2010. Modeling and forecasting electricity consumption by functional data analysis. *The Journal of Energy Markets*, 3:315.
- Antoch, J., L. Prchal, M. De Rosa, and P. Sarda, 2010, Electricity consumption prediction with functional linear regression using spline estimators, *Journal of Applied Statistics*, 37, 2027-2041.
- Bosq, D. *Linear processes in function spaces: Theory and applications*. Springer-Verlag, New York, 2000.
- Breunig, C. and Johannes, J. ,2015, Adaptive estimation of functionals in nonparametric instrumental regression. forthcoming in *Econometric Theory* (arXiv:1109.0961)
- Cardo, H., F. Ferraty, and P. Sarda (2003) Spline estimators for the functional linear model, *Statistica Sinica*, 13, 571-591.
- Carrasco, M. and J. P. Florens, 2000, Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, 16, 797-834.
- Carrasco, M., J. P. Florens, and E. Renault, 2007, Linear Inverse Problems in Structural Econometrics: Estimation based on spectral decomposition and regularization, *Handbook of Econometrics*, Vol. 6B, edited by J.J. Heckman and E.E. Leamer.
- Centorrino, S., 2014, Data-driven selection of the regularization parameter in additive nonparametric instrumental regressions, mimeo, Stony Brook University.
- Centorrino, S. and J.-P. Florens, 2014, Nonparametric instrumental variable estimation of binary response models, mimeo, Stony Brook University.
- Centorrino, S., F. Fève, and J.-P. Florens, 2013, Implementation simulations and bootstrap in nonparametric instrumental variable estimation, mimeo, Stony Brook University.
- Comte and Johannes (2012) Adaptive functional linear regression. *Annals of Statistics*, 40(6), 2765-2797
- Conley, T., L. Hansen, E. Luttmer, J. Scheinkman (1997) Short-Term Interest Rates as Subordinated Diffusions, *The Review of Financial Studies*, 10(3).
- Crambes, C, A. Kneip, and P. Sarda, 2009, Smoothing spline estimators for functional linear regression, *The Annals of Statistics*, 37, 35-72.
- Crambes, C. and A. Mas, 2013, Asymptotics of prediction in functional linear regression with functional outputs, *Bernoulli*, 19, 2627-2651.

- Cuevas, A., M. Febrero, and R. Fraiman, 2002, Linear functional regression: the case of fixe design and functional response. *The Canadian Journal of Statistics*. Vol 30., N.2, 285-300.
- Darolles, S., Y. Fan, J.-P. Florens, and E. Renault, 2011, Nonparametric Instrumental Regression, *Econometrica*, 79, 1541-1565.
- Dauxois, J., A. Pousse, and Y. Romain, 1982, Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference, *Journal of Multivariate Analysis*, 12, 136-154.
- Engl, H., M. Hanke, and A. Neubauer, 2000, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht.
- Eubank, R., 1988, *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York.
- Faruqui, A., S. Sergici, N. Lessem, F. Denton, B. Spencer, and C. King (2013) Report prepared for Ontario Power Authority, *Impact Evaluation of Ontario's Time-of-Use Rates: First Year Analysis*.
- Ferraty, F. and P. Vieu, 2006, *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations*, Springer-Verlag, London.
- Gabrys, R., L. Horvath, and P. Kokoszka, 2010, Tests for Error Correlation in the Functional Linear Model, *Journal of the American Statistical Association*, 105:491, 1113-1125.
- Goldenshluger and Lepski (2011) Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Annals of Statistics*, 39, 1608-1632.
- Hall, P. and J. Horowitz, 2007, Methodology and convergence rates for functional linear regression, *The Annals of Statistics*, 35, 70-91.
- Hansen, L.P. and T. Sargent, 1980. Formulating and estimating dynamic linear rational expectations models, *Journal of Economic Dynamics and Control*, Elsevier, vol. 2(1), pages 7-46.
- Hansen, L.P. and T. Sargent, 1982. Formulating and estimating continuous time rational expectations models, Staff Report 75, Federal Reserve Bank of Minneapolis.
- Hansen, L.P. and T. Sargent, 1983. Identification of continuous time rational expectations models from discrete time data, Staff Report 73, Federal Reserve Bank of Minneapolis.
- Harmann, S., Kokoszka, P., 2010. Weakly dependent functional data. *Ann. Statist.* 38,

1845-1884.

Harmann, S., Kokoszka, P., 2012. Functional time series. In: Rao, C.R., Subba Rao, T. (Eds.), Handbook of Statistics. In: Time Series, vol. 30. Elsevier.

He, G., H. G. Müller, and J. L. Wang, 2000, Extending Correlation and Regression from Multivariate to Functional Data, Asymptotics in Statistics and Probability, 1-14.

Horowitz, J. and S. Lee, 2007, Nonparametric Instrumental Variables Estimation of a Quantile Regression Model, *Econometrica*, Vol. 75, No. 4, 1191-1208.

Horvath, L. and P. Kokoszka, 2012. Inference for Functional Data with Applications, Springer, New York, NY, USA.

Horvath, L., M. Huskova and G. Rice, 2013. Test of independence for functional data, *Journal of Multivariate Analysis*, Elsevier, vol. 117(C), pages 100-119.

Horvath, L., P. Kokoszka and G. Rice, 2014. Testing stationarity of functional time series, *Journal of Econometrics*, Elsevier, vol. 179(1), pages 66-82.

Kennan, J. (1979), The Estimation of Partial Adjustment Models with Rational Expectations, *Econometrica*, vol.47

Kress, R. (1999), *Linear Integral Equations*, Springer.

Liebl, D. Modeling and forecasting electricity spot prices: A functional data perspective. *Ann. Appl. Stat.* 7 (2013), no. 3, 1562–1592.

McLaughlin, L., Jia, L., Yu, Z. and Tong, L., 2011, Thermal Dynamic for Home Energy Management: A Case Study. Cornell Univ., Ithaca, NY, USA, Tech. Rep. ACSP-TR-10-11-01.

Muth, J. (1961). "Rational Expectations and the Theory of Price Movements", *Econometrica* 29, pp. 315-335.

Patrick, R. and F. Wolak. 1997. Estimating customer-level demand for electricity under real time pricing. Document available from <http://www.stanford.edu/~wolak>.

Ramsay, J.O. and B.W. Silverman, 2005, *Functional Data Analysis*, 2nd edition, Springer, New York.

Rudin, W., 1991. *Functional Analysis*, 2nd Edition, McGraw Hill.

Vilenkin, N.I.A, 1968. *Special functions and the theory of group representations*, American Mathematical Society.

Weron, R. 2006. "Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach," HSC Books, Hugo Steinhaus Center, Wroclaw University of Technology.

Statistics Canada, 2008, EnviroStats "Thermostat use in Canadian homes". Fall 2008

Vol.2, no.3, Catalogue no.16-002-X.

Yao, F., H. G. Müller, and J. L. Wang, 2005, Functional Linear Regression Analysis for Longitudinal Data, *The Annals of Statistics*, Vol. 33, N.6, 2873-2903.

Yu, Z., McLaughlin, L., Jia, L., Murphy-Hoye, M.C., Pratt, A. and Tong, L., 2012, Modeling and Stochastic Control for Home Energy Management. *Proc. IEEE PES Gen. Meet.*, July.