Functional linear regression with functional response*

David Benatia
Université de Montréal

Marine Carrasco
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 L2 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, medecine, 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 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 of values 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) $i = 1, 2, ..., n$ 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 unknow 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 $cov\left(Z, U\right) = 0$. This assumption will be relaxed in Section 7.

For illustration, consider the following example

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

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

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

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

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

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

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

Using (1), we have

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

Hence, we have the following relationships:

$$C_{YZ} = \Pi V_Z, \tag{2}$$
$$C_{ZY} = V_Z \Pi^* \tag{3}$$

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

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

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

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

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

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

for $\psi \in \mathcal{E}$. An estimator of Π^* can not be obtained directly by solving $\hat{C}_{ZY} = \widehat{V}_Z\Pi^*$ because the initial equation $C_{ZY} = V_Z\Pi^*$ is an ill-posed problem in the sense that V_Z is invertible only on a subset of $\mathcal{E}$ and its inverse is not continuous. Note that $\hat{V}_Z$ has finite rank equal to n and hence is not invertible. A Moore-Penrose generalized inverse 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

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

and that of Π is defined by

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

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

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

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

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

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

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

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

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

Note that

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

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

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

It follows from (6) that $B^* \varphi_j = \frac{1}{n} \sum_i \left\langle f_i, \varphi_j \right\rangle z_i$ 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

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

and

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

It follows that

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

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

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

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

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

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

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

In summary, a necessary and sufficient condition for the identification of Π is that V_Z is injective.

2.4 Computation of the estimator

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

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

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

$$\frac{1}{n}\sum_{i=1}^n \langle z_l, z_i\rangle \langle y_i, \psi\rangle = \alpha\left\langle z_l, \hat{\Pi}_\alpha^*\psi\right\rangle + \frac{1}{n}\sum_{i=1}^n \langle z_l, z_i\rangle \left\langle z_i, \hat{\Pi}_\alpha^*\psi\right\rangle, \; l = 1, 2, ..., n, \tag{8}$$

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

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

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

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

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

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

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, $i = 1, 2, ..., n$, we obtain

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

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 $E(U_i) = 0$, $cov(U_i, U_j | Z_1, Z_2, ..., Z_n) = 0$ 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 $\sum_j \left\langle K\phi_j, \phi_j \right\rangle < \infty$ for any basis $\left(\phi_j\right)$. If K is self-adjoint positive definite, it is equivalent to say that the sum of the eigenvalues of K is finite. Given V_U is a covariance operator, V_U is trace-class if and only if $E\left(\|U_i\|^2\right) < \infty$.

The notation $\|\|_{HS}$ refers to the Hilbert-Schmidt norm of operators. An operator K is Hilbert-Schmidt (noted HS) if $\|K\|_{HS}^2 \equiv \sum_j \left\langle K\phi_j, K\phi_j \right\rangle < \infty$ for any basis $\left(\phi_j\right)$. 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 $E\left(\|V_i\|^4\right) < \infty$, 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

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

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

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

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

$$\begin{aligned}
\hat{\Pi}_\alpha^* - \Pi^* &= \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZY} - \Pi^* \\
&= \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZU} \qquad (12)\\
&\quad + \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{V}_Z \Pi^* - (\alpha I + V_Z)^{-1} V_Z \Pi^* \qquad (13)\\
&\quad + (\alpha I + V_Z)^{-1} V_Z \Pi^* - \Pi^*. \qquad (14)
\end{aligned}$$

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 $E\left(U_i | Z_i\right) = 0$.

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

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

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, $\sqrt{n}\left(\hat{V}_Z - V_Z\right)$ 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

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

As n goes to infinity, $\hat{\Pi}^*_\alpha - \Pi^*_\alpha$ converges to zero and is $\sqrt{n}-$asymptotically normal. The first two terms of the r.h.s are $O_p\left(1/\sqrt{n}\right)$ and will affect the asymptotic distribution. This distribution is not simple.

We 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,

$$\sqrt{n}\left(\hat{C}_{ZY} - \hat{V}_Z \Pi^*_0\right) \xrightarrow{d} \mathcal{N}\left(0, K_{ZU}\right)$$

where

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

and $\left(x \otimes y\right)(f) = \langle x, f \rangle y$ and $\left(\Pi_1 \widetilde{\otimes} \Pi_2\right) T = \langle T, \Pi^*_1 \rangle_{\mathcal{H}} \Pi_2$ (see Dauxois, Pousse, and Romain, 1982)

Let $\left\{\phi_j : j = 1, 2, ..., q\right\}$ 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:

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

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

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

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

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

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

where $\hat{\Pi}^{(-i)}_\alpha$ has been computed using all observations except for the ith 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), $i = 1, 2, ..., n$ 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 $i = 1, 2, ..., n$. 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.

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

Proposition 3 *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, $E((Y - \Pi Z)\langle W, .\rangle) = 0$. It follows that

$$C_{YW} = \Pi C_{ZW} \tag{15}$$

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

$$C_{WY} = C_{WZ}\Pi^* \tag{16}$$

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

We need the following identification conditions:

Assumption 8. C_{WZ} is injective.

Under this assumption, Π is uniquely defined from (15). To see this, assume that there are two solutions Π_1 and Π_2 to (15). It follows that $(\Pi_1 - \Pi_2)\, C_{ZW} = 0$ or equivalently $C_{WZ}\left(\Pi_1^* - \Pi_2^*\right) = 0$. Hence the range of $\left(\Pi_1^* - \Pi_2^*\right)$ belongs to the null space of C_{WZ}. However, under Assumption 6, the null space of C_{WZ} is reduced to zero and thus the range of $\left(\Pi_1^* - \Pi_2^*\right)$ is equal to zero. It follows that $\Pi_1^*\varphi - \Pi_2^*\varphi = 0$ for all φ, hence $\Pi_1^* = \Pi_2^*$.

To construct an estimator of Π^*, we first apply the operator C_{ZW} on the l.h.s and r.h.s of Equation (16) to obtain

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

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

$$\hat{\Pi}_\alpha^* = \left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW}\hat{C}_{WY}. \tag{17}$$

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 $\hat{\Pi}_\alpha^*$ is practice. From (17), 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

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

Taking the inner product with z_l yields n equations

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

with n unknowns $\left\langle z_j,\hat{\Pi}_\alpha^*\psi\right\rangle$, $j=1,2,...,n$. 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:

$$\begin{aligned}&\hat{\Pi}_\alpha^*-\Pi^*\\ =\;&\left(\alpha I+\hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW}\hat{C}_{WY}-\Pi^* &(18)\\ =\;&\left(\alpha I+\hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW}\hat{C}_{WU} &(19)\\ &+\left(\alpha I+\hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW}\hat{C}_{WZ}\Pi^*-(\alpha I+C_{ZW}C_{WZ})^{-1}C_{ZW}C_{WZ}\Pi^* &(20)\\ &+(\alpha I+C_{ZW}C_{WZ})^{-1}C_{ZW}C_{WZ}\Pi^*-\Pi^*. &(21)\end{aligned}$$

Proposition 4 *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 to $L^2[0,1]$ to $L^2[0,1]$

with kernel $\pi(s,t) = 1 - |s-t|^2$.[1] We consider an Ornstein-Uhlenbeck process with zero mean and mean reversion rate equal to one to represent the error function. It is described by the differential equation $dU(s) = -U(s)ds + \sigma_u dG_u(s)$, for $s \in [0,1]$ and where G_u is a Wiener process and σ_u denotes the standard deviation of its increments dG_u. Note that this error function is stationary.

We study the model

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

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

8.1 Exogenous predictor functions

We consider the design function

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

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

The numerical simulation is performed as follows:

1. Construct both a pseudo-continuous interval of $[0,1]$, denoted $\mathcal{T}$, consisting of 1000 equally-spaced discrete steps, and a discretized interval of $[0,1]$, denoted $\tilde{\mathcal{T}}$, consisting of only 100 equally-spaced discrete steps.

2. Generate n predictor functions $z_i(t)$ and error functions $u_i(s)$, where $t, s \in \mathcal{T}$ so as to obtain pseudo-continuous functions.

3. Generate the n response functions $y_i(s)$ using the specified model where $s \in \mathcal{T}$.

4. Generate the sample of n discretized pairs of functions $(\tilde{z}_i, \tilde{y}_i)$ by extracting the corresponding values of the pairs (z_i, y_i) for all $t, s \in \tilde{\mathcal{T}}$.

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

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 $\|\hat{\Pi}_\alpha - \Pi\|^2_{HS} = \int\limits_{\tilde{\mathcal{T}}}\int\limits_{\tilde{\mathcal{T}}}(\hat{\pi}_\alpha(s,t) - \pi(s,t))^2 dtds$ 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).[2] 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 $\|\Pi - \Pi_\alpha\|^2_{HS}$, which is a measure of the squared bias introduced by the regularization method.[3] The last two columns of Table 1 report the true global (R^2) and extended local ($\tilde{R}^2$) functional coefficients of determination, defined as

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

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

[2] In 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.

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

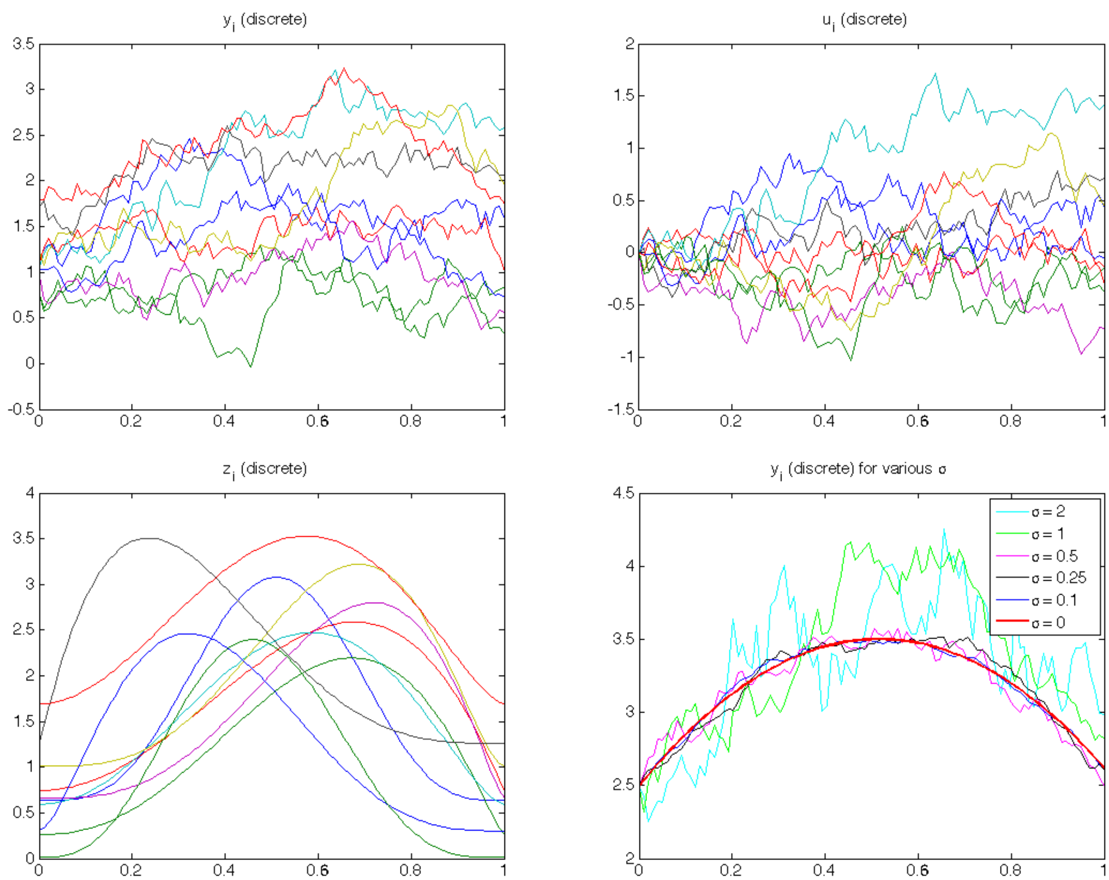

Figure 1: Examples of simulated functions (top left: discretized y_i; top right: discretized u_i for $\sigma_U = 1$, bottom left: discretized z_i, bottom right: a single y_i for various σ_u).

which are directly related to those proposed in Yao, Muller and Wang (2005).[4]

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

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

Note: Standard deviations are reported in parentheses.

Simulations results are in line with the theoretical results. We observe that, for a fixed α, the MSE converges to the squared bias and its variance shrinks 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.

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 (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.

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

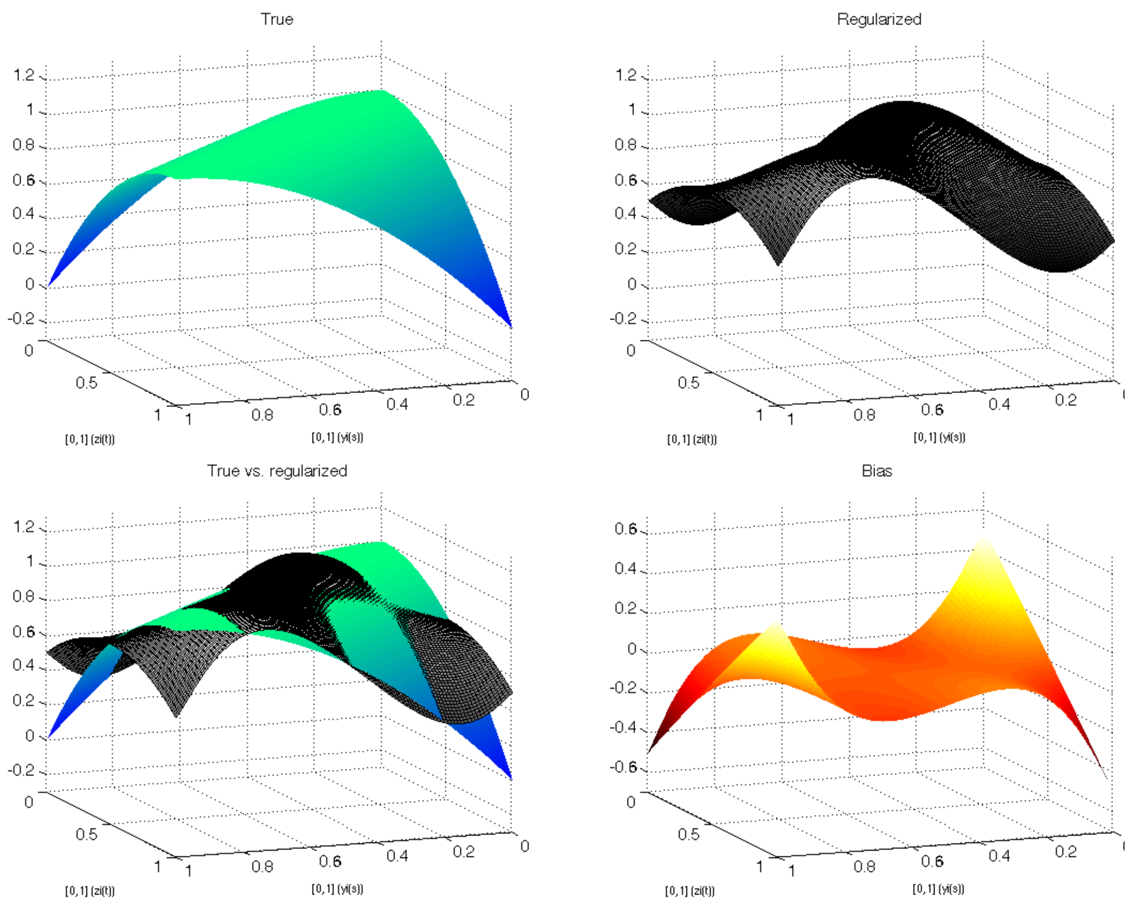

Figure 2: True kernel vs. regularized kernel (top left: True; top right: Regularized, bottom left: True vs. regularized, bottom right: Bias).

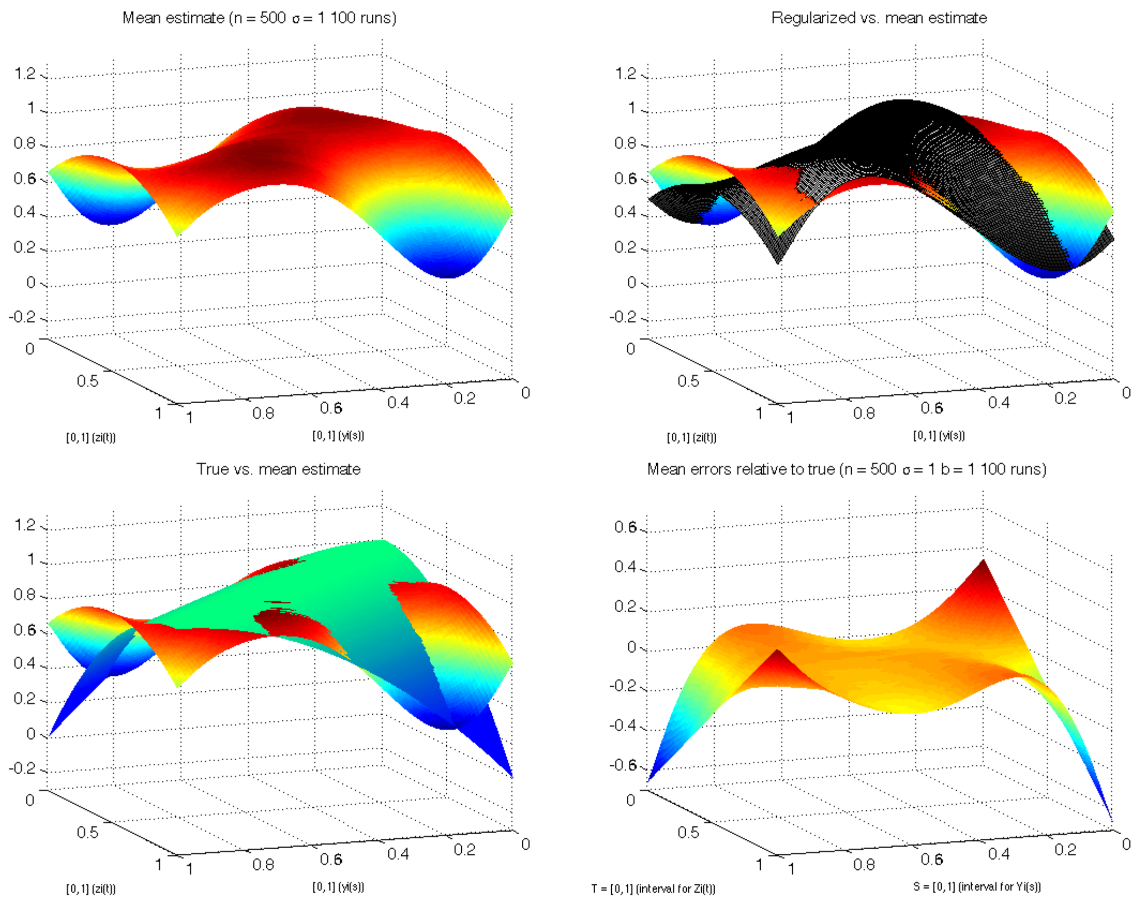

Figure 3: True kernel vs. mean estimate (100 runs with $n = 500$, $\sigma_u = 1$) (top left: Mean estimate, top right: Regularized vs. mean estimate, bottom left: True vs. mean estimate, bottom right: Mean errors)

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

8.2 Endogenous predictor functions

We consider the design function

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

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

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

for $t \in [0,1]$, $\alpha_i, \beta_i \sim iid\, U[2,5]$ and $\eta_i \sim iid\, N(0,1)$, for all $i = 1, ..., n$. 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 $\int\limits_{\mathcal{S}} \mathrm{var}(Y(s))ds$ is unchanged as b varies.[5] 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] This assumption allows to keep the variance of Y stable when varying instrument strength. It implies $c = \sqrt{1 + (1-b^2)\frac{\int\limits_{\mathcal{S}}(\mathrm{var}(\Pi W(s))ds}{\int\limits_{\mathcal{S}}\mathrm{var}(\Pi\varepsilon(s))ds}}$.

5. Generate the sample of n discretized pairs of functions $(\tilde{w}_i, \tilde{z}_i, \tilde{y}_i)$ by extracting the corresponding values of the pairs (w_i, z_i, y_i) for all $t, s \in \tilde{\mathcal{T}}$.

6. Estimate Π using the regularization method on the sample of n triplets of functions $(\tilde{w}_i, \tilde{z}_i, \tilde{y}_i)$ and a fixed smoothing parameter $\alpha = .01$.

7. Repeat steps 2-5 100 times and calculate the MSE by averaging the quantities $\|\hat{\Pi}_\alpha - \Pi\|_{HS}^2 = \int_{\tilde{\mathcal{T}}}\int_{\tilde{\mathcal{T}}} (\hat{\pi}_\alpha(s,t) - \pi(s,t))^2 dtds$ over all repetitions.

Table 2 reports the MSE for 4 different sample sizes ($n = 50, 100, 500, 1000$) and 4 values of b when estimating the model without accounting for the endogeneity of Z. 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 $\int_{\mathcal{S}} \text{var}(Y(s))ds$ 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.	Sample sizes				Squared bias	Coef. of d.	
	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$\lVert\Pi - \Pi_\alpha\rVert^2_{HS}$	R^2	R^2_{FS}
$b = 0.25$	.2383	.1710	.0752	.0542	.0060	.0175	.0246
$(c = 2.3)$	(.2019)	(.1422)	(.0779)	(.0209)			
$b = 0.5$	.1040	.0619	.0315	.0276	.0027	.0737	.1092
$(c = 1.96)$	(.0859)	(.0349)	(.0099)	(.0053)			
$b = 0.75$	.0682	.0444	.0242	.0216	.0011	.1767	.2683
$(c = 1.55)$	(.0364)	(.0203)	(.0044)	(.0028)			
$b = 1$	.0466	.0330	.0211	.0199	.0006	.3287	.5048
$(c = 1)$	(.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.

A Appendix. Proofs

Proof of Proposition 1.

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

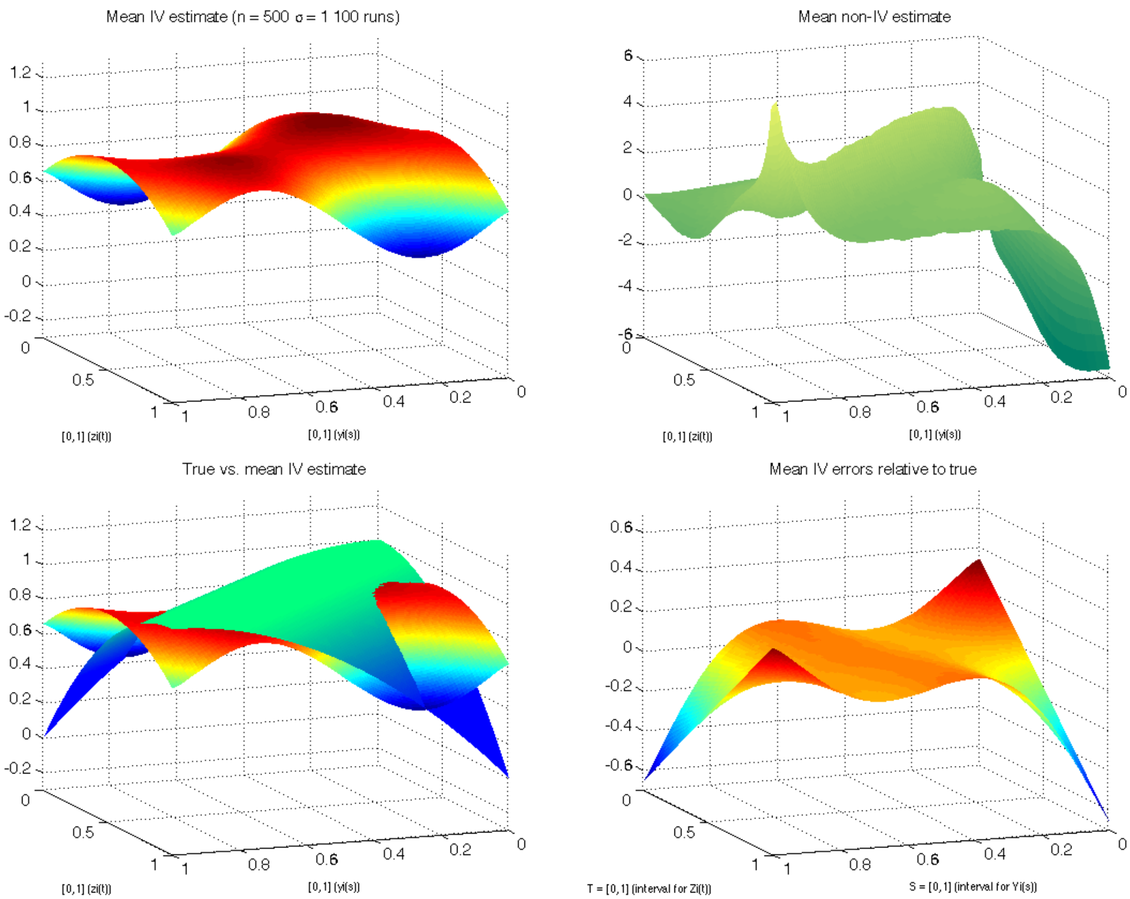

Figure 4: True kernel vs. mean IV estimate (100 runs with $n = 500$, $\sigma_u = 1$ and $b = 0.75$) (top left: Mean estimated IV; top right: Mean estimated non-IV, bottom left: True vs. mean IV estimate, bottom right: Mean IV errors)

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 5 *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 5.

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

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

Lemma 6 *Let A be a random operator from $\mathcal{E}$ to $\mathcal{F}$.*

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

Proof of Lemma 6. We have

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

We turn to the proof of Proposition 2. Applying Lemma 5 on the decomposition (12), (13), and (14), we have

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

We study the first term of the r.h.s. By Lemma 6,

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

Note that

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

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

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

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

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

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

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

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

Hence,

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

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

Now, we have

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

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

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

Now, we turn our attention toward the term (14). We have

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

by Assumption 4. Let $\left\{\lambda_j,\varphi_j\right\}$ be the eigenvalues and orthonormal eigenfunctions of V_Z.

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

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

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

by Carrasco, Florens, and Renault (2007, Proposition 3.11). Consequently, $\left\| \alpha \left(\alpha I + V_Z\right)^{-1} V_Z^{\beta/2} R \right\|_{HS}^2 = O\left(\alpha^{\beta \wedge 2}\right)$.

This concludes the proof of Proposition 2.

Proof of Proposition 3.

We have

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

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

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

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

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

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

Hence,

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

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

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 **3**.

Proof of Proposition 4. Using the fact that $\|a+b+c\|_{HS}^2 \leq 3\left(\|a\|_{HS}^2+\|b\|_{HS}^2+\|c\|_{HS}^2\right)$, we can evaluate the terms (19), (20), and (21) separately. The proof follows closely that of Proposition 2. Let $\mathbf{Z}$ and $\mathbf{W}$ be the sets $(Z_1, Z_2, ..., Z_n)$ and $(W_1, W_2, ..., W_n)$.

$$\begin{aligned}&E\left(\|(19)\|_{HS}^2|\mathbf{Z},\mathbf{W}\right) \\ =\ & tr\left\{\left(\alpha I+\hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW}E\left(\hat{C}_{WU}\hat{C}_{UW}|\mathbf{Z},\mathbf{W}\right)\hat{C}_{WZ}\left(\alpha I+\hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\right\}.\end{aligned}$$

Using

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

we obtain

$$E\left(\|(19)\|_{HS}^2\right) \leq \frac{C}{n\alpha}$$

for some constant C.

The proof regarding the rates of convergence of (20) and (21) is similar to that of Proposition 2 and is not repeated here.

REFERENCES

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.

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.

Ferraty, F. and P. Vieu, 2006, Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations, Springer-Verlag, London.

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.

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.

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

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

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.