Functional instrumental variable regression with an application to estimating the impact of immigration on native wages

Dakyung Seong\textsuperscript{a\textdagger} Won-Ki Seo\textsuperscript{a}

\textsuperscript{a} School of Economics, University of Sydney, Australia

Abstract

Functional linear regression gets its popularity as a statistical tool to study the relationship between function-valued response and exogenous explanatory variables. However, in practice, it is hard to expect that the explanatory variables of interest are perfectly exogenous, due to, for example, the presence of omitted variables and measurement error. Despite its empirical relevance, it was not until recently that this issue of endogeneity was studied in the literature on functional regression, and the development in this direction does not seem to sufficiently meet practitioners’ needs; for example, this issue has been discussed with paying particular attention on consistent estimation and thus the distributional properties of the proposed estimators still remain to be further explored. To fill this gap, this paper proposes new consistent FPCA-based instrumental variable estimators and develops their asymptotic properties in detail. We also provide a novel test for examining if various characteristics of the response variable depend on the explanatory variable in our model. Simulation experiments under a wide range of settings show that the proposed estimators and test perform considerably well. We apply our methodology to estimate the impact of immigration on native wages.

\textbf{JEL codes:} C13, C36

\textbf{Keywords:} Functional regression, Instrumental variables, Endogeneity, Regularization

1 Introduction

The recent developments in data collection and storage technologies ignite studies on how to use more complicated observations such as curves, probability density functions, or images. This area of study, commonly called functional data analysis, has become popular in statistics, and researchers in various fields of study, including economics, have been benefited from advances in this area. In particular, for practitioners who are interested in studying the relationship between two or more such variables, functional linear models are of central importance, and crucial contributions on this

\textsuperscript{\textdagger}The authors are thankful to Morten Ø. Nielsen and seminar participants at the University of Sydney for their helpful comments. Data and R code to replicate the empirical results in Tables 3 and 4 are available on the authors’ websites.

\textsuperscript{†}Corresponding author. Address: School of Economics, University of Sydney, Camperdown, 2006, NSW, Australia. E-mail addresses: dakyung.seong@sydney.edu.au (D. Seong), won-ki.seo@sydney.edu.au (W.-K. Seo)
The existing statistical approaches for estimating the functional linear model, including those proposed in the aforementioned literature, are mostly established under the assumption that the explanatory variable of interest is exogenous, meaning that it is uncorrelated with the regression error. However, this assumption is not likely to hold in practice; that is, explanatory variables are often *endogenous*. The issue of endogeneity may be especially relevant in the context of functional linear models since functional observations used in analysis are almost always constructed by smoothing their discrete realizations that are not often densely observed (see, e.g., Yao et al., 2005). If this being the case, the functional observations inevitably involve small or large measurement errors, which in turn leads to violation of the exogeneity condition at least to some degree (see, e.g., Section 2 and Section 6.2). This issue may hamper practitioners from applying the functional linear model.

If the size of the literature on functional linear regression under the exogeneity condition provides any indication, the developments that have been made so far to deal with endogeneity do not seem to sufficiently satisfy practitioners’ needs. For example, although a few papers, such as Benatia et al. (2017) and Chen et al. (2022), study the issue of endogeneity in the functional linear model, not much is known about asymptotic distributions of their estimators and how to implement statistical inference on the parameter of interest; this may limit the practical applicability of the functional endogenous linear model. We will fill this gap to some extent by providing new estimators and inferential methods based on the asymptotic properties of our estimators, and this is a crucial point where the present paper is differentiated from the existing ones concerning the issue of endogeneity.

Specifically, this paper provides new estimation results for the functional endogenous linear model based on (i) the functional principal component analysis (FPCA) and (ii) the instrumental variable (IV) approach. The former has been widely adopted by researchers dealing with functional data in practice (see e.g., Ramsay and Silverman, 2005; Shang, 2014), and the latter has also been widely adopted in order to address endogeneity not only in the conventional Euclidean space setting (see e.g., Bekker, 1994; Chao and Swanson, 2005; Newey and Windmeijer, 2009), but also in the setting involving functional observations (see, e.g., Carrasco, 2012; Florens and Van Bellegem, 2015; Benatia et al., 2017; Chen et al., 2022; Babii, 2021). However, the application of the FPCA to the functional endogenous linear model has not been fully explored.

We consider the case where the response variable $y_t$, explanatory variable $x_t$, and instrumental variable $z_t$ are all function-valued; of course, with slight modifications, our results to be subsequently given can be adjusted for the case where $y_t$ is scalar- or vector-valued. Unlike in the most of the aforementioned papers, we do not require the variables of interest to be iid, but allow those to exhibit some weak dependence so that our methodology can be applied to various empirical examples. Given that many functional observations which have been considered in the literature toward applications in fields of energy, environmental and financial economics tend to allow time series dependence, this extension may be attractive to practitioners. Among the papers mentioned in the
preceding paragraph, the study by Chen et al. (2022) is most closely related to the present paper in the sense that they consider FPCA-based consistent estimation of function-on-function regression models with endogeneity introduced by measurement errors; Benatia et al. (2017) earlier considered a similar model and proposed a consistent estimation method, but their theoretical results are obtained from a quite different theoretical methodology (ridge-type regularization applied in an infinite dimensional setting). As mentioned above, we complement these studies by providing new FPCA-based estimators and in-depth discussion on their asymptotic properties.

Technically, we view the function-valued variables of interest as random variables taking values in a Hilbert space of square-integrable functions, and then propose our FPCA-based functional IV estimator (FIVE). As is well known in the literature, estimation of a model involving function-valued random variables is not straightforward because some important sample operators, such as the covariance of such a random variable, are not invertible over the entire Hilbert space(s). We circumvent this issue by employing a rank-regularized inverse of such an operator, and this is the point where we make use of the FPCA. The reason why we focus on this regularization scheme comes not from its theoretical superiority, but merely from its popularity in the literature. Other schemes such as ridge-type regularization (e.g., Florens and Van Bellegem, 2015; Benatia et al., 2017) may be alternatively adopted, and are expected to have their own merits (see Remark 3).

This paper studies in depth the asymptotic properties of the proposed estimator. It is first shown that, under some mild conditions on the data generating process of \( \{y_t, x_t, z_t\}_{t \geq 1} \), the FIVE achieves the weak (convergence in probability) and strong (almost sure convergence) consistencies as long as the regularization parameter, which is introduced for a rank-regularized inverse of a certain sample operator used to construct the FIVE, decays to zero at an appropriate rate. We then establish more detailed asymptotic properties of the FIVE under some nonrestrictive assumptions on the eigenstructure of the cross-covariance operator of the explanatory variable \( x_t \) and the IV \( z_t \). By doing so, we can see how the cross-covariance structure of \( x_t \) and \( z_t \) and the choice of the regularization parameter affect the convergence of the FIVE toward its true counterpart. In addition to these results, we also show that the FIVE is asymptotically normal in a pointwise sense if it is centered at a certain operator that is slightly biased from the true parameter of interest; moreover, if certain additional conditions are satisfied, such a bias becomes asymptotically negligible and thus, in this case, the FIVE centered at the true parameter becomes asymptotically normal. The asymptotic normality results given in this paper are quite different from similar results given in a finite dimensional setting in the sense that the convergence rate is (i) possibly random and (ii) not uniformly given over the entire Hilbert space on which our estimators are defined. This result implies that the proposed estimators do not weakly converge to any elements in the usual operator topology, which is similar to what Mas (2007) earlier found in the context of functional autoregressive (AR) models of order 1.

Based on our study of the FIVE, we also propose a different but closely related estimator, called the functional two-stage least square estimator (F2SLSE) and obtain its asymptotic properties in a similar manner. We discuss how our estimators and their asymptotic properties can be used to
implement usual statistical inference on the parameter of interest, and moreover, propose a novel significance test for examining if the considered explanatory variable affects various characteristics of the response variable; the proposed test is easy-to-implement, and its asymptotic null distribution can be easily approximated via Monte Carlo simulations.

To see how the asymptotic properties of our estimators and test are revealed in finite samples, we implement Monte Carlo experiments under various simulation designs. The simulation results are quite satisfactory; overall, it seems that our estimators can be good alternatives or sometimes complements to some existing estimators that are closely related to ours. In addition, our significance test is shown to have excellent size control and power in the considered simulation designs.

As an empirical illustration, we study the impact of immigration on native wages. To be specific, we employ a model that is similar to those considered by, e.g., Dustmann et al. (2013) and Sharpe and Bollinger (2020). The previous literature in this area, including Ottaviano and Peri (2012), Card (2009), and the aforementioned articles, show that an inflow of immigrants differently affects native wages depending on levels of skill (captured by, for examples, education and experience) of both natives and immigrants. We, in the present paper, investigate such heterogeneous effects using our functional linear model, which is initiated by viewing both the labor supply and the native wage as functions of a certain measure of occupation-specific skills (will be detailed in Section 7). This approach has a couple of advantages compared to that taken in the earlier literature. For example, in our functional setting, an inflow of immigrants in an occupation requiring a particular skill level is allowed to have an impact on the wages of natives in other occupations. Furthermore, in the previous literature, workers of various levels of skill are often classified into a few skill groups before analysis, which is necessitated to reduce dimensionality of the considered model (see Example 1 and Section 7). However, such a pre-classification, which may affect estimation results and the interpretation of those, is not required in our approach. Using the methodology developed in this paper, we find evidence supporting the presence of heterogeneous effects of immigration.

The remainder of this paper is organized as follows. Section 2 introduces a functional endogenous linear model and provides some motivating examples. In Sections 3 and 4, we define our estimators and discuss their asymptotic properties. The significant test is proposed in Section 5 in which its asymptotic property is also established. Section 6 reports Monte Carlo simulation results and Section 7 details our empirical example. Section 8 concludes.

2 Functional endogenous linear model

2.1 Endogeneity and motivating examples

We suppose that random functions $y_t$, $x_t$ and $u_t$ satisfy the following relationship: for $t \geq 1$,

$$y_t = c_y + Ax_t + u_t,$$

(2.1)

where $c_y$ is the intercept function and $A$ is a linear operator satisfying certain conditions to be clarified. In (2.1), the response variable $y_t$, the explanatory variable $x_t$ and the error variable $u_t$ are
random functions, which are technically understood as random variables taking values in separable Hilbert spaces. Appendix A1 briefly introduces the definitions of a Hilbert-valued random variable $X$, its expectation (denoted $\mathbb{E}[X]$), covariance operator (denoted $\mathcal{C}_X := \mathbb{E}[(X - \mathbb{E}[X]) \otimes (X - \mathbb{E}[X])]$), and cross-covariance operator with another Hilbert-valued random variable $Y$ (denoted $\mathcal{C}_{XY} := \mathbb{E}[(X - \mathbb{E}[X]) \otimes (Y - \mathbb{E}[Y])]$), where $\otimes$ signifies the tensor product defined in (A1.1) in Appendix A1. These notions are natural generalizations of a random vector and its expectation, covariance matrix, and cross-covariance matrix with another random vector, respectively.

We say that the explanatory variable $x_t$ is endogenous if it is correlated with the error $u_t$, i.e., the cross-covariance of $x_t$ and $u_t$, given by the operator $\mathbb{E}[(x_t - \mathbb{E}[x_t]) \otimes (u_t - \mathbb{E}[u_t])]$, is nonzero. The present paper focuses on estimation and inference of the functional linear model in the presence of endogeneity. Below we provide specific and practical examples that motivate this model of interest.

**Example 1** (Effects of immigration on the native labor market). We briefly illustrate our empirical example in Section 7. To examine the effect of immigration on native wages, we let $w_t(s)$ (resp. $h_t(s)$) be the occupation-specific log wage of natives (resp. the occupation-specific share of immigrants) at time $t$, where $s \in [0,1]$ identifies occupations by quantifying some work-related abilities that are required by each occupation (this will be detailed in Section 7). In many applications, researchers classify occupations into $n$ groups, $S_{1(n)}, \ldots, S_{n(n)}$, according to the value of $s$; for example, we may consider three groups of occupations requiring low, mid, and high levels of skill by partitioning $[0,1]$ into $[0,1/3]$, $(1/3, 2/3]$, and $(2/3, 1]$. Then, the following model may be employed to examine the effect of immigration for each group $S_{i(n)}$: $w_{t,i(n)} = c_{i(n)} + \beta_{i(n)} h_{t,i(n)} + u_{t,i(n)}$, where $w_{t,i(n)}$ (resp. $h_{t,i(n)}$) is the within average of $w_t(s)$ (resp. $h_t(s)$) for the group $S_{i(n)}$ at time $t$. The parameter $c_{i(n)}$ represents the group-specific fixed effect and $\beta_{i(n)}$ is the group-specific effect of immigration. This model can be generalized in two ways. First, as suggested by Peri and Sparber (2009) and Llull (2018), an inflow of immigrants in a particular occupation requiring skill level $s \in S_{i(n)}$ could have an impact on the labor market outcomes of native workers in other occupations, and thus, the above model can be generalized to accommodate such an effect. Second, instead of pre-classifying workers into a few groups, we may treat $w_t(s)$ (resp. $h_t(s)$) for different values of $s \in [0,1]$ as discrete realizations of the curve of skill-specific native wage $w_t$ (resp. the curve of skill-specific share of immigrants $h_t$) that is continuous in $s \in [0,1]$. Combining these views, we extend the above model as follows:

$$w_t = c_y + \mathcal{A} h_t + u_t,$$

where $c_y$ is the intercept function and the linear operator $\mathcal{A}$ is the parameter of interest. This generalization seems to be appealing, but, unfortunately, estimating this model involves challenging issues in practice. One of the main concerns, we want to stress here, is related to the endogenous occupational adjustment of workers (see, e.g., Llull, 2018), which results in violation of the exogeneity condition; this will be further discussed in Section 7.

**Example 2** (Functional AR model with measurement errors). The functional AR model has been used in many applications involving functional data. We, in this example, consider the functional
AR model where each observation is contaminated by measurement errors; this may be understood as a special case of the model considered in Chen et al. (2022). A natural example can be found in some recent literature on forecasting of probability density functions; see e.g., Kokoszka et al. (2019). Since true probability density functions are not observable in practice, they need to be replaced by appropriate nonparametric estimates that involve estimation errors. As we show in Section 6.2 via Monte Carlo study, this error could make estimation results of some existing estimators less accurate. Beyond this specific case, it seems to be quite common in practice that the true functional realization \( y_t^0 \) cannot be observed and thus has to be replaced by an estimate \( y_t \), obtained by smoothing its discrete realizations. In these cases, it may be natural to assume that \( y_t \) contains a measurement error \( e_t \), i.e., \( y_t = y_t^0 + e_t \). If a sequence of functions \( \{y_t^0\}_{t \geq 1} \) satisfies the stationary AR law of motion given by \( y_t^0 = Ay_{t-1}^0 + \epsilon_t \) for \( t \geq 1 \), with \( \mathbb{E}[\epsilon_t] = 0 \) and \( \mathbb{E}[y_{t-1}^0 \otimes \epsilon_t] = 0 \), we have

\[
y_t = Ay_{t-1} + u_t, \quad \text{where} \quad u_t = e_t - \varepsilon_{t-1} + \epsilon_t.
\]

In this case, \( \mathbb{E}[y_{t-1} \otimes u_t] \neq 0 \) in general, and hence \( y_{t-1} \) is endogenous. To see this in detail, we assume for simplicity that \( e_t \) is uncorrelated with \( \varepsilon_{\ell} \) for every \( \ell \neq t \), and also uncorrelated with \( y_{\ell}^0 \) and \( \varepsilon_{\ell} \) for every \( \ell \); this may not be restrictive in practice, and an analogous assumption was employed by Chen et al. (2022) in a similar context. Then, it can be shown that \( \mathbb{E}[y_{t-1} \otimes u_t] = A\mathbb{E}[e_{t-1} \otimes e_{t-1}] \), which is a nonzero operator in general.

It is expected from Example 2 that endogeneity can arise in many practical applications of the functional linear model, where \( x_t \) is incompletely observed. In such a case, the exogeneity condition is likely violated. This type of endogeneity, induced by error-contaminated explanatory variables, was earlier studied by Chen et al. (2022) (see Remark 2 for more details).

As well expected from the literature on the standard linear simultaneous equation model, endogeneity should be properly addressed for consistent estimation of the regression operator. A widely used strategy to do this is the IV approach, which is pursued in our Hilbert space setting.

### 2.2 Model, some simplifying assumptions, and notation

This paper concerns the case where the response variable \( y_t \), the endogenous explanatory variable \( x_t \), and the IV \( z_t \) (to be detailed shortly) are infinite-dimensional random variables taking values in separable Hilbert spaces. We hereafter conveniently assume that all of \( y_t, x_t \) and \( z_t \) take values in \( \mathcal{H} \), the Hilbert space of square-integrable functions defined on the unit interval \([0, 1]\), where the inner product \( \langle \cdot, \cdot \rangle \) is defined by \( \langle \zeta_1, \zeta_2 \rangle = \int_0^1 \zeta_1(s)\zeta_2(s)ds \), for \( \zeta_1, \zeta_2 \in \mathcal{H} \) and \( \| \cdot \| = \langle \cdot, \cdot \rangle^{1/2} \) defines the norm of \( \mathcal{H} \). This setup in fact encompasses a more general scenario where \( y_t, x_t \) and \( z_t \) take values in different separable Hilbert spaces of infinite dimension, say \( \mathcal{H}_y, \mathcal{H}_x \), and \( \mathcal{H}_z \). This is because that these spaces are all isomorphic to \( \mathcal{H} \) (see e.g., Corollary 5.5 in Conway, 1994), and thus there is no loss of generality by assuming that \( \mathcal{H}_y = \mathcal{H}_x = \mathcal{H}_z = \mathcal{H} \). We further assume for convenience that \( y_t, x_t \) and \( z_t \) have zero means, i.e., \( \mathbb{E}[y_t] = \mathbb{E}[x_t] = \mathbb{E}[z_t] = 0 \); this assumption naturally makes \( c_y \) in (2.1) be suppressed to zero. The extension to the case where the means are unknown and needed to be estimated is straightforward. After adopting all such simplifying assumptions, the
functional endogenous linear model, which will subsequently be considered, is given as follows: for a linear operator $A : \mathcal{H} \mapsto \mathcal{H}$,

$$y_t = Ax_t + u_t, \quad \text{where } \mathbb{E}[x_t \otimes u_t] \neq 0 \quad \text{and} \quad \mathbb{E}[u_t] = 0. \quad (2.2)$$

As mentioned earlier, our approach to address the endogeneity of $x_t$ relies on the use of a proper IV $z_t$. Later in Sections 3 and 4, we will consider a sequence of random functions $z_t$ satisfying

$$\mathbb{E}[x_t \otimes z_t] \neq 0 \quad \text{and} \quad \mathbb{E}[u_t \otimes z_t] = 0. \quad (2.3)$$

The above conditions are naturally understood as the relevance and validity of the IV $z_t$, respectively, in our setting; however, those are very elementary requirements for $z_t$ to be a proper IV, and far from sufficient for deriving our theoretical results to be given. We will further detail about the required conditions in Sections 3 and 4.

To facilitate the subsequent discussions, it may be helpful to briefly review some of the basic concepts in $\mathcal{H}$ and introduce essential notation. We first let $\mathcal{L}_H$ denote the space of bounded linear operators acting on $\mathcal{H}$, equipped with the operator norm $\|T\|_\infty = \sup \|\zeta\| \|T\zeta\|$. For any $T \in \mathcal{L}_H$, $T^*$ denotes the adjoint of $T$, which is an element of $\mathcal{L}_H$ and satisfies that $\langle T\zeta_1, \zeta_2 \rangle = \langle \zeta_1, T^*\zeta_2 \rangle$ for all $\zeta_1, \zeta_2 \in \mathcal{H}$. Moreover, we let rank $T$ and ker $T$ denote the range and kernel of $T$, respectively. $T$ is said to be nonnegative if $\langle T\zeta, \zeta \rangle \geq 0$ for all $\zeta \in \mathcal{H}$, and positive if the inequality is strict. An element $T \in \mathcal{L}_H$ is called compact if there exist two orthonormal bases $\{\zeta_j\}_{j \geq 1}$ and $\{\zeta_j\}_{j \geq 1}$ of $\mathcal{H}$, and a sequence of real numbers $\{a_j\}_{j \geq 1}$ tending to zero, such that $T = \sum_{j=1}^{\infty} a_j \zeta_j \otimes \zeta_j$. In this expression, it may be assumed that $\zeta_j = \zeta_j$ and $a_1 \geq a_2 \geq \ldots \geq 0$ if $T$ is self-adjoint, i.e., $T = T^*$, and nonnegative (Bosq, 2000, p.35). In this case, $a_j$ becomes an eigenvalue of $T$ and hence $\zeta_1$ is the corresponding eigenfunction, and moreover, we may define $T^{1/2}$ by replacing $a_j$ with $\sqrt{a_j}$. It is well known that the covariance of a $\mathcal{H}$-valued random variable is self-adjoint, nonnegative and compact if exists. A linear operator $T$ is called a Hilbert-Schmidt operator if its Hilbert-Schmidt norm $\|T\|_{HS} = (\sum_{j=1}^{\infty} \|T\zeta_j\|^2)^{1/2}$ is finite, where $\{\zeta_j\}_{j \geq 1}$ is an arbitrary orthonormal basis of $\mathcal{H}$. It is well known that $\|T\|_\infty \leq \|T\|_{HS}$ holds and the collection of Hilbert-Schmidt operators consists of a strict subclass of $\mathcal{L}_H$; see Section 1.5 of Bosq (2000).

3 Functional IV estimator

This section discusses estimation of the model (2.2) given observations $\{y_t, x_t, z_t\}_{t=1}^T$. We first propose the FIVE in detail and study its asymptotic properties.

3.1 The proposed estimator

For notational convenience, we use $C_{zz}$, $C_{xz}$, $C_{yz}$ and $C_{uz}$ to denote the following operators,

$$C_{zz} = \mathbb{E}[z_t \otimes z_t], \quad C_{xz} = \mathbb{E}[x_t \otimes z_t], \quad C_{yz} = \mathbb{E}[y_t \otimes z_t], \quad \text{and} \quad C_{uz} = \mathbb{E}[u_t \otimes z_t].$$
Similarly, let \( \hat{C}_{zz}, \hat{C}_{xz}, \hat{C}_{yz} \) and \( \hat{C}_{uz} \) denote their sample counterparts that are computed as follows,

\[
\hat{C}_{zz} = \frac{1}{T} \sum_{t=1}^{T} z_t \otimes z_t, \quad \hat{C}_{xz} = \frac{1}{T} \sum_{t=1}^{T} x_t \otimes z_t, \quad \hat{C}_{yz} = \frac{1}{T} \sum_{t=1}^{T} y_t \otimes z_t, \quad \text{and} \quad \hat{C}_{uz} = \frac{1}{T} \sum_{t=1}^{T} u_t \otimes z_t.
\]

We then find from (2.2) and (2.3) that

\[
\hat{C}_{yz} = \frac{1}{T} \sum_{t=1}^{T} y_t \otimes z_t = \mathcal{A} \hat{C}_{xz} = \mathcal{A} \hat{E}[z_t \otimes x_t] \quad \text{and hence},
\]

\[
\hat{C}_{yz} \hat{C}_{xz} = \mathcal{A} \hat{C}_{xz} \hat{C}_{xz}.
\]  

As discussed in Mas (2007) and Benatia et al. (2017), \( \mathcal{A} \) is a uniquely identified bounded linear operator if and only if \( \ker \hat{C}_{xz} = \{0\} \). This condition will be maintained throughout the rest of the paper, and we note that all the eigenvalues of \( \hat{C}_{xz} \) are positive under the condition (Mas, 2007, Remark 2.1). In the sequel, we thus may let \( \{ \lambda_j^2 \}_{j \geq 1} \) denote the collection of the eigenvalues of \( \hat{C}_{xz} \) ordered from the largest to the smallest, and represent \( \hat{C}_{xz} \) as its spectral decomposition given by

\[
\hat{C}_{xz} = \sum_{j=1}^{\infty} \lambda_j^2 f_j \otimes f_j,
\]

where \( f_j \) is the eigenfunction corresponding to \( \lambda_j^2 \). Given (3.1), it may be natural to consider an estimator \( \hat{A} \) that satisfies the equation \( \hat{C}_{yz} \hat{C}_{xz} = \hat{A} \hat{C}_{xz} \hat{C}_{xz} \), obtained by replacing \( \hat{C}_{yz} \) and \( \hat{C}_{xz} \) with their sample counterparts. However, it is generally impossible to directly compute the estimator \( \hat{A} \) from this equation since \( \hat{C}_{xz} \hat{C}_{xz} \) is not invertible over the entire Hilbert space \( \mathcal{H} \). We circumvent this issue by employing a regularized inverse of \( \hat{C}_{xz} \hat{C}_{xz} \) which may be understood as the well defined inverse on a strict subspace of \( \mathcal{H} \); this approach has been widely adopted in the literature on functional linear models, see e.g., Bosq (2000), Mas (2007), and Park and Qian (2012) to name a few.

To this end, we first note that \( \hat{C}_{xz} \hat{C}_{xz} \) is nonnegative, self-adjoint and compact and hence allows the following representation:

\[
\hat{C}_{xz} \hat{C}_{xz} = \sum_{j=1}^{\infty} \lambda_j^2 \hat{f}_j \otimes \hat{f}_j,
\]

where \( \{ \hat{\lambda}_j^2, \hat{f}_j \}_{j \geq 1} \) is the pairs of eigenvalues and eigenfunctions, and \( \hat{\lambda}_1^2 \geq \ldots \geq \hat{\lambda}_T^2 \geq 0 = \hat{\lambda}_{T+1}^2 = \ldots \). We then define \( K \) as the random integer determined by the threshold parameter \( \alpha > 0 \) such that

\[
K = \#\{j : \hat{\lambda}_j^2 > \alpha \}.
\]  

Using the first \( K \) eigenfunctions of \( \hat{C}_{xz} \hat{C}_{xz} \), its rank-regularized inverse, denoted \( (\hat{C}_{xz} \hat{C}_{xz})_K^{-1} \), and the FIVE, denoted \( \hat{A} \), are defined as follows:

\[
\hat{A} = \hat{C}_{yz} \hat{C}_{xz} (\hat{C}_{xz} \hat{C}_{xz})_K^{-1}, \quad \text{where} \quad (\hat{C}_{xz} \hat{C}_{xz})_K^{-1} = \sum_{j=1}^{K} \hat{\lambda}_j^{-2} \hat{f}_j \otimes \hat{f}_j.
\]  

The largest eigenvalue of the regularized inverse \( (\hat{C}_{xz} \hat{C}_{xz})_K^{-1} \) is bounded above by \( \alpha^{-1} \) and thus the regularized inverse is a well defined bounded linear operator for every \( \alpha > 0 \). It is worth
mentioning that the FIVE becomes equivalent to the estimator proposed by Park and Qian (2012) in the case where \( z_t = x_t \) and \( K \) is deterministically chosen by researchers (see Remark 1 below), so our estimator may be understood as an extension of their estimator. We also note that the FIVE \( \hat{A} \) may be viewed as a sample-analogue of \( A \) satisfying (3.1) in the sense that \( \hat{A} \) is the solution to \( \hat{C}_{yz} = \hat{A}\hat{C}_{xz} \), on the restricted domain given by \( \hat{H}_K = \text{span}\{\hat{f}_j\}_{j=1}^K \).

Computation of \( \hat{A} \) from data can be done using the FPCA, see Section 3.4 for details.

**Remark 1.** The regularized inverse \( (\hat{C}_{xz}^*\hat{C}_{zz})_K^{-1} \) is obtained by restricting the domain and codomain of \( \hat{C}_{xz}^*\hat{C}_{zz} \) to the space spanned by its first \( K \) eigenfunctions. Such a regularized inverse operator has been considered in the literature; see e.g., Bosq (2000), Mas (2007), and Park and Qian (2012). It should be noted that \( K \) in this paper is by construction a random variable associated with the choice of \( \alpha \), while it is directly chosen by practitioners and hence regarded as deterministic in the foregoing articles. However, even in the latter situation, practitioners are generally recommended to choose \( K \) taking the eigenvalues of \( \hat{C}_{xz}^*\hat{C}_{zz} \) into account, and thus viewing \( K \) as a random variable seems to be natural. This different view on \( K \) helps practitioners more directly control the degree of instability, measured by the largest eigenvalue, of the regularized inverse by the parameter \( \alpha \) that they choose, and, moreover, makes our asymptotic approach be differentiated from those in the aforementioned papers.

### 3.2 General asymptotic properties

As may be deduced from our construction of the FIVE, \( \hat{A} = 0 \) is imposed outside a subspace of which dimension increases as \( \alpha \) gets smaller. Thus, for \( \hat{A} \) to be a consistent estimator of \( A \) defined on the entire space \( \mathcal{H} \), the regularization parameter \( \alpha \) given in (3.2) needs to shrink to zero. Taking this into consideration, we investigate the asymptotic properties of the FIVE when \( T \to \infty \) and \( \alpha \to 0 \) jointly.

Throughout this section, the following assumptions are employed: below, \( \mathfrak{F}_t \) denotes the filtration given by \( \mathfrak{F}_t = \sigma \{ \{ z_s \}_{s \leq t+1}, \{ u_s \}_{s \leq t} \} \), and \( \hat{C}_{uu} = T^{-1} \sum_{t=1}^T u_t \otimes u_t \).

**Assumption 1.** (a) (2.2) holds, (b) \( \{ x_t, z_t \}_{t \geq 1} \) is stationary and geometrically strongly mixing in \( \mathcal{H} \times \mathcal{H} \), \( \mathbb{E}[\|x_t\|^2] < \infty \), and \( \mathbb{E}[\|z_t\|^2] < \infty \), (c) \( \mathbb{E}[u_t|\mathfrak{F}_{t-1}] = 0 \), (d) \( \mathbb{E}[u_t \otimes u_t|\mathfrak{F}_{t-1}] = C_{uu} \), and \( \sup_{1 \leq t \leq T} \mathbb{E}[\|u_t\|^{2+\delta}|\mathfrak{F}_{t-1}] < \infty \) for \( \delta > 0 \), (e) \( A \) satisfying (3.1) is identified in \( \mathcal{L}_H \) and Hilbert-Schmidt (i.e., \( \ker \hat{C}_{xz}^* \hat{C}_{zz} = \{ 0 \} \) and \( \|A\|_{HS} < \infty \)), (f) \( \lambda_1^2 > \lambda_2^2 > \ldots > 0 \), (g) \( \|\hat{C}_{xz} - \hat{C}_{xz}\|_{HS} \), \( \|\hat{C}_{zz} - C_{zz}\|_{HS} \), and \( \|\hat{C}_{uz} - C_{uz}\|_{HS} \) are \( O_p(T^{-1/2}) \), (h) \( \|\hat{C}_{uu} - C_{uu}\|_{HS} = o_p(1) \).

By Assumption 1.(b), we allow \( \{ x_t, z_t \}_{t \geq 1} \) to be a weakly dependent sequence; this is because (i) we want to accommodate various empirical examples such as those given in Horváth and Kokoszka (2012, Chapters 13-16) by not restricting our attention to the iid case and (ii) the variables, which will be considered in our empirical application (Section 7), naturally exhibit time series dependence. In Assumptions 1.(c) and 1.(d), the error term \( u_t \) is assumed to be a homoscedastic martingale difference sequence. Assumption 1.(c) states the exogeneity condition required for the IV \( z_t \) in this setting; see Example 3 (and also Remark 2), where a possible example of such an IV is presented for the model given in Example 2. In Assumption 1.(d), we impose some requirements on the moments.
of \( u_t \). We here note that, if \( \{ z_t, u_t \}_{t \geq 1} \) is an iid sequence, as often assumed in the literature, Assumptions 1.(c) and 1.(d) reduce to the following:

\[
E[u_t | z_t] = 0, \quad E[u_t \otimes u_t | z_t] = C_{uu}, \quad \text{and} \quad E[\| u_t \|^{2+\delta} | z_t] < \infty \text{ for some } \delta > 0.
\]

The Hilbert-Schmidt condition of \( \mathcal{A} \) given in Assumption 1.(e) would become redundant if we considered a finite dimensional Hilbert space, but in our setting it imposes a nontrivial mathematical condition on \( \mathcal{A} \). In addition, as discussed earlier in Remark 2.1 of Mas (2007), it should also be noted that, under Assumption 1.(e), \( C_{xz} \neq 0 \) and, moreover, the eigenvalues of \( C_{xz}^* C_{xz} \) are necessarily positive. We further assume that these eigenvalues are distinct in Assumption 1.(f). Assumptions 1.(e) and 1.(f) are employed for mathematical convenience and may not be restrictive in practice; in fact, similar assumptions have been employed in the literature on functional linear models, see e.g., Bosq (2000, Section 8.3) and Park and Qian (2012). Lastly, in Assumptions 1.(g) and 1.(h) high-level conditions on limiting behaviors of some sample operators are given, and these are also for mathematical convenience. We first note that \( \{ x_t \otimes z_t - C_{xz} \}_{t \geq 1}, \{ z_t \otimes z_t - C_{zz} \}_{t \geq 1}, \) and \( \{ u_t \otimes z_t - C_{uz} \}_{t \geq 1} \) are sequences in the Hilbert space of Hilbert-Schmidt operators, denoted by \( \mathcal{S}_H \) (see Section S4 in the Supplementary Material). If those sequences are iid (resp. geometrically strongly mixing), then Assumption 1.(g) holds once \( E[\| x_t \| \| z_t \|^{v}] \), \( E[\| z_t \|^{2v}] \), and \( E[\| u_t \| \| z_t \|^{v}] \) are finite for some \( v \geq 2 \) (resp. \( v \geq 2+\delta \) for some \( \delta > 0 \)) (Bosq, 2000, Theorems 2.7 and 2.17); such primitive sufficient conditions can also be found for martingale differences (Bosq, 2000, Theorem 2.16) and weakly stationary sequences (Bosq, 2000, Theorems 2.18). We also observe that \( \{ u_t \otimes u_t - C_{uu} \}_{t \geq 1} \) is a martingale difference sequence in \( \mathcal{S}_H \), and some primitive sufficient conditions for Assumption 1.(h) can be found in e.g., Theorems 2.11 and 2.14 of Bosq (2000).

**Example 3.** Consider Example 2 in Section 2.1. If the sequence of measurement errors \( \{ e_t \}_{t \geq 1} \) satisfies that \( E[e_t \otimes e_{t-\ell}] = 0 \) for \( \ell \neq 0 \) and \( E[e_t \otimes y_{t-\ell-1}] = E[e_{\ell} \otimes e_{t-\ell}] = 0 \) for \( \ell \geq 0 \), then \( y_{t-\ell} \) for \( \ell > 1 \) satisfies the exogeneity condition given by Assumption 1.(c). For example, \( y_{t-2} \) satisfies that

\[
E[y_{t-2} \otimes y_{t}] = \mathcal{A} E[y_{t-2} \otimes y_{t-1}] + E[y_{t-2} \otimes u_t] = \mathcal{A} E[y_{t-2} \otimes y_{t-1}] + E[y_{t-2} \otimes u_t].
\]  

(3.4)

The above equation reveals a theoretical connection between our approach and the modified Yule-Walker method, which was introduced to deal with uncorrelated measurement errors in AR models. In a univariate AR(1) case, the modified Yule-Walker estimator can be obtained by replacing the population moments in (3.4) by their sample counterparts; see Walker (1960) and Staudenmayer and Buonaccorsi (2005, Section 4.4).

**Remark 2.** As may be expected from Example 3, if \( \{ y_t, x_t \}_{t \geq 1} \) is a time series satisfying (2.2), the lagged explanatory variables \( \{ x_{t-\ell} \}_{t \geq 1} \) for \( \ell = 1, \ldots, L \) may be suitable candidates for \( z_t \). Chen et al. (2022) noted this and proposed \( \sum_{\ell=1}^{L} x_{t-\ell} \) as the IV to obtain a consistent estimator of the regression operator \( \mathcal{A} \). In this regard, our model is related to that of Chen et al. (2022), but the theory and methodology that are subsequently pursued in this paper move in an apparently different direction.

We now provide the asymptotic properties of the estimator \( \hat{\mathcal{A}} \) when \( \alpha^{-1} \) and \( T \) grow jointly.
without bound. To this end, we consider the following decomposition of $\hat{A} - A$:

$$\hat{A} - A = (\hat{A} - A\hat{\Pi}_K) - A(I - \hat{\Pi}_K),$$ (3.5)

where $\hat{\Pi}_K$ denotes the orthogonal projection defined by $\hat{\Pi}_K = \sum_{j=1}^{K} \hat{f}_j \otimes \hat{f}_j$ and $I$ is the identity operator acting on $H$. Given that the FIVE is computed on the restricted domain $\text{ran}(I - \hat{\Pi}_K)$ (note that $\hat{A} = 0$ on $\text{ran}(I - \hat{\Pi}_K)$ by construction) the first term of (3.5) may be understood as the deviation of $\hat{A}$ from $A$ on $\text{ran}(\hat{\Pi}_K)$, and thus this term is hereafter called the deviation component on the restricted domain (the DR component). On the other hand, the second term $A(I - \hat{\Pi}_K)$ may be understood as the bias induced by the fact that $\hat{A}$ is enforced to zero on the $\text{ran}(I - \hat{\Pi}_K)$.

We thus call this term the regularization bias component (the RB component). Our first result below shows that both DR and RB components are asymptotically negligible and thus $\hat{A}$ becomes weakly consistent once the regularization parameter $\alpha$ decays to zero at an appropriate rate: in the next theorem, we let $\tau(\alpha)$ be a random function that increases without bound as $\alpha \to 0$, which is defined by $\tau(\alpha) = \sum_{j=1}^{K} \tau_j$, where $\tau_j = 2\sqrt{2} \max\{(\lambda^2_{j-1} - \lambda^2_j)^{-1}, (\lambda^2_j - \lambda^2_{j+1})^{-1}\}$.

**Theorem 1.** Suppose that Assumption 1 is satisfied, $T^{-1/2}\tau(\alpha) \xrightarrow{p} 0$ and $T^{-1}\alpha^{-1} \to 0$ as $\alpha \to 0$ and $T \to \infty$. Then

$$\|\hat{A} - A\hat{\Pi}_K\|_{\text{op}}^2 = O_p(T^{-1}\alpha^{-1}) \quad \text{and} \quad \|A(I - \hat{\Pi}_K)\|_{\text{op}}^2 = o_p(1).$$

The following is an immediate consequence of Theorem 1 and Assumption 1.(h).

**Corollary 1.** Suppose that the assumptions in Theorem 1 are satisfied, and let $\hat{u}_t = y_t - \hat{A}x_t$. Then $\|T^{-1} \sum_{t=1}^{T} \hat{u}_t \otimes \hat{u}_t - C_{uu}\|_{\text{op}} \xrightarrow{p} 0$.

The condition imposed on $\tau(\alpha)$ in Theorem 1 does not place any essential restrictions on the eigenvalues of $C_{xz}^*C_{xz}$. Given that $\tau(\alpha)$ increases as $\alpha^{-1}$ (and thus $K$) gets larger, the condition, together with the requirement that $\alpha T \to \infty$, merely tells us that $\alpha^{-1}$ needs to grow with a sufficiently slower rate than $T$ for the weak consistency of the FIVE. In fact, under some additional conditions, the strong (almost sure) consistency of the estimator can also be derived; we need more mathematical preliminaries to present this result, and thus leave the discussion to Section S4.1 in the Supplementary Material.

**Remark 3.** The result in Theorem 1 is, at least to some extent, related to a similar consistency result given by Benatia et al. (2017) for their functional IV estimator. In order to obtain an estimator from (3.1), the authors employ the ridge regularized inverse of $\hat{C}_{xz}^*\hat{C}_{xz}$, while we use a rank-regularized inverse of $\hat{C}_{xz}^*\hat{C}_{xz}$. This makes a significant difference in asymptotic approaches to establish consistency in the two papers. For example, our result is based on the FPCA, and thus we require the eigenvalues of $C_{xz}^*C_{xz}$ to be distinct, which is not required in Benatia et al. (2017). In addition, Benatia et al.’s (2017) approach restricts the range of $A$ to a certain subspace of $H$, called the $\beta$-regularity space, while we need to restrict the increasing rate of $\alpha^{-1}$ or $K$ depending on $\tau(\alpha)$. 

11
Under stronger assumptions than what we require for the weak consistency of \( \hat{\Theta} \), we can further find that (i) the decaying rate of \( \hat{\Theta} - \Theta \) is not uniform over the entire Hilbert space \( \mathcal{H} \), and (ii) the choice of \( \alpha \) can affect the decaying rates of the DR and RB components in the opposite directions. These results are given as consequences of the following asymptotic normality result of the DR component; in the theorem below, \( (C_{xz}^* C_{zz})^{-1}_K \) denotes the operator given by \( \sum_{j=1}^K \lambda_j^{-2} f_j \otimes f_j \) and \( N(0, \mathcal{G}) \) denotes Gaussian random element in \( \mathcal{H} \) with covariance operator \( \mathcal{G} \).

**Theorem 2.** Suppose that Assumption 1 is satisfied, \( \alpha^{-1/2}T^{-1/2} \tau(\alpha) \stackrel{P}{\to} 0 \) and \( T^{-1} \alpha^{-1} \to 0 \) as \( \alpha \to 0 \) and \( T \to \infty \). Then the following hold for any \( \zeta \in \mathcal{H} \).

(i) \( \sqrt{T} \theta_K(\zeta)(\hat{\Theta} - \Theta \hat{\Pi}_K) \zeta \overset{d}{\to} N(0, C_{uu}) \), where \( \theta_K(\zeta) = \langle \zeta, (C_{xz}^* C_{zz})^{-1}_K C_{xz}^* C_{zz} C_{zz}(C_{xz}^* C_{zz})^{-1}_K \zeta \rangle \).

(ii) If \( \tilde{\theta}_K(\zeta) := \langle \zeta, (C_{xz}^* C_{zz})^{-1}_K C_{xz}^* \hat{C}_{zz} \hat{C}_{zz} C_{zz}(C_{xz}^* \hat{C}_{zz})^{-1}_K \zeta \rangle \), then \( \| \tilde{\theta}_K(\zeta) - \theta_K(\zeta) \| \stackrel{P}{\to} 0. \)

Depending on the choice of \( \zeta \), \( \theta_K(\zeta) \) may be convergent or divergent in probability (see Remark 4 below), and thus the convergence rate of the DR component \( (\hat{\Theta} - \Theta \hat{\Pi}_K) \zeta \) depends on \( \zeta \). This finding is not completely new; similar results were formerly observed by Mas (2007) and Hu and Park (2016) in the context of functional AR(1) models. If \( \theta_K(\zeta) \) is convergent in probability, then \( (\hat{\Theta} - \Theta \hat{\Pi}_K) \zeta \) converges at \( \sqrt{T} \)-rate, otherwise it converges at a slower rate given by \( \sqrt{T}/\theta_K(\zeta) \) which is random (note that \( K \) is random in our setting). As noted by Mas (2007), this discrepancy in convergence rates implies that (i) there exists no sequence of normalizing constants \( c_T \) such that \( c_T(\hat{\Theta} - \Theta \hat{\Pi}_K) \zeta \) weakly converges to a well defined limiting distribution uniformly in \( \zeta \in \mathcal{H} \), and thus it is impossible that \( \hat{\Theta} - \Theta \hat{\Pi}_K \) weakly converges to a well defined bounded linear operator in the topology of \( \mathcal{L}\mathcal{H} \), and (ii) this statement is also true if \( \hat{\Pi}_K \) is replaced by \( \Theta \) (see Theorem 3.1 of Mas, 2007). Moreover, Theorem 2 clearly shows that the regularization parameter \( \alpha \) induces a trade-off between the decaying rates of the DR and RB components if \( \theta_K(\zeta) \) is not convergent as \( K \) tends to infinity. If \( K \) relative to \( T \) increases by a smaller choice of \( \alpha \), obviously, the operator norm of the RB component tends to be attenuated at a faster rate. On the other hand, this change, at the same time, makes the DR component decay at a slower rate, which follows from that \( \theta_K(\zeta) \) monotonically increases in \( K \).

**Remark 4.** As a simple way to see if \( \theta_K(\zeta) \) can be either convergent or divergent depending on the choice of \( \zeta \), it is useful to assume that \( x_t = z_t + v_t \), where \( \{v_t\}_{t \geq 1} \) satisfies \( \mathbb{E}[x_t | \Pi K] = \hat{A} \zeta \), which can be consistently estimated by \( \hat{\Theta} \). Let \( \{\zeta_K\} \) be a sequence of random elements given by \( \zeta_K = \hat{\Pi}_K \zeta \). Then \( \zeta_K \) is given by the orthogonal projection of the new perturbation \( \zeta \) onto the subspace on which the sample cross-covariance of \( x_t \) and \( z_t \) is the most explained (in a certain sense, see Remark 5) among all the subspaces of dimension \( K \); that is, \( \zeta_K \) is the best linear approximation of \( \zeta \) based
on covariation of the explanatory and instrumental variables. Therefore, $\zeta_K$ may be interpreted as a nice approximation showing how a hypothetical perturbation $\zeta$ can be revealed given the data set, and we thus call $\zeta_K$ a data-supporting approximation of $\zeta$. The following is an immediate consequence of Theorem 2: under the assumptions employed in Theorem 2

$$|\theta_K(\zeta_K) - \theta_K(\zeta)| \xrightarrow{P} 0 \quad \text{and} \quad \sqrt{T/\theta_K(\zeta_K)}(\hat{A} - A)\zeta_K \xrightarrow{d} N(0, C_{uu}).$$

(3.6)

Based on this result, we may implement standard statistical inference on various characteristics of $A\zeta_K$, which may be naturally interpreted and thus useful for practitioners. We illustrate this by constructing a confidence interval for the random variable given by $\langle A\zeta_K, \psi \rangle$ for some $\psi \in \mathcal{H}$. In fact, various characteristics of $A\zeta_K$ may be written in this form; for example, if $\psi(s) = 1\{s_1 \leq s \leq s_2\}$ then $\langle A\zeta_K, \psi \rangle = \int_{s_1}^{s_2} A\zeta_K(s)ds$ means the locally (if $s_1 \neq 0$ or $s_2 \neq 1$) or globally (if $s_1 = 0$ and $s_2 = 1$) aggregated marginal effect of $\zeta_K$ on $y_t$. We then consider the interval whose endpoints are given as follows,

$$\langle \hat{A}\zeta_K, \psi \rangle \pm \bar{\Phi}^{-1}(1 - \bar{\varpi}/2)\sqrt{\hat{\theta}_K(\zeta_K)}\langle \hat{C}_{uu}^{-1}\psi, \psi \rangle/T,$$

(3.7)

where $\bar{\Phi}^{-1}(\cdot)$ is the quantile function of the standard normal distribution and $\hat{C}_{uu} = T^{-1}\sum_{t=1}^{T}\hat{u}_t \otimes \hat{u}_t$. Based on Theorem 2 and Corollary 1, the intervals that are repeatedly constructed as in (3.7) are expected to include $\langle A\zeta_K, \psi \rangle$ with $100(1 - \bar{\varpi})\%$ of probability for a large $T$. Of course, (3.7) may not be quite satisfactory for practitioners who want to consider a purely hypothetical perturbation $\zeta$ without data-supporting approximation; however, given that (i) the discrepancy between $\zeta$ and $\zeta_K$ caused by the noninvertibility of $\hat{C}_{xzz}^*\hat{C}_{xzz}$, which inevitably arises in our functional setting, and (ii) the magnitude of the discrepancy is expected to be small since it is, anyhow, asymptotically negligible, a small bias caused by data-supporting approximation may be understood as a cost to implement standard inference based on asymptotic normality in our setting. Furthermore and more importantly, it will be shown in Section 3.3 that, if certain conditions, which are not that restrictive, are satisfied, then the convergence result given in Theorem 2.(i) holds if $A\hat{\Pi}_K$ is replaced by $A$ (see Remark 10); this, of course, implies that (3.7) can be understood as a confidence interval for $\langle A\zeta, \psi \rangle$ with no data-supporting approximation.

**Remark 5.** Note that $\|\hat{C}_{xzz}\|_{HS} = \sum_{j=1}^{\infty} \overline{\lambda}_j^2 = \sum_{j=1}^{\infty} \|\hat{C}_{xzz}\zeta_j\|^2$ holds for any arbitrary orthonormal basis $\{\zeta_j\}_{j \geq 1}$ of $\mathcal{H}$. We then may define the proportion of the sample cross-covariance operator explained by the first $K$ orthonormal vectors as

$$\sum_{j=1}^{K} \|\hat{C}_{xzz}\zeta_j\|^2 / \sum_{j=1}^{\infty} \overline{\lambda}_j^2 = \sum_{k=1}^{K} \sum_{j=1}^{\infty} \hat{\lambda}_j^2 \langle \hat{f}_j, \zeta_k \rangle^2 / \sum_{j=1}^{\infty} \hat{\lambda}_j^2.$$

Provided that $\{\hat{\lambda}_j^2, \hat{f}_j\}_{j \geq 1}$ is the sequence of the eigenelements of $\hat{C}_{xzz}^*\hat{C}_{xzz}$, it is deduced from the results given in Horváth and Kokoszka (2012, Theorem 3.2 and Section 3.2) that the above quantity is bounded above by $\sum_{j=1}^{K} \hat{\lambda}_j^2 / \sum_{j=1}^{\infty} \hat{\lambda}_j^2$, and this upper bound is attained if and only if $\zeta_k = \pm \hat{f}_j$ for $j = 1, \ldots, K$. This shows that, among all the subspaces of dimension $K$, ran $\hat{\Pi}_K$ is the unique one that explains the most proportion of the squared Hilbert-Schmidt norm of $\hat{C}_{xzz}$.
3.3 Refinements of the general asymptotic results

In Section 3.2, we established some general asymptotic properties of the FIVE, which do not require any specific assumptions on the eigenstructure of the cross-covariance of \( x_t \) and \( z_t \) other than the assumption of distinct eigenvalues. The results given in the previous section tell us that the FIVE is a reasonable estimator in this functional setting. However, what can be learned from Theorems 1 and 2 is not rich enough; we only know that the FIVE is consistent (Theorem 1) and its DR component is asymptotically normal in a pointwise sense (Theorem 2) if \( \alpha \) decays to zero at a sufficiently slow rate. We in this section investigate the asymptotic behavior of the FIVE in more detail under some additional assumptions that are not restrictive in practice. By doing so, we will obtain useful refinements of Theorems 1 and 2. The specific assumptions that we need are given as follows: in the assumption below, we let \( v_t(j, \ell) = \langle x_t, f_j \rangle \langle z_t, \xi_\ell \rangle - \mathbb{E}[\langle x_t, f_j \rangle \langle z_t, \xi_\ell \rangle] \) for \( j, \ell \geq 1 \).

**Assumption 2.** Assumption 1 holds, and there exist constants \( c_0 > 0, \rho > 2, \varsigma > 1/2, \) and \( \gamma > 1/2 \) satisfying the following: (a) \( \lambda_j^2 \leq c_0 j^{-\rho} \), (b) \( \lambda_j^2 - \lambda_{j+1}^2 \geq c_0^{-1} j^{-\rho-1} \), (c) \( |\langle A f_j, \xi_\ell \rangle| \leq c_0 j^{-\varsigma} \ell^{-\gamma} \), (d) \( \mathbb{E}[v_t(j, \ell)v_{t-s}(j, \ell)] \leq c_0 s^{-m}\mathbb{E}[v_t^2(j, \ell)] \) for \( m > 1 \), and furthermore, \( \mathbb{E}[\|\langle x_t, f_j \rangle z_t \|^2] \leq c_0 \lambda_j^2 \) and \( \mathbb{E}[\|\langle z_t, \xi_\ell \rangle x_t \|^2] \leq c_0 \lambda_j^2 \).

Assumptions 2.(a) and 2.(b) restrict the eigenstructure of \( C_{xz} \) (or equivalently \( C_{xz}^* C_{xz} \)), which are adapted from similar conditions in Hall and Horowitz (2007) and Imaizumi and Kato (2018),\(^1\) which does not seem to be restrictive. Assumption 2.(c) is a very natural condition given that \( \langle A f_j, \xi_\ell \rangle \) must be square-summable with respect to both \( j \) and \( \ell \); in this assumption, it is worth mentioning that \( \varsigma \) is the parameter determining smoothness of \( A \) on \( \text{ran} C_{xz}^* C_{xz} \). As may be deduced from the definition of \( v_t(j, \ell) \) and Assumption 1.(b), \( \{v_t(j, \ell)\}_{t \geq 1} \) is a stationary sequence in \( \mathbb{R} \) for each \( j \) and \( \ell \), and the former condition of Assumption 2.(d) states that its lag-\( s \) autocovariance function decays at a sufficiently fast rate; this condition is satisfied for a wide class of stationary processes. Note that \( \mathbb{E}[\|\langle x_t, f_j \rangle z_t \|^2] \) and \( \mathbb{E}[\|\langle z_t, \xi_\ell \rangle x_t \|^2] \) naturally decrease as \( j \) gets larger and its decay rate is restricted by Assumption 2.(d). Specifically, we require the second moments of \( \|\langle x_t, f_j \rangle z_t \| \) and \( \|\langle z_t, \xi_\ell \rangle x_t \| \) as functions of \( j \) have a constant multiple of \( \lambda_j \) as their upper envelope; a similar condition for the iid case can be found in e.g. Hall and Horowitz (2007).

The following theorem refines the result given in Theorem 1 under Assumption 2.

**Theorem 3.** Suppose that Assumption 2 is satisfied and \( \alpha^{-1} = o(T^{\rho/(2\rho+2)}) \). Then, \( \|\hat{A} - A\hat{\Pi}_K\|_{op}^2 = O_p(T^{-1}\alpha^{-1}) \) as in Theorem 1, and

\[
\|A(I - \hat{\Pi}_K)\|_{op}^2 = O_p(T^{-1}\alpha^{-1}\max\{1, \alpha^{(2\varsigma-3)/\rho}\} + \alpha^{(2\varsigma-1)/\rho}).
\]

Thus, \( \|\hat{A} - A\|_{op} = o_p(1) \) for any \( \rho > 2 \) and \( \varsigma > 1/2 \).

Some comments on the requirement \( \alpha^{-1} = o(T^{\rho/(2\rho+2)}) \) are first in order. This condition is needed in our proof of Theorem 3 to deal with estimation errors associated with \( \hat{\lambda}_j \) (see Remark

---

\(^1\) In fact, Assumption 2.(a) and Assumption 2.(b) can be replaced the following conditions: \( |\lambda_j| \leq c_0 j^{-\rho/2} \) and \( |\lambda_j - |\lambda_{j+1}| \geq c_0 j^{-\rho/2-1} \) for \( \rho > 2 \). These conditions are directly comparable with similar conditions given for the eigenvalues of \( \mathbb{E}[x_t \otimes x_t] \) in Hall and Horowitz (2007) and Imaizumi and Kato (2018).
7). This may be replaced by a sufficient and more convenient condition given by $\alpha^{-1} = o(T^{1/3})$, which does not depend on the value of $\rho$ under Assumption 2 requiring $\rho > 2$.

Theorem 3 not only gives us a more detailed consistency result than that given in Theorem 1, but also better clarifies how certain parameters appearing in Assumption 2 can affect the convergence rate of the FIVE. Specifically, in the above theorem, the convergence rate is characterized by the regularization parameter $\alpha$, the smoothness $\varsigma$ of $A$ on $\text{ran} C^*_xz Cxz$, and the decaying rate $\rho$ of $\lambda_j$ (as a function of $j$). From (3.8), it is quite obvious that if $\alpha$ decays to zero at a slower rate (i.e., $\alpha^{-1}$ diverges to infinity at a faster rate) then the RB component slowly decays to zero as well; this is a quite natural property that can also be deduced from our earlier discussion following Theorem 2 in Section 3.2. Moreover, it can be shown that the convergence rate of the FIVE is generally positively (resp. negative) related to $\varsigma$ (resp. $\rho$); the former is immediately seen from (3.8), and the latter is discussed in detail in Remark 8.

**Remark 6.** In fact, the results given in Theorems 1 and 3 hold if $\| \cdot \|_{op}$ is replaced by $\| \cdot \|_{HS}$, which can be seen in our proofs of those theorems (see Section S1 in the Supplementary Material).

**Remark 7.** The requirement $\alpha^{-1} = o(T^{\rho/(2\rho+2)})$ is used to ensure the existence of a small constant, say $c_\varsigma$, such that $\mathbb{P}(|\tilde{\lambda}_j - \lambda_j| \geq c_\varsigma |\lambda_j - \lambda_\ell|$ for all $1 \leq j \leq K$ and $\ell \neq j \rightarrow 1$; the detailed reason why this result can be obtained from the requirement on $\alpha$ is given in our proof of Theorem 3, which is quite similar to the discussion given by Imaizumi and Kato (2018) following their Theorem 3.

**Remark 8.** If the eigenvalues of $C^*_xz Cxz$ decay to zero at a fast rate (and thus the eigenvalues of $C^*_xz \hat{C}xz$ tend to do so), then the rank-regularized inverse $(\hat{C}^*_xz \hat{C}xz)^{-1}$ tends to be more unstable (unless $K$ becomes smaller) than in the case with more slowly decaying eigenvalues. Thus, it is expected that the convergence rate of the FIVE generally becomes slower if $\rho$ increases. This can be seen from the asymptotic results given in Theorem 3. Let $\rho$ change with the other parameters fixed. In particular, it is assumed that $\alpha^{-1}$ satisfies the condition $\alpha^{-1} = o(T^{\rho/(2\rho+2)})$ both before and after any change in $\rho$ and thus no adjustment in $\alpha$ is required; note that this can always be done by letting $\alpha^{-1} = o(T^{1/3})$ if necessary. From (3.8) it can be shown that the RB component is (i) $O_p(T^{-1} \alpha^{-1}) + O_p(\alpha^{(2\varsigma-1)/\rho})$ if $\varsigma \geq 3/2$ and (ii) $O_p(T^{-1} \alpha^{-(\rho-2\varsigma+3)/\rho}) + O_p(\alpha^{(2\varsigma-1)/\rho})$ if $\varsigma \in (1/2, 3/2)$. In case (i), the decaying rate of the second term is negatively related to $\rho$, and thus an increase in $\rho$ does not yield a faster convergence rate. In case (ii), the decaying rates of both terms are negatively related to $\rho$ since $(\rho - 2\varsigma + 3)/\rho$ is positive and strictly increasing in $\rho$.

We next refine our pointwise normality result under Assumption 2. To this end, it is convenient to decompose the RB component again as follows:

$$
\mathcal{A}(\hat{\Pi}_K - \mathcal{I}) = \mathcal{A}(\hat{\Pi}_K - \Pi_K) + \mathcal{A}(\Pi_K - \mathcal{I}),
$$

(3.9)

where $\Pi_K = \sum_{j=1}^K f_j \otimes f_j$, and this may be understood as the population counterpart of $\hat{\Pi}_K$. The next theorem refines the results given in Theorem 2.
Theorem 4. Suppose that Assumption 2 is satisfied, $\zeta \in \mathcal{H}$ satisfies $\langle f_j, \zeta \rangle \leq c \zeta j^{-\delta} \zeta$ for some $c \zeta > 0$ and $\delta \zeta > 1/2$, and

$$\alpha^{-1} = o(\min\{T^{\rho/(3\rho-2\delta-1)}, T^{\rho/(2\rho+2)}\}).$$  \hfill (3.10)

Then, Theorem 2 holds and

$$\|A(\tilde{\Pi}_K - \Pi_K)\zeta\|_{o_p} = \begin{cases} O_p(T^{-1/2}) & \text{if } \rho/2 + 2 < \varsigma + \delta, \\ O_p(T^{-1/2} \max\{\log \alpha^{-1}, \alpha^{-(\rho/2-\varsigma-\delta+1)/\rho}\}) & \text{if } \rho/2 + 2 \geq \varsigma + \delta, \end{cases}$$

$$\|A(\Pi_K - \mathcal{I})\zeta\|_{o_p} = O_p(\alpha^{(\varsigma+\delta-1)/\rho}).$$

Theorem 4 refines Theorem 2 by providing a detailed asymptotic order of the RB component. Some remarks on the theorem are given in Remarks 9 and 10 below; particularly, in the latter remark, an improvement of the asymptotic normality result in Theorem 2 is discussed.

Remark 9. The decaying rate of $\alpha$ required for Theorem 4 depends on both $\rho$ and $\delta$. If $\delta$ is sufficiently large so that $2\delta \geq \rho - 1$, then (3.10) can be simplified to the following condition: $\alpha^{-1} = o(T^{1/3})$. Moreover, the condition $\delta > 1/2$ is natural since $\langle f_j, \zeta \rangle$ must be square-summable with respect to $j$.

Remark 10 (Pointwise asymptotic normality of the FIVE). A consequence of Theorem 4 is that, if $A$ is smooth enough and $\langle f_j, \delta \zeta \rangle$ decays to zero at a sufficiently fast rate as $j$ increases, then $\sqrt{T/\theta_K(\zeta)}\|A(\tilde{\Pi}_K - \mathcal{I})\zeta\|_{o_p} \overset{p}{\rightarrow} 0$ and thus the result given in Theorem 2(i) can be strengthened to the following:

$$\sqrt{T/\theta_K(\zeta)}(\tilde{A} - A)\zeta \overset{d}{\rightarrow} N(0, C_{uu}).$$  \hfill (3.11)

In this case, of course, (3.7) is understood as a confidence interval for $\langle A\zeta, \psi \rangle$. In particular, if (i) $\varsigma + \delta > \rho/2 + 2$ and (ii) $T\alpha^{2\varsigma+2\delta-1} = O(1)$ (note that both conditions are easier to hold if $\varsigma$ and $\delta$ are large), we have

$$\sqrt{T/\theta_K(\zeta)}\|A(\tilde{\Pi}_K - \mathcal{I})\zeta\|_{o_p} = O_p(1/\sqrt{\theta_K(\zeta)}).$$

The above quantity converges to zero if $\theta_K(\zeta) \overset{p}{\rightarrow} \infty$, which is likely to happen in practice for many possible choices of $\zeta$; for example, if we assume that $\zeta$ is arbitrarily and randomly chosen from $\mathcal{H}$, $\mathbb{P}\{\theta_K(\zeta) < c < \infty\} \rightarrow 0$ as $K \rightarrow \infty$ since $\theta_K$ is only convergent on a strict subspace of $\mathcal{H}$. Given that $\delta > 1/2$, the aforementioned conditions for (3.11) are satisfied if (i)' $\varsigma > \rho/2 + 3/2$ and (ii)' $T\alpha^{2\varsigma} = O(1)$ regardless of the value of $\delta$. The former condition (i)' requires $A$ to have a sufficient smoothness depending on the decaying rate of the eigenvalues of $C_{zz}^2C_{zz}$, and this seems not to be restrictive; in fact, in the literature on functional linear models, it is common to impose such a smoothness condition on $A$ depending on the eigenvalues of a certain covariance or cross-covariance operator (e.g. Hall and Horowitz, 2007; Imaizumi and Kato, 2018; Chen et al., 2022).
3.4 Computation

We here briefly describe how to compute $\hat{A}$ from observations $\{y_t, x_t, z_t\}_{t=1}^T$. Specifically, for each $T$, $\hat{A}$ is a finite rank operator acting on the Hilbert space of square-integrable functions defined on $[0, 1]$, so it allows the following representation (Gohberg et al., 2013, Chapter 8):

$$\hat{A}v(s_1) = \int_0^1 \hat{\kappa}(s_1, s_2)v(s_2)ds_2, \quad s_1, s_2 \in [0, 1],$$

(3.12)

where $v$ is any arbitrary random or nonrandom element $v$ taking values in $\mathcal{H}$ (for example, $v$ can be $x_t$ or any fixed element in $\mathcal{H}$). Therefore, computation of $\hat{A}$ reduces to obtaining an explicit formula for the integral kernel $\hat{\kappa}(s_1, s_2)$ for $s_1, s_2 \in [0, 1]$. We here present a way to compute this integral kernel from the eigenelements of $\hat{C}_{xz}^*\hat{C}_{xz}$ and $\hat{C}_{xz}\hat{C}_{xz}^*$, which can be obtained by the standard functional principal component method; see e.g., Ramsay and Silverman (2005, Chapter 8) and Horváth and Kokoszka (2012, Chapter 3).

Let $\{\xi_j\}_{j \geq 1}$ be the collection of the eigenfunctions of $\hat{C}_{xz}\hat{C}_{xz}^*$, and then note that $\hat{C}_{xz}\hat{f}_j = \lambda_j \xi_j$ and $\hat{C}_{xz}^*\xi_j = \hat{\lambda}_j \hat{f}_j$, where $\lambda_j = \langle \xi_j, \xi_j \rangle$ (Bosq, 2000, Section 4.3). Then, $\hat{A}$ is given by

$$\hat{A} = \hat{C}_{xz}^*\hat{C}_{xz}(\hat{C}_{xz}\hat{C}_{xz}^*)^{-1} = \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^K \lambda_j^{-1} \langle \xi_j, z_t \rangle \hat{f}_j \otimes y_t.$$  

(3.13)

It is quite obvious from (3.13) that the integral kernel $\hat{\kappa}(s_1, s_2)$ for $s_1, s_2 \in [0, 1]$ is given by $T^{-1} \sum_{t=1}^T \sum_{j=1}^K \lambda_j^{-1} \langle \xi_j, z_t \rangle \hat{f}_j(s_1)y_t(s_2)$, and this can be equivalently expressed as follows,

$$\hat{\kappa}(s_1, s_2) = T^{-1} \tilde{F}_K(s_1)' \text{diag}(\hat{\lambda}_j^{-1}) \tilde{G}_K Y_T(s_2),$$

(3.14)

where $Y_T(s) = (y_1(s), \ldots, y_T(s))'$, $\tilde{F}_K(s) = (\tilde{f}_1(s), \ldots, \tilde{f}_K(s))'$ for $s \in [0, 1]$, and $\tilde{G}_K$ is the $K \times T$ matrix whose $(i, t)$-th element is given by $\langle \xi_i, z_t \rangle$. Thus, for each choice of $s_1$ and $s_2$, $\hat{\kappa}(s_1, s_2)$ can be obtained by simple matrix multiplications. In view of (3.13) and (3.14), $\hat{A}v$, for any arbitrary random or nonrandom element $v$ taking values in $\mathcal{H}$, is computed as follows: $\hat{A}v(s) = T^{-1} \tilde{F}_K(v)' \text{diag}(\hat{\lambda}_j^{-1}) \tilde{G}_K Y_T(s)$, where $s \in [0, 1]$ and $\tilde{F}_K(v)' = (\langle \tilde{f}_1, v \rangle, \ldots, \langle \tilde{f}_K, v \rangle)'$.

Computing the FIVE requires choosing $\alpha$ defined in (3.2). The eigenvalue $\hat{\lambda}_j^2$ of $\hat{C}_{xz}\hat{C}_{xz}$ depends on the scales of $x_t$ and $z_t$, and this needs to be considered in choosing $\alpha$. In practice, it thus may be of interest to have a scale-invariant choice of $\alpha$. This can be done by computing the contribution of each of the eigenvalues to a magnitude of the operator $\hat{C}_{xz}$ and then viewing $\alpha$ as the threshold parameter for such computed contributions. We illustrate an easy-to-implement way here. Define $\hat{r}_k = \lambda_k^2 / \sum_{j=1}^\infty \xi_j^2$. Since $\|\hat{C}_{xz}\|_{HS} = \sum_{j=1}^\infty \lambda_j^2$, the ratio $\hat{r}_k$ computes the contribution of the $k$-th eigenvalue to the squared Hilbert-Schmidt norm of $\hat{C}_{xz}$. Of course, the above quantity does not depend on the scales of $x_t$ and $z_t$, and hence, a scale-invariant version of (3.2) may be written as

$$K = \# \{ j : \hat{r}_k > \alpha_T \},$$

(3.15)

where $\alpha_T \in (0, 1)$ depends only on $T$, and shrinks to zero as $T$ increases. An alternative way is
to directly choose $K$, instead of $\alpha$, as the minimal number of the eigenvalues whose sum exceeds a pre-specified proportion of $\|\hat{C}_{xz}\|_{\text{HS}}^{2}$. (Of course, even in this case, it is more natural to understand $K$ as a random variable.) To be more specific, let
\[
\hat{R}_{k} = \sum_{j=1}^{k} \frac{\lambda_{j}^{2}}{\sum_{j=1}^{\infty} \lambda_{j}^{2}} \quad \text{and} \quad K = \min_{k}\{k : \hat{R}_{k} > (1 - \alpha T)\},
\] (3.16)
where $\alpha T \in (0, 1)$ is similarly defined. This choice is obviously scale-invariant as well.

It is also possible to pursue a data-driven selection of $\alpha T$ in (3.15) and (3.16), such as a cross-validation approach, proposed by Benatia et al. (2017) developed in an iid setting. Such a procedure may be adapted for dependent non iid data, but this will not be further studied in this paper.

4 Extension: the functional two-stage least square estimator

In the endogenous linear model in Euclidean space setting, the two-stage least square estimator (2SLSE) is often preferred to the IV estimator because of its estimation efficiency. We in this section consider an extension of this estimator. To the best of the authors’ knowledge, such an estimator has not been considered in the context of function-on-function regression, although a similar estimator was studied by Florens and Van Bellegem (2015) in the case where the dependent variable is scalar-valued. For this estimator, it is assumed that $x_{t}$ and $z_{t}$ satisfy the so-called first-stage relationship given as follows: for a certain linear operator $B$,
\[
x_{t} = Bz_{t} + v_{t}, \quad \text{where} \quad E[z_{t} \otimes v_{t}] = 0 \quad \text{and} \quad E[v_{t}] = 0.
\] (4.1)
If we considered $\mathcal{H} = \mathbb{R}^{n}$, then the 2SLSE is defined as follows:
\[
\tilde{A}^{c} = \tilde{C}_{yz}^{-1} \tilde{C}_{xz} \left(\tilde{C}_{xz}^{*} \tilde{C}_{zz}^{-1} \tilde{C}_{xz}\right)^{-1},
\] (4.2)
and it is widely known that $\tilde{A}^{c}$ has various desirable properties as an estimator of $A$. Coming back to our functional setting, it is not difficult to see that the use of the standard 2SLSE is problematic since it involves $\tilde{C}_{zz}^{-1}$ and $(\tilde{C}_{xz}^{*} \tilde{C}_{zz}^{-1} \tilde{C}_{xz})^{-1}$ which are not well defined as bounded linear operators.

To have a well-behaved analogue of the 2SLSE in our setting, we regularize those inverses as we did in Section 3 and propose an alternative estimator. To this end, we hereafter let $T_{K}^{-1}$ denote the regularized inverse of a compact operator $T$ based on its first $K$ eigenelements (this is defined in the same way as $(\tilde{C}_{xz}^{*} \tilde{C}_{zz}^{-1} \tilde{C}_{xz})^{-1}$ given in (3.3)). Our proposed estimator is defined as follows:
\[
\tilde{A} = \tilde{P} \tilde{Q}_{K_{2}}^{-1}, \quad \text{where} \quad \tilde{P} = \tilde{C}_{xz}^{*} \tilde{C}_{zz}^{-1} \tilde{C}_{xz}^{*} \tilde{Q}_{K_{1}}^{-1}\tilde{C}_{xz}^{*} \text{ and} \quad \tilde{Q} = \tilde{C}_{xz}^{*} \tilde{C}_{zz}^{-1} \tilde{C}_{xz},
\]
and, if we let $\{\hat{\mu}_{j}\}_{j \geq 1}$ (resp. $\{\hat{\nu}_{j}\}_{j \geq 1}$) be the ordered (from the largest to the smallest) eigenvalues of $\tilde{C}_{zz}$ (resp. $\tilde{Q}$),\(^{2}\) then $K_{1}$ and $K_{2}$ are defined as
\[
K_{1} = \#\{j : \hat{\mu}_{j}^{2} > \alpha_{1}\} \quad \text{and} \quad K_{2} = \#\{j : \hat{\nu}_{j}^{2} > \alpha_{2}\}.
\]
\(^{2}\)The eigenvalues of $\tilde{C}_{zz}$ and $\tilde{Q}$ are almost surely positive since they are nonnegative self-adjoint by construction.
Note that by definition, \( K_2 \leq K_1 \) holds almost surely. We conveniently call \( \tilde{A} \) the F2SLSE.

To investigate the asymptotic properties of the F2SLSE, it is necessary to establish some preliminary results and fix notation. First, we note that the operator \( Q \) defined by \( Q = C_{zz}^{-1}C_{xz} \) can be understood as a well defined compact operator. To be more specific, Lemma S2 (and the following discussion given in Section S2 of the Supplementary Material) shows that \( C_{zz}^{-1/2}C_{xz} = R_{xz}C_{xz}^{1/2} \) for some unique bounded linear operator \( R_{xz} \), which may be understood as the correlation operator of \( x_t \) and \( z_t \), and thus \( Q = C_{xz}^{1/2}R_{xz}^*R_{xz}C_{xz}^{1/2} \). From similar arguments, it can be easily shown that the operator \( P = C_{yz}C_{zz}^{-1}C_{xz} \) is also well defined. We then let \( \{\mu_j, g_j\}_{j \geq 1} \) (resp. \( \{\nu_j, h_j\}_{j \geq 1} \)) be the eigenvalues of \( C_{zz} \) (resp. \( Q \)), i.e.,

\[
C_{zz} = \sum_{j=1}^{\infty} \mu_j g_j \otimes g_j \quad \text{and} \quad Q = \sum_{j=1}^{\infty} \nu_j h_j \otimes h_j.
\]

Since \( C_{zz} \) and \( Q \) are self-adjoint and nonnegative, \( \mu_j \) and \( \nu_j \) are all nonnegative. We know from (2.2) and (2.3) that the population relationship \( P = AQ \) holds, and also from our earlier discussion on identification in Section 3.1, \( A \) is uniquely identified if and only if ker \( Q = \{0\} \).

4.1 General asymptotic properties

This section discusses the asymptotic properties of the F2SLSE under the following assumption:

**Assumption 3.** (a) Assumptions 1.(a)-1.(d) and (4.1) hold, (b) \( A \) satisfying \( P = AQ \) is identified in \( L^2 \) and Hilbert-Schmidt (i.e., ker \( Q = \{0\} \) and \( \|A\|_{HS} < \infty \)), (c) \( \mu_1 > \mu_2 > \ldots > 0 \) and \( \nu_1 > \nu_2 > \ldots > 0 \), (d) Assumptions 1.(g)-1.(h) hold and \( \hat{C}_{vz} = T^{-1} \sum_{t=1}^{T} v_t \otimes z_t \) satisfies that \( \|\hat{C}_{vz}\|_{HS} = O_p(T^{-1/2}) \).

Under Assumption 3.(a), our model can be seen as a linear simultaneous equation model involving function-valued random variables. In Assumption 3.(b), \( A \) is assumed to be uniquely identified, and we know in this case that all the eigenvalues of \( Q \) become positive. Assumption 3.(c) is introduced for mathematical convenience, and this makes two seemingly different identification requirements in Assumptions 1.(c) and 3.(b) become equivalent. Recall that \( A \) satisfying \( C_{yz}C_{xz}^* = AC_{xz}^*C_{xz} \) (resp. \( P = AQ \)) is identified if and only if \( C_{xz} \) (resp. \( C_{zz}^{-1/2}C_{xz} \)) is injective. If Assumption 3.(c) is satisfied and hence \( C_{zz} \) is injective, then ker \( C_{xz} \) (\( = \ker C_{xz}^* \ker C_{xz} \)) becomes identical to ker \( C_{zz}^{-1/2}C_{xz} \) (\( = \ker Q \)). Assumption 3.(d) states high-level conditions employed to shorten our proof of the main results; some primitive sufficient conditions can be found in Chapter 2 of Bosq (2000).

As in Section 3.2, we also consider the following decomposition:

\[
\tilde{A} - A = (\tilde{A} - A\tilde{\Pi}_{K_2}) - A(I - \tilde{\Pi}_{K_2}), \tag{4.3}
\]

where \( \tilde{\Pi}_{K_2} \) denotes the orthogonal projection defined by \( \tilde{\Pi}_{K_2} = \sum_{j=1}^{K_2} \tilde{h}_j \otimes \tilde{h}_j \) and \( \{\tilde{h}_j\}_{j=1}^{K_2} \) is the collection of the eigenvectors of \( \tilde{Q} \) corresponding to the first \( K_2 \) leading eigenvalues. The two terms in (4.3) are similarly interpreted as in the case of the FIVE (see (3.9)), and we thus call the first (resp. the second) term the DR (resp. RB) component.
We first show that both DR and RB components can be asymptotically negligible (and thus \( \hat{A} \) can be weakly consistent) if the regularization parameters \( \alpha_1 \) and \( \alpha_2 \) decay to zero at appropriate rates: in the theorem below, we let \( \tau_{1,j} = 2\sqrt{2} \max\{ (\mu_{j-1} - \mu_j)^{-1}, (\mu_j - \mu_{j+1})^{-1} \} \) and \( \tau_{2,j} = 2\sqrt{2} \max\{ (\nu_{j-1} - \nu_j)^{-1}, (\nu_j - \nu_{j+1})^{-1} \} \).

**Theorem 5.** Suppose that Assumption 3 is satisfied, and \( T^{-1/2} (\sum_{j=1}^{K_1} \mu_j \tau_{1,j}) \sum_{j=1}^{K_2} \tau_{2,j} ) \xrightarrow{p} 0 \), \( (\sum_{j=K_1+1}^{\infty} \mu_j)(\sum_{j=1}^{K_2} \tau_{2,j}) \xrightarrow{p} 0 \), \( \alpha_1 \alpha_2^{-1} \rightarrow 0 \), and \( T^{-1} \alpha_1^{-1} \rightarrow 0 \) as \( \alpha_1 \rightarrow 0 \), \( \alpha_2 \rightarrow 0 \) and \( T \rightarrow \infty \). Then
\[
\| \hat{A} - A \hat{\Pi}_{K_2} \|_{\text{op}}^2 = O_p(T^{-1} \alpha_1^{-1/2} \alpha_2^{-1/2}) \quad \text{and} \quad \| A(T - \hat{\Pi}_{K_2}) \|_{\text{op}}^2 = o_p(1).
\]

An immediate consequence of Theorem 5 is given as follows:

**Corollary 2.** Suppose that the assumptions in Theorem 2 are satisfied and let \( \hat{u}_t = y_t - \hat{A} x_t \). Then
\[
\| T^{-1} \sum_{t=1}^{T} \hat{u}_t \otimes \hat{u}_t - C_{uu} \|_{\text{op}} \xrightarrow{p} 0.
\]

As in the case of the FIVE, the conditions imposed on the quantities \( (\sum_{j=1}^{K_1} \mu_j \tau_{1,j}) (\sum_{j=1}^{K_2} \tau_{2,j}) \) and \( (\sum_{j=K_1+1}^{\infty} \mu_j)(\sum_{j=1}^{K_2} \tau_{2,j}) \) are understood not as special restrictions on the eigenvalues, but as requirements on the decaying rates of \( \alpha_1 \) and \( \alpha_2 \) in our asymptotic theory. Specifically, the condition on the former quantity merely requires \( \alpha_1 \) and \( \alpha_2 \) to decrease slowly so that \( K_1 \) and \( K_2 \) tend to grow with sufficiently slower rates than \( T \). Moreover, given the fact that, for fixed \( \alpha_2 \), \( (\sum_{j=K_1+1}^{\infty} \mu_j)(\sum_{j=1}^{K_2} \tau_{2,j}) \) can be arbitrarily small by choosing \( \alpha_1 \) small enough, the condition on the latter quantity merely tells us that the decaying rate of \( \alpha_1 \) needs to be sufficiently higher than that of \( \alpha_2 \). In addition to the weak consistency given by Theorem 5, the strong consistency of the F2SLSE can be derived under some additional conditions; this is discussed in Section S4.2 in the Supplementary Material.

We also obtain an asymptotic normality result similar to that given by Theorem 2 for the FIVE:

**Theorem 6.** Suppose that the assumptions in Theorem 5 are satisfied, \( T^{-1/2} \alpha_1^{-1/2} \sum_{j=1}^{K_1} \tau_{1,j} \xrightarrow{p} 0 \), \( T^{-1/2} \alpha_2^{-1/2} (\sum_{j=1}^{K_1} \mu_j \tau_{1,j}) (\sum_{j=1}^{K_2} \tau_{2,j}) \xrightarrow{p} 0 \), \( \alpha_2^{-1/2} (\sum_{j=K_1+1}^{\infty} \mu_j)(\sum_{j=1}^{K_2} \tau_{2,j}) \xrightarrow{p} 0 \), \( \alpha_1 \alpha_2^{-1} \rightarrow 0 \), and \( T^{-1} \alpha_1^{-1} \rightarrow 0 \) as \( \alpha_1 \rightarrow 0 \), \( \alpha_2 \rightarrow 0 \) and \( T \rightarrow \infty \). Then the following hold for any \( \zeta \in \mathcal{H} \).

1. \( \sqrt{T/\phi_{K_2}(\zeta)} (\hat{A} - A \hat{\Pi}_{K_2}) \zeta \overset{d}{\rightarrow} N(0, C_{uu}) \), where \( \phi_{K_2}(\zeta) = \langle \zeta, \mathcal{Q}_{K_2}^{-1} \zeta \rangle \).
2. If \( \hat{\phi}_{K_2}(\zeta) := \langle \zeta, \hat{\mathcal{Q}}_{K_2}^{-1} \zeta \rangle \), then \( \hat{\phi}_{K_2}(\zeta) - \phi_{K_2}(\zeta) \xrightarrow{p} 0 \).

As shown, we need more stringent requirements on the decaying rates of \( \alpha_1 \) and \( \alpha_2 \). This is mainly due to that the F2SLSE involves the doubly regularized inverse \( \hat{\mathcal{Q}}_{K_2}^{-1} \), which may be ill-behaved if \( \alpha_1 \) and \( \alpha_2 \) do not decay at sufficiently slow rates. This implies that there is no reason why the F2SLSE is generally preferred to the FIVE in this functional setting, unlike what we can expect from the general preference for the 2SLSE by practitioners in Euclidean space setting.

\(^3\)Moreover, the two estimators have different RB terms whose magnitudes depend on various parameters (e.g., the eigenvalues of \( C_s^2 C_s^2 \) and \( \mathcal{Q} \), and thus the RB component of the FIVE can have a smaller asymptotic order.
4.2 Refinements of the general asymptotic results

We in this section provide refinements of Theorems 5 and 6 under the following assumption: below, we let \( \hat{v}_t(j, \ell) = \langle z_t, g_j(z_t, g_\ell) \rangle - \mathbb{E}[\langle z_t, g_j(z_t, g_\ell) \rangle] \) for \( j, \ell \geq 1 \).

**Assumption 4.** Assumption 3 holds, and there exist constants \( c_0 > 0 \), \( \rho_\mu > 2 \), \( \rho_\nu > 2 \), \( \varsigma_\mu > 1/2 \), \( \varsigma_\nu > 1/2 \), \( \gamma_\mu > 1/2 \) and \( \gamma_\nu > 1/2 \) satisfying the following: (a) \( \mu_j^2 \leq c_0 j^{-\rho_\mu} \) (b) \( \mu_j^2 - \mu_{j+1}^2 \geq c_0^{-1} j^{-\rho_\nu}(-1) \), (c) \( \nu_j^2 \leq c_0 j^{-\rho_\nu} \) (d) \( \nu_j^2 - \nu_{j+1}^2 \geq c_0^{-1} j^{-\rho_\nu}(-1) \) (e) \( \langle h_j, A_\ell \rangle \leq c_0 j^{-\gamma_\nu} \ell^{-\varsigma_\nu} \), (f) \( \langle h_j, B_\ell \rangle \leq c_0 j^{-\gamma_\nu} \ell^{-\varsigma_\nu} \) and \( \gamma_\mu \geq \rho_\nu/4 \), (g) \( \mathbb{E}[\hat{v}_t(j, \ell)\hat{v}_{t-s}(j, \ell)] \leq c_0 s^{-m} \mathbb{E}[\hat{v}_t^2(j, \ell)] \) for \( m > 1 \), \( \mathbb{E}[\|z_t, g_j(z_t, g_\ell)\|_2^2] \leq c_0 \mu_j^2 \), and \( \mathbb{E}[\|x_t, h_j(z_t)\|_2^2] \leq c_0 \|C_{zz} h_j\|_2^2 \).

The conditions are somewhat similar to those in Assumption 2, and thus we omit detailed comments except the following two points: (i) from a technical point of view, Assumption 4.(g) is similar to Assumption 2.(d) employed for our study of the FIVE and helps us obtain convergence rates of the eigenelements of \( \hat{Q} \) (which are crucial inputs to our main results for the F2SLSE), and (ii) Assumption 4.(c) is redundant but included as a part of the assumption to stress similarity between our requirements on the eigenstructures of \( C_{zz} \) and \( Q \). In fact, under Assumptions 4. and 4.(f), we have \( \nu_j^2 = \langle Q h_j, h_j \rangle^2 = \langle BC_{zz} B^* h_j, h_j \rangle^2 = (\sum_{\ell=1}^{\infty} \mu_\ell \langle h_j, B_\ell \rangle^2)^2 \leq \tilde{c}_0 j^{-4\gamma_\nu} \leq c_0 j^{-\rho_\nu} \) for some \( \tilde{c}_0 > 0 \). This result not only verifies the redundancy of Assumption 4.(c), but also shows why the condition \( \gamma_\mu \geq \rho_\nu/4 \) is naturally required in Assumption 4.(f).

Our next result refines Theorem 5 by providing a more detailed result on the RB component.

**Theorem 7.** Suppose that Assumption 4 is satisfied, \( \alpha_1^{-1} = o(T^{\rho_\nu/(2\rho_\nu+2)}) \) and \( \alpha_2^{-1} = o(\alpha_1^{-\rho_\nu/(2\rho_\nu+2)}) \). Then, \( \|\hat{A} - A\Pi_{K_2}\|^2_{op} = O_p(T^{-1/2}\alpha_1^{-1/2}\alpha_2^{-1/2}) \) as in Theorem 5, and

\[
\|\hat{A}(I - \hat{\Pi}_{K_2})\|^2_{op} = O_p(\alpha_1^{2\alpha_2^{-3}/\rho_\nu} + \alpha_2^{2\varphi_2^{-1}/\rho_\nu}).
\]

Thus, \( \|\hat{A} - A\|_{op} = o_p(1) \) for any \( \rho_\mu > 2 \), \( \rho_\nu > 2 \), \( \varsigma_\mu > 1/2 \) and \( \varsigma_\nu > 1/2 \).

The convergence rate of the RB component, described in the above theorem, depends not only on the regularization parameters, but on smoothness of \( A \) as in the case of the FIVE. However, the convergence rate described in (4.4) is generally slower than that of the FIVE, and this is somewhat expected from the fact that the F2SLSE involves a doubly regularized (and thus less stable) inverse. Despite this disadvantage of the F2SLSE over the FIVE, our simulation results support that the F2SLSE performs comparably well among a set of the competing estimators (including the FIVE), and thus this estimator can also be used in practice.

Using Assumption 4, Theorem 6 can also be refined as follows: in the theorem below, we, as in (3.9), consider the decomposition of the RB component given by

\[
\hat{A}(\hat{\Pi}_{K_2} - I) = \hat{A}(\Pi_{K_2} - \hat{\Pi}_{K_2}) + \hat{A}(\Pi_{K_2} - I),
\]

where \( \Pi_{K_2} = \sum_{j=1}^{K} h_j \otimes h_j \) is understood as the population counterpart of \( \hat{\Pi}_{K_2} \).
Theorem 8. Suppose that Assumption 2 is satisfied, $\zeta \in \mathcal{H}$ satisfies $\langle h_j, \zeta \rangle \leq c_\zeta j^{-\delta_\zeta}$ for some $c_\zeta > 0$ and $\delta_\zeta > 1/2$, and the following hold:

$$
\alpha_1^{-1} = o(T^{\rho_\nu/(2\rho_\nu+2)}), \quad \alpha_2^{-1} = o\left(\min\{\alpha_1^{-\rho_\nu/(3\rho_\nu-2\delta_\zeta+1)}, \alpha_1^{-\rho_\nu/(2\rho_\nu+2)}\}\right). \quad (4.5)
$$

Then Theorem 6 holds and

$$
\| \mathcal{A}(\tilde{\Pi}_{K_2} - \Pi_{K_2})\zeta \|_{op} = \begin{cases} O_p(\alpha_1^{1/2}) & \text{if } \rho_\nu/2 + 3/2 < c_\nu + \delta_\zeta, \\ O_p(\alpha_1^{1/2}\max\{\log \alpha_2^{-1}, \alpha_2^{-(\rho_\nu/2-\delta_\zeta+2)/\rho_\nu}\}) & \text{if } \rho_\nu/2 + 3/2 \geq c_\nu + \delta_\zeta. 
\end{cases}
$$

Obtaining a pointwise asymptotic normality result that is not dependent on the RB component as in the case of the FIVE (Remark 10) requires more stringent conditions, which will be detailed in Remark 11 below.

Remark 11 (Pointwise asymptotic normality of the F2SLSE). In order to strengthen the result given by Theorem 6(i) using Theorem 8 as in the case of the FIVE, we need more stringent conditions. For example, suppose as in Remark 10 that $\mathcal{A}$ is smooth enough so that $c_\nu > \rho_\nu/2 + 3/2$ and $T^{\rho_\nu/2 - \delta_\zeta - 1} = O(1)$. We then know from Theorem 8 that $\sqrt{T/\phi_{K_2}(\zeta)} \| \mathcal{A}(\tilde{\Pi}_{K_2} - \Pi_{K_2})\zeta \| = O_p(\sqrt{T/\phi_{K_2}(\zeta)}(\mathcal{A}(\Pi_{K_2} - A)\zeta \overset{d}{\to} N(0, C_{uu}))$, if $\phi_{K_2}(\zeta)$ diverges to infinite at a faster rate than that of $T^{\rho_\nu}$. In the case of the FIVE and under an analogous smoothness condition, recall that only $\theta_K(\zeta) \overset{p}{\to} \infty$ is needed to obtain a similar result; see Remark 10.

5 Significance testing in functional endogenous linear model

Practitioners are often interested in testing if various characteristics of $y_t$ depend on $x_t$. Noting that various characteristics of $y_t$ can be written as $\langle y_t, \psi \rangle$ for $\psi \in \mathcal{H}$, we in this section develop a significant test for examining if $\langle y_t, \psi \rangle$ is affected by $x_t$. Specifically, for any $\psi \in \mathcal{H}$, observe that

$$
\langle y_t, \psi \rangle = \langle x_t, \mathcal{A}^* \psi \rangle + \langle u_t, \psi \rangle.
$$

We then want to test the following null and alternative hypotheses:

$$
H_0 : \mathcal{A}^* \psi = 0 \quad \text{v.s.} \quad H_1 : \mathcal{A}^* \psi \neq 0. \quad (5.1)
$$

The null hypothesis means that the characteristic $\langle y_t, \psi \rangle$ of $y_t$ does not linearly depend on $x_t$. Note that $\hat{C}_{yz} \psi = \hat{C}_{xz} A^* \psi + \hat{C}_{uz} \psi$, and hence $\hat{C}_{yz} \psi$ reduces to $\hat{C}_{uz} \psi$ if the null is true; moreover, in this case, $\sqrt{T} \hat{C}_{uz} \psi$ turns out to weakly converge to a $\mathcal{H}$-valued Gaussian random element under relevant assumptions. Using this property, we develop a significance test, which is described by Theorem 9.

Theorem 9. Suppose that (i) $C_{uu}$ is positive definite, (ii) either Assumption 1 or Assumption 3 holds, (iii) $\hat{A}$ is the FIVE (if Assumption 1 holds) or the F2SLSE (if Assumption 3 holds) and the other assumptions for $\| \hat{A} - A \|_{op} \overset{p}{\to} 0$ are satisfied (see Theorems 1, 3, 5 and 7). Let $\tilde{u}_t = y_t - \hat{A} x_t$
and let $J$ denote $T\|\hat{c}_{\psi}^{-1}\hat{C}_{y\psi}\|^2$, where $\hat{c}_{\psi}^2 = \langle T^{-1}\sum_{t=1}^{T} \tilde{u}_t \otimes \tilde{u}_t(\psi), \psi \rangle$. Then, the following hold (below $x_j \sim \text{iid } N(0,1)$).

(i) $J \xrightarrow{d} \sum_{j=1}^{\infty} \mu_j x_j^2$ under $H_0$ of (5.1) while $J \xrightarrow{p} \infty$ under $H_1$ of (5.1).

(ii) If $D \to \infty$ and $D = o(T^{1/2})$, then \( \sum_{j=1}^{\infty} \mu_j x_j^2 = \sum_{j=1}^{D} \tilde{\mu}_j x_j^2 + o_p(1) \).

Note that the asymptotic null distribution of the proposed statistic does not depend on $\psi \in \mathcal{H}$, but does depend on all the eigenvalues of $\mathcal{C}_{zz}$; this means that there are infinitely many nuisance parameters. However, we can approximate the limiting distribution using the estimated eigenvalues of $\mathcal{C}_{zz}$ as described in Theorem 9.(ii) and hence its approximate quantiles can be easily computed by Monte Carlo simulations. Thus, implementation of the proposed test in practice is quite straightforward, which will be illustrated in Sections 6 and 7.

Remark 12. The test proposed in Theorem 9 can obviously be extended to examine the following hypotheses:

\[ H_0 : A^* \psi = 0 \quad \text{v.s.} \quad H_1 : A^* \psi \neq 0, \]

for any $\psi_0 \in \mathcal{H}$. The extension only requires redefining $J$ as $T\|c_{\psi}^{-1}(\hat{C}_{y\psi} - \hat{C}_{xz} \psi_0)\|^2$, and this does not make any change in the convergence results given by (i) and (ii) of Theorem 9.

6 Simulation Study

We investigate the finite sample performance of our estimators via Monte Carlo studies. In all simulation experiments considered in this section, the number of replications is set to 1,000. The response, explanatory, and instrumental variables that will be considered in this section have nonzero means, and hence those are demeaned before computing the estimators of $A$.

6.1 Experiment 1: Functional linear simultaneous equation model

To begin with, we consider the following data generating process (DGP) which is a slight modification of that considered by Benatia et al. (2017): for $t \geq 1$,

\[
\begin{align*}
y_t &= A x_t + u_t, \\
x_t &= \vartheta z_t + v_t,
\end{align*}
\]

where $u_t = 0.8v_t + 0.6\varepsilon_t$, each of \{\(v_t\)\}_{t \geq 1} and \{\(\varepsilon_t\)\}_{t \geq 1} is an iid sequence of standard Brownian bridges, and $E[v_t \otimes \varepsilon_t] = 0$ for all $t, \ell \geq 1$. The regression operator $A$ is the integral operator with kernel $\kappa_A(s_1, s_2) = 1 - |s_1 - s_2|^2$ for $s_1, s_2 \in [0,1]$. The IV $z_t$ is given as follows: for $s \in [0,1]$, \( z_t(s) = \tilde{z}(s; a_t, b_t) + \eta_t(s) \) and \( \tilde{z}(s; a_t, b_t) = \frac{\Gamma(a_t)\Gamma(b_t)}{\Gamma(a_t + b_t)} s^{a_t - 1}(1 - s)^{b_t - 1}, \)

where $a_t$ and $b_t$ are randomly drawn from the uniform distribution $U[2,5]$ for each $t$. That is, $z_t$ is obtained by adding an additive noise $\eta_t$ to the beta density function with parameters $a_t$ and $b_t$. In practice, the functional variables $y_t$, $x_t$ and $z_t$ may be only partially observed at discrete points.
We thus assume that, for each \( t \), only their discrete realizations at 50 equally-spaced points of \([0,1]\) are available and then construct the functional observations by smoothing such discrete realizations using 31 Fourier basis functions as in Ramsay and Silverman (2005, Chapter 5).

The IV in (6.1) is similar to that used for the simulation experiments implemented by Benatia et al. (2017), in which the additive noise \( \eta_t \) is a constant function whose level is determined by a standard normal random variable, i.e., \( \eta_t(s) = q_t \) for all \( s \in [0,1] \) where \( q_t \sim \text{iid } \mathcal{N}(0,1) \) across \( t \). In our simulation experiments, we allow \( \eta_t \) to have a more general form; to be specific, three different forms of \( \eta_t \) will be considered to see how the performance of our estimators varies depending on those. In every design, we let \( \eta_t = \sum_{j=1}^{31} \sigma_j q_{t,j} \xi_j \) where \( \{\xi_j\}_{j \geq 1} \) is the Fourier basis functions of \( \eta \) and \( q_{t,j} \sim \text{iid } \mathcal{N}(0,1) \) across \( t \) and \( j \). We first consider the case where \( \sigma_j = c_1 \sigma_{\eta} \) for \( j \leq 2 \) and \( \sigma_j = c_1 \sigma_{\eta} (0.1)^{j-2} \) for \( j > 2 \); this is called the sparse design. Secondly, we set \( \sigma_j = \sigma_{\eta}(0.9)^{j-1} \) and call this setting the exponential design. In the last specification, which we call the geometric design, \( \sigma_j \) is set to \( c_2 \sigma_{\eta} j^{-1} \). The parameter \( \sigma_{\eta} \) is set to 0.5 and 0.9, and the constants \( c_1 \) and \( c_2 \) are chosen in order to have the same Hilbert-Schmidt norm of \( \mathbb{E}[\eta_t \otimes \eta_t] \) in the three designs. The first-stage coefficient \( \vartheta \) is chosen so that the first-stage functional coefficient of determination (see, Yao et al., 2005), defined by \( \mathbb{E}[\|\vartheta z_t\|^2]/\mathbb{E}[\|x_t\|^2] \); is set to 0.5; this can be easily done since \( \mathbb{E}[\|\nu_t\|^2] = 1/6 \), \( \mathbb{E}[\|\eta_t\|^2] = \sum_{j=1}^{31} \sigma_j^2 \), and \( \mathbb{E}[\|\tilde{z}_t(s; \alpha_t, \beta_t)\|^2] \) can be approximated by the mean of a large number of simulated IID realizations of \( \tilde{z}_t(s; \alpha_t, \beta_t) \).

### 6.1.1 Estimation results

We first compare the performance of our estimators with two existing estimators. The first is the IV estimator (RIVE) suggested by Benatia et al. (2017). To compute the estimators, we consider \( \delta_n T^{-0.4}\|\hat{C}_{xz}\|_{HS}^2 \) as candidates for the regularization parameters of the FLSE, FIVE, and RIVE, where \( \delta_n \in \mathcal{M} \) and \( \mathcal{M} \) consists of 20 equally-spaced points of \([0.1, T^{0.2}]\). Among such candidates, the regularization parameter is set to the value that minimizes the empirical mean squared error (MSE) for each estimator (of course, for the FLSE, \( z_t = x_t \) and thus \( \hat{C}_{xz} = \hat{C}_{xx} \)). The first regularization parameter \( \alpha_1 \) for the F2SLSE is similarly chosen, but \( \|\hat{C}_{xz}\|_{HS}^2 \), used in defining candidate values of a regularization parameter, is replaced by \( \|\hat{C}_{xx}\|_{HS}^2 \). Once \( \alpha_1 \) is chosen, then the second parameter \( \alpha_2 \) is similarly chosen from \( \delta_n(\alpha_1/\|\hat{C}_{zz}\|_{HS}^2)^{1/2}\|\hat{Q}_{K_1}\|_{HS} \) for \( \delta_n \) taking values in the set of 20 equally-spaced points of \([T^{0.05}, T^{0.2}]\). This setup enforces \( \alpha_1 \) to decay at a faster rate than that of \( \alpha_2 \).

In the top of Table 1, we report the MSE of each estimator with the sample size, \( T \), set to 200 and 500; for any estimator \( \tilde{A} \), the empirical MSE is computed as the squared Hilbert-Schmidt norm of \( \tilde{A} - A \). We first note that the IV estimators outperform the FLSE for all the considered cases. This is obviously because the FLSE does not properly address the endogeneity of \( x_t \). In particular, the FIVE and F2SLSE report smaller MSEs than those of the other estimators in the exponential design, whereas their performance is similar to that of the RIVE in the other two designs.

Overall, our estimators perform comparably with the RIVE in terms of MSE; however, note that our estimators and related asymptotic results provide a way to examine various hypotheses.
on $\mathcal{A}$. Before moving on to another simulation study, we thus examine the coverage probability of the interval (3.7) which is computed from the FIVE. Note that the interval is expected to contain the random quantity $\langle \mathcal{A}\hat{\Pi}_{K}\zeta, \psi \rangle$ with $(100 - \omega)\%$ of probability; moreover, if certain conditions are satisfied (see Remark 10) the interval (3.7) can be understood as the $(100 - \omega)\%$ confidence interval for $\langle \mathcal{A}\zeta, \psi \rangle$ which is nonrandom. Based on Theorem 6 (and also Remark 11), we may construct a similar interval that is computed from the F2SLSE and expected to include $\langle \mathcal{A}\hat{\Pi}_{K}\zeta, \psi \rangle$ (and also $\langle \mathcal{A}\zeta, \psi \rangle$ under certain conditions are satisfied) with $(100 - \omega)\%$ of probability; the coverage of this confidence interval will also be examined in this experiment. In order to compute the coverage probabilities, we let $\psi = \ell_1$ and let $\zeta$ be randomly generated by $\zeta = \sum_{j=1}^{11} \bar{q}_{1,j} \ell_j$ for each realization of the DGP, where $\{\ell_{j}\}_{j \geq 1}$ is the polynomial basis with the constant basis function $\ell_1$ and $\bar{q}_{j} \sim \text{iid} N(0, j^{-4})$ across $j$. The simulation results are reported at the bottom of Table 1. We note that, in all the considered cases, the coverage probabilities for $\langle \mathcal{A}\hat{\Pi}_{K}\zeta, \psi \rangle$ or $\langle \mathcal{A}\hat{\Pi}_{K_{2}}\zeta, \psi \rangle$ are close to the nominal level, which supports our findings in Theorems 2 and 6. Moreover, even if the reported coverage probabilities for $\langle \mathcal{A}\zeta, \psi \rangle$ tend to be worse than those for $\langle \mathcal{A}\hat{\Pi}_{K}\zeta, \psi \rangle$ or $\langle \mathcal{A}\hat{\Pi}_{K_{2}}\zeta, \psi \rangle$, they are still reasonably close to the nominal level of 95%. This is what can be expected from Remarks 10 and 11. In unreported simulations, we further experimented with different choices of $\zeta$ and $\psi$, but found no significant difference.

### 6.1.2 Test results

Under the same DGP, we now explore the finite sample performance of the test for examining the null and alternative hypotheses (5.2), which is proposed in Remark 12 of Section 5. The test

| $\sigma_\eta$ | Sparse Design | Exponential Design | Geometric Design |
|---------------|---------------|-------------------|-----------------|
|               | 0.5           | 0.9               | 0.5             | 0.9             | 0.5             | 0.9             |
| $T$           | 200 500       | 200 500           | 200 500         | 200 500         | 200 500         | 200 500         |
| Empirical MSE |               |                   |                 |
| FIVE          | 0.043 0.030   | 0.042 0.030       | 0.111 0.057     | 0.108 0.055     | 0.046 0.033     | 0.046 0.034     |
| F2SLSE        | 0.043 0.030   | 0.042 0.030       | 0.168 0.081     | 0.142 0.076     | 0.045 0.033     | 0.045 0.033     |
| RIVE          | 0.041 0.030   | 0.040 0.029       | 0.141 0.082     | 0.134 0.079     | 0.045 0.031     | 0.043 0.031     |
| FLSE          | 0.289 0.293   | 0.266 0.269       | 0.506 0.503     | 0.486 0.484     | 0.237 0.234     | 0.216 0.214     |
| Coverage prob. |               |                   |                 |
| $\langle \mathcal{A}\hat{\Pi}_{K}\zeta, \psi \rangle$ or $\langle \mathcal{A}\hat{\Pi}_{K_{2}}\zeta, \psi \rangle$ | FIVE 0.947 0.950 0.943 0.949 | 0.923 0.938 0.929 0.936 | 0.948 0.951 0.951 0.952 |
| F2SLSE        | 0.947 0.950   | 0.943 0.949       | 0.907 0.930     | 0.905 0.927     | 0.950 0.954     | 0.949 0.956     |
| Coverage prob. |               |                   |                 |
| $\langle \mathcal{A}\zeta, \psi \rangle$ | FIVE 0.944 0.940 0.943 0.941 | 0.932 0.950 0.937 0.945 | 0.942 0.946 0.938 0.944 |
| F2SLSE        | 0.944 0.940   | 0.943 0.941       | 0.855 0.923     | 0.882 0.930     | 0.941 0.948     | 0.937 0.946     |

Notes: Based on 1,000 replications. In the top, each cell reports the empirical mean squared error (MSE) of four estimators: FIVE, F2SLSE, Benatia et al.’s (2017) RIVE and Park and Qian’s (2012) FLSE. The last four rows report the coverage probabilities of the designated quantities; the nominal level is 95%.
Figure 1: Rejection probability of $J$ when $A^*\psi = \psi_0 + c_\zeta \tilde{\psi}$

(a) Sparse Design   (b) Exponential Design   (c) Noisy Design

Notes: Based on 1,000 Monte Carlo replications. The rejection probability of $J$ when $\hat{c}_\psi$ is computed from the FIVE is reported according to the value of $c_\zeta \in \{0, 0.01, \ldots, 0.5\}$.

The statistic $J$ is computed with $\hat{c}_\psi$ obtained from the FIVE (see Theorem 9 and Corollary 1). For each realization of the DGP, the critical value at 5% significance level is obtained from 500 Monte Carlo simulations of the distribution given in Theorem 9.(ii) with $D = \lceil T^{1/3} \rceil$. The finite sample properties of the test are investigated by computing its rejection probabilities when $A^*\zeta = \psi_0 + c_\zeta \tilde{\psi}$ holds, where $\zeta$ is defined in Section 6.1.1, $c_\zeta \in \{0, 0.01, \ldots, 0.5\}$, and $\tilde{\psi}$ is a perturbation element with unit norm and randomly generated for each realization of the DGP. Specifically, $\tilde{\psi} = \bar{\psi}/\|\bar{\psi}\|$ and $\bar{\psi} = \sum_{j=1}^{11} \bar{q}_{3,j} \xi_j$, where $\bar{q}_{3,j} \sim \text{iid} \ N(0, 0.5^{2(j-1)})$ across $j$ and $\mathbb{E}[\bar{q}_{1,i} \bar{q}_{3,j}] = 0$ for all $i$ and $j$. This section only considers the case where $\sigma_\eta = 0.5$; in unreported simulations, we also investigated the performance of the test (i) when $\hat{c}_\psi$ is computed from the F2SLSE and (ii) when $\sigma_\eta$ is given by 0.9, but found no significant difference.

Figure 1 shows the rejection probability of the test depending on $c_\zeta$. The dashed and solid lines indicate the rejection probabilities when $T = 200$ and 500, respectively, and the dotted horizontal lines indicate the nominal size of the test. As expected, the proposed test exhibits a higher power as $c_\zeta$ gets deviated from zero, and it seems that the power of the test more rapidly increases in the sparse design compared to those in the other designs. Moreover, in all the considered cases, the test has excellent size control, as it can be seen from the case where $c_\zeta = 0$.

Overall, the results reported in this section show that our estimators and test perform well in the considered simulation designs.

6.2 Experiment 2: AR(1) model of probability density functions

In this section, we implement a quite different experiment to examine the performance of the proposed estimators in the AR(1) model of probability density functions. What is mainly different from the earlier experiments given in Section 6.1 is that endogeneity is not explicitly imposed, but implicitly introduced by estimation errors.

We let $\{p_t^o\}_{t \geq 1}$ be a sequence of probability densities whose support is the unit interval $[0, 1]$, and consider the linear prediction model of $p_t^o$ given $p_{t-1}^o$. Each density may be treated as a random variable taking values in $\mathcal{H}$, but the collection of probability densities in $\mathcal{H}$ is not a linear subspace due to the following nonlinear constraints that any density element $\zeta \in \mathcal{H}$ must satisfy: (i) $\zeta(s) \geq 0$ for all $s \in [0, 1]$ and (ii) $\int_0^1 \zeta(s)ds = 1$. As pointed by Delicado (2011), Petersen and Müller (2016),
Hron et al. (2016), Kokoszka et al. (2019), and Zhang et al. (2021), neglecting these nonlinear constraints may bring about statistical issues. As a way to circumvent such issues, we consider the centered-log-ratio (clr) transformation proposed by Egozcue et al. (2006): for \( s \in [0,1] \),

\[
\begin{align*}
p_t^o(s) &\rightarrow y_t^o(s) = \log p_t^o(s) - \int \log p_t^o(s) ds.
\end{align*}
\] (6.2)

Then \( \{y_t^o\}_{t \geq 1} \) becomes a sequence in \( \mathcal{H}_c \), the collection of all \( \zeta \in \mathcal{H} \) satisfying \( \int_0^1 \zeta(s) ds = 0 \), and \( \mathcal{H}_c \) is obviously a Hilbert space. Furthermore, the clr transformation turns out to be an isomorphism between a certain Hilbert space of probability density functions (called a Bayes Hilbert space) and \( \mathcal{H}_c \), and its inverse transformation is given by \( y_t^o(s) \mapsto \exp(y_t^o(s))/\int_0^1 \exp(y_t^o(s)) ds \). By this property, the linear prediction model of \( p_t^o \) given \( p_{t-1}^o \) may be recast into that of \( y_t^o \) given \( y_{t-1}^o \) in \( \mathcal{H}_c \), where we no longer need to take the nonlinear constraints for densities into account. We thus hereafter consider the following AR(1) prediction model for \( \{y_t^o\}_{t \geq 1} \):

\[
y_t^o = c_y + \mathcal{A}(y_{t-1}^o - c_y) + \varepsilon_t,
\]

where \( y_{t-1}^o \) and \( \varepsilon_t \) are uncorrelated. To mimic situations commonly encountered in practice, we assume that \( p_t^o \) (and thus \( y_t^o \)) is not observed, but only random samples \( \{s_{i,t}\}_{i=1}^n \) drawn from \( p_t^o \) are available. If so, by replacing the density \( p_t^o \) or the log-density \( \log p_t^o \) with its proper nonparametric estimate in (6.2), we may obtain an estimate \( y_t \) of \( y_t^o \), and then, as shown in Example 2, \( \{y_t\}_{t \geq 1} \) satisfies the following:

\[
y_t = c_y + \mathcal{A}(y_{t-1} - c_y) + u_t,
\] (6.3)

where \( y_{t-1} \) and \( u_t \) are generally correlated due to errors arising from the nonparametric estimation. Given (6.3), we will compute the FIVE and F2SLSE by assuming that \( y_{t-2} \) is a proper IV, as in Example 3. Of course, this assumption may not be true depending on how estimation errors are generated. Even with this possibility, it may be of interest to practitioners, who are very often have no choice but to replace \( p_t^o \) or \( \log p_t^o \) with a standard nonparametric estimate, to see if this naive use of our estimators can make any actual improvements in estimating \( \mathcal{A} \). This is the purpose of simulation experiments in this section.

In our simulation experiments given below, we estimate \( y_t^o \) from \( n \) random samples that are generated from \( p_t^o \) by (i) the local likelihood density estimation method proposed by Loader (1996) (see Appendix A2 for more details) and (ii) the standard kernel density estimation method with the Gaussian kernel and Silverman’s rule-of-thumb bandwidth (Silverman, 1998). As will be detailed in Appendix A2, a fundamental difference between these two methods is in that the former is more suitable for estimation of log-densities than the latter, and hence it seems to be a better approach to estimate \( y_t^o \) (Seo and Beare, 2019, Section 4.2). Nevertheless, the latter is considered as well because of its popularity in empirical studies. Once \( y_t \) is computed, then it is represented by the first 30 nonconstant Fourier basis functions (note that constant functions are included in the orthogonal complement of \( \mathcal{H}_c \)) for implementation of the FPCA.

We let \( c_y \) be the clr transformation of the normal density function with mean 0.5 and variance...
Below we consider two different specifications of $\sigma_j$, which are respectively called the sparse design and the exponential design; in the exponential design, $\sigma_j = 0.1(0.9)^{j-1}$, and in the sparse design, $\sigma_j = c_\sigma$ for $j \leq 2$ and $\sigma_j = c_\sigma(0.1)^{j-2}$ for $j > 2$, where $c_\sigma$ is chosen so that the Hilbert-Schmidt norms of $E[\varepsilon_t \otimes \varepsilon_t]$ in both designs are equal. These two designs are respectively obtained by setting $\sigma_q$ to 0.1 in the sparse and exponential designs considered for $\eta_q$ in Section 6.1, and the reason why we choose a relatively smaller scale of $\sigma_j$ in this experiment is only to avoid as much as possible that the simulated densities have shapes that are rarely observed in practice (e.g., densities that are U-shaped or highly multimodal).

We let $\mathcal{A}$ be defined by $\sum_{j=1}^{\infty} a_j \xi_j \otimes \xi_j$, and, for each realization of the DGP, the coefficients $\{a_j\}_{j \geq 1}$ are independently determined across $j$ as follows,

\[ a_1 \sim U[0.4, 0.9], \quad a_2 \sim U[0.4, 0.9], \quad a_j = a_{u,j}(0.5)^{j-2} \quad \text{and} \quad a_{u,j} \sim \text{iid} \ U[0, 0.9] \quad \text{for} \quad j \geq 3. \]

Note here that we let the first two coefficients $a_1$ and $a_2$ be bounded below by 0.4, which is to ensure that the operator norm of the cross-covariance operator of $y_{t-1}^q$ and $y_{t-2}^q$ is bounded away from zero. If this quantity is close to zero, then the employed IV may become ‘weak’ and this case is not considered in the present paper.

Table 2 reports the empirical MSEs of the estimators that were considered in Section 6.1 when $n = 100$ and 150 (recall that $n$ is the number of random samples drawn from the distribution $p_1^\phi$ to estimate $\log p_1^\phi$ or $p_1^\phi$). The IV estimators tend to exhibit smaller empirical MSEs than the

---

**Table 2: Simulation results ($a_1, a_2 \geq 0.4$)**

|                | Sparse Design | Exponential Design |
|----------------|---------------|--------------------|
|                | 100           | 150               | 100          | 150          |
| $T$            |               |                   |              |              |
| 200            | 0.206         | 0.158             | 0.187        | 0.152        | 0.400 | 0.227 | 0.315 | 0.194 |
| 500            | 0.204         | 0.157             | 0.186        | 0.151        | 0.392 | 0.219 | 0.310 | 0.189 |
| Loader’s       |               |                   |              |              |
| FIVE           | 0.214         | 0.161             | 0.192        | 0.153        | 0.404 | 0.237 | 0.329 | 0.198 |
| F2SLSE         | 0.255         | 0.228             | 0.208        | 0.187        | 0.427 | 0.354 | 0.333 | 0.267 |
| RIVE           | 0.272         | 0.218             | 0.223        | 0.188        | 0.395 | 0.257 | 0.314 | 0.215 |
| FLSE           | 0.272         | 0.215             | 0.223        | 0.186        | 0.396 | 0.247 | 0.315 | 0.209 |
| Silverman’s    |               |                   |              |              |
| FIVE           | 0.268         | 0.214             | 0.225        | 0.186        | 0.379 | 0.249 | 0.313 | 0.209 |
| F2SLSE         | 0.326         | 0.291             | 0.251        | 0.226        | 0.418 | 0.351 | 0.322 | 0.258 |
| RIVE           | 0.326         | 0.291             | 0.251        | 0.226        | 0.418 | 0.351 | 0.322 | 0.258 |
| FLSE           | 0.326         | 0.291             | 0.251        | 0.226        | 0.418 | 0.351 | 0.322 | 0.258 |

Notes: Based on 1,000 replications. Each cell reports the empirical mean squared error (MSE) of the four considered estimators: FIVE, F2SLSE, Benatia et al.’s (2017) RIVE and Park and Qian’s (2012) FLSE.

---

$0.25^2$ that is truncated on $[0, 1]$. In addition, $\varepsilon_t = \sum_{j=1}^{\infty} \sigma_j q_{t,j} \xi_j$, where $\{\xi_j\}_{j \geq 1}$ is the Fourier basis functions except for the constant basis function, and $q_{t,j} \sim \text{iid} \ N(0, 1)$ across $t$ and $j$. Below we consider two different specifications of $\sigma_j$, which are respectively called the sparse design and the exponential design; in the exponential design, $\sigma_j = 0.1(0.9)^{j-1}$, and in the sparse design, $\sigma_j = c_\sigma$ for $j \leq 2$ and $\sigma_j = c_\sigma(0.1)^{j-2}$ for $j > 2$, where $c_\sigma$ is chosen so that the Hilbert-Schmidt norms of $E[\varepsilon_t \otimes \varepsilon_t]$ in both designs are equal. These two designs are respectively obtained by setting $\sigma_q$ to 0.1 in the sparse and exponential designs considered for $\eta_q$ in Section 6.1, and the reason why we choose a relatively smaller scale of $\sigma_j$ in this experiment is only to avoid as much as possible that the simulated densities have shapes that are rarely observed in practice (e.g., densities that are U-shaped or highly multimodal).

---

\[ a_1 \sim U[0.4, 0.9], \quad a_2 \sim U[0.4, 0.9], \quad a_j = a_{u,j}(0.5)^{j-2} \quad \text{and} \quad a_{u,j} \sim \text{iid} \ U[0, 0.9] \quad \text{for} \quad j \geq 3. \]

---

\[ ^4 \text{in actual computation, } \varepsilon_t \text{ can be approximated by } \sum_{j=1}^{L} \sigma_j q_{t,j} \xi_j \text{ for some large } L. \text{ We set } L \text{ to 50 in this example and found no significant difference even from big changes in } L \text{ as long as } L \geq 50. \]

\[ ^5 \mathcal{A} \text{ is approximated by } \sum_{j=1}^{\infty} a_j \xi_j \otimes \xi_j \text{ in actual computation as in the case of } \varepsilon_t \]
FLSE, which is consistent with the simulation results given in Section 6.1. Such good comparative performance of the IV estimators is more noticeable when $n$ is small and $T$ is large. This is what can be conjectured from our earlier discussion; as $n$ gets smaller, $y_t$ becomes a less accurate estimate of $y_t^o$, and hence the estimators that address the possible endogeneity caused by estimation errors will work better. Even in comparison with the RIVE, our estimators perform quite well in the considered simulation designs; the FIVE or F2SLSE exhibits the smallest MSE in most of the cases (see also Table A1 in Appendix reporting the simulation results for the case where the lower bound of $a_1$ and $a_2$ increases to 0.6). However, the RIVE sometimes works the best (in Table 2, this happens when kernel density estimation with Silverman’s rule-of-thumb bandwidth is employed), so it is hard to conclude the relative performance between these IV estimators; this may depend on various factors such as the DGP and the method of density estimation. Thus, it would be advisable to use those IV estimators complementarily in practice.

7 Empirical application: effect of immigration on native wages

Estimation of the effect of immigrant inflows on the labor market outcomes of natives of heterogeneous skills has received due attention from both researchers and policymakers. However, it seems that researchers in this area have not reached an agreement on the effect of immigration; as is well known in the literature (e.g., Dustmann et al., 2016), this is because their findings depend on various empirical specifications such as the assumption on the substitutability of workers of different characteristics. Notwithstanding such different conclusions in the literature, researchers agreed that an inflow of immigrants has heterogeneous effects on the labor market outcomes of native workers depending on their skills. This conclusion is supported by empirical evidence obtained by investigating the effect of immigration after classifying workers into groups according to some observable characteristics representing their skills; see, e.g., Card (2009), Borjas et al. (2011), Ottaviano and Peri (2012), and Glitz (2012) to mention a few. For example, Ottaviano and Peri (2012) showed that native- and foreign-born workers of similar skills, proxied by years of education and experience, are imperfectly substitutable. This in turn leads to their conclusion that the effects of an inflow of immigrants on the wages of heterogeneously skilled natives are overall positive, but the magnitude of such an effect is quite dependent on workers’ levels of skill, see, Table 6 in Ottaviano and Peri (2012).

In addition to such empirical evidence on heterogeneous effects of immigration, it has been suggested that natives may adjust their occupations or regions of residence, in responding to an inflow of new workers (see, e.g., Llull, 2018). It is thus advisable to assume that an inflow of immigrants of a particular skill level can affect the labor market outcomes of all natives, and this needs to be considered in empirical analysis. However, doing so would be challenging, since (i) a worker’s skill cannot be observed and has to be proxied by observable characteristics and (ii) as pointed out by Borjas (1987) and Llull (2018), workers’ occupational or regional adjustments in responding to an inflow of immigrants may generate self-selection bias.

In this section, we apply our methodology to this problem with taking the aforementioned challenges into consideration. Specifically, we use national level data and generalize a widely used
empirical model by viewing the variables of interest as functions depending on a certain measure of occupation-specific skill to be defined shortly. We then construct a sort of predicted share of immigrants as an instrument to control for the endogeneity related to the self-selection bias. Our final goal is not only to illustrate the empirical relevance of our estimators but also to contribute to the literature on immigration by using a state-of-the-art econometric model.

Our measure of skill is similar to Peri and Sparber’s (2009) measure of occupation-specific relative provision of communication versus manual skills. Specifically, we use the O*Net ability survey data,\(^6\) in which the importance of each of 52 distinct abilities required by each occupation is quantified. Using the data, we construct the communication skill measure \(c_j^\circ\) and the manual skill measure \(m_j^\circ\) for each occupation \(j\), where the definitions of communication and manual skills are equivalent to the extended definitions of those in Table A.1 of Peri and Sparber (2009). We merge the values of \(c_j^\circ\) and \(m_j^\circ\) to individuals in the 2000 census using the monthly US Current Population Survey (CPS) data. Then, as done similarly by Peri and Sparber (2009), the measure of occupation-specific skill intensity \((s_j)\) is obtained by converting the value of \(c_j^\circ/m_j^\circ\) to its percentile score \((s_{jt}^\circ)\) for each month in 2000 and averaging the monthly scores for each occupation \(j\), i.e., \[s_j = 12^{-1} \sum_{t=1}^{12} s_{jt}^\circ.\] The number of distinct skill levels, \(s_j\), is 223, and, by construction, each occupation is uniquely identified by the skill score \(s_j \in [0,1].\)^7 In Table A2 in Appendix, we report occupations with the lowest and highest scores of relative communication skill provision.

We merge the percentile scores of relative communication skill provision to individuals in the monthly CPS data running from January 1996 to December 2019. The CPS data, which can be downloaded from the Integrated Public Use Microdata Series (IPUMS), provide information on various characteristics of individuals: hourly wage, citizenship status, age, employment status, and occupation. We focus on individuals who (i) are aged between 18 and 64 years, (ii) are not self-employed, and (iii) have positive income. Immigrants are defined by those who are not a citizen or are a naturalized citizen. The skill-dependent labor supply of immigrants \((\ell_{nt}^\circ(s_j))\) and that of natives \((\ell_{nt}^\circ(s_j))\) are computed by the total hours of work per week (weighted by the variable WTFINL in the CPS) provided by the foreign- and native-born workers for each \(s_j\). The skill-dependent native wage is computed by weighted (using the variable EARNWT in the CPS) averaging weekly wages of native workers\(^8\) in the occupation corresponding to \(s_j\), and its logged value \((w_t^\circ(s_j))\) is used for the analysis.

The empirical models used in the labor economics literature (e.g., Dustmann et al., 2013; Sharpe and Bollinger, 2020) can be written as follows: \[
\Delta w_t^\circ(s_j) = \beta_j^\circ \Delta h_t^\circ(s_j) + u_t^\circ(s_j),
\] where \(\Delta w_t^\circ(s_j) = w_t^\circ(s_j) - w_{t-1}^\circ(s_j)\), \(u_t^\circ(s_j)\) denotes the disturbance term, \(\beta_j^\circ\) is the parameter of interest, the explana-

\(^6\)Version 24, provided by the US Department of Labor

\(^7\)A similar rescaling procedure is taken by Peri and Sparber (2009). Specifically, they first converted the values of \(c_j^\circ\) and \(m_j^\circ\) into their percentile scores in 2000 for each \(j\), and then their measure of relative provision of communication versus manual skills is given by the ratio of the percentile scores for each \(j\). In contrast, we first take the ratio of \(c_j^\circ\) and \(m_j^\circ\) and then convert the ratio into the percentile score for each \(j\). This is to ensure that our measure of relative communication skill provision takes values in \([0,1]\).

\(^8\)The weekly wage of a native worker is computed as (hourly wage) \(\times\) (usual hours of work), and the variables required to compute this quantity are also available in the CPS.
tory variable $\Delta h_t^s(s_j)$ is the first difference of $h_t^s(s_j)$, and $h_t^s(s_j) = \ell_t^s(s_j)/\ell_{nt}^s(s_j) + \ell_t^{s_b}(s_j))$. The above model may not be satisfactory to practitioners, since, in the model, an inflow of immigrants in the occupation corresponding to $s_j$ only affects the wages of natives in the occupation requiring the same skill level, which seems to be restrictive. To resolve this issue, one may instead allow spillover effects across occupations, but this requires researchers to estimate too many parameters; for example, if we allow a spillover effect from the occupation corresponding to $s_i$ to another occupation corresponding to $s_j$ for any arbitrary $i, j \in \{1, \ldots, 223\}$, then there are 223^2 elements to be estimated. Furthermore, for a fixed value of $s_j$, $w_t^s(s_j)$ and $h_t^s(s_j)$ are often not observed for every $t$. A possible strategy to resolve these issues may be classifying workers in occupations corresponding to various values of $s_j$ into a few groups so that dimensionality of the model can be reduced, as discussed in Example 1. Another possibility, which will be subsequently pursued, is to view observations $w_t^s(s_j)$ and $h_t^s(s_j)$ for each $t$ as imperfect realizations of curves $w_t$ and $h_t$, and use our methodology developed in the previous sections. To this end, we first estimate each of those curves with the standard Nadaraya-Watson estimator employing the second-order Gaussian kernel and the bandwidth minimizing the least square cross validation criterion. The smoothed curves are represented by 15 cubic B-Spline functions and are denoted by $w_t$ and $h_t$, respectively. Then, we estimate the following model:

$$
\Delta w_t = \mathcal{A} \Delta h_t + u_t,
$$

where $\Delta w_t = w_t - w_{t-1}$, $\Delta h_t = h_t - h_{t-1}$, and $\Delta h_t$ is generally correlated with $u_t$ due to the aforementioned self-selection bias. To estimate (7.1), we use the changes in the imputed share of immigrants as an IV, which has been employed in various contexts, including Card (2009), Peri and Sparber (2009), and Autor and Dorn (2013). Specifically, the imputed share of immigrants in the occupation corresponding to $s_j$, denoted $z_t^s(s_j)$, is defined as follows:

$$
z_t^s(s_j) = \frac{\ell_t^s(s_j)}{\ell_{nt}^s(s_j) + \ell_t^{s_b}(s_j)} \quad \text{and} \quad \ell_t^s(s_j) = \frac{1}{12} \sum_{b=1}^{B} \sum_{t=1}^{12} \ell_{1994}^{b}(s_j) \ell_{1994,t}^{b},
$$

where $b$ denotes the country of birth of immigrants, $\ell_{1994,t}^{b}(s_j)$ is the labor supply of immigrants in the occupation corresponding to $s_j$ from the country $b$ in the month $t$ of the year 1994, and $\ell_{1994,t}^{b}$ is its aggregation over $s_j$. The use of $z_t^s(s_j)$ as an IV relies on the assumption that incoming workers are likely to choose their occupations that are known to provide favorable environments to them (see e.g., Llull, 2018). The curve of imputed shares of immigrants, denoted $z_t$, is obtained by smoothing $z_t^s(s_j)$, and the instrument, denoted $\Delta z_t$, is the first difference of $z_t$.

The smoothed curves are reported in Figure 2 where each of them is a function of relative communication skill provision scores. The solid lines in Figure 2 indicate the mean functions of $w_t$, $h_t$, and $z_t$. Figure 2a shows that native workers tend to be better paid if they are in occupations needing relatively higher communication skills. On the other hand, the share of immigrants tends to be lower in such occupations, and so does the imputed share of immigrants; this may not be surprising because natives would have a comparative advantage in communication intensive tasks.

Before discussing estimation results, recall that the earlier literature mostly relies on the strategy

31
of reducing dimensionality of the model by classifying workers into a few groups according to a measure of their skill. We will make a comparison of our estimation results with those based on this strategy. To this end, we let \( \zeta_j(J) = J \times 1 \{ j - 1 < Js \leq j \} \). Then, the inner product \( \langle \Delta w_t, \zeta_j(J) \rangle \) computes the average of changes in the log wages of natives in the occupations of which \( s \) is between \( (j - 1)/J \) and \( j/J \), at time \( t \). We then estimate the following using the standard 2SLSE,

\[
\Delta w_{t(J)} = \beta \Delta h_{t(J)} + u_{t(J)},
\]

(7.2)

where \( \Delta w_{t(J)} = (\langle \Delta w_t, \zeta_1(J) \rangle, \ldots, \langle \Delta w_t, \zeta_J(J) \rangle)' \) and \( \Delta h_{t(J)} = (\langle \Delta h_t, \zeta_1(J) \rangle, \ldots, \langle \Delta h_t, \zeta_J(J) \rangle)' \). The IV that is used to compute the 2SLSE is \( \Delta z_{t(J)} = (\langle \Delta z_t, \zeta_1(J) \rangle, \ldots, \langle \Delta z_t, \zeta_J(J) \rangle)' \).

For illustration purposes, in Figure 3, we report the time series of \( \langle \Delta w_t, \zeta_1(3) \rangle \), \( \langle \Delta w_t, \zeta_2(3) \rangle \), and \( \langle \Delta w_t, \zeta_3(3) \rangle \). It seems that the time series for each group is stationary and has its own characteristics. For example, the time series for workers in the occupations requiring medium communication skills (Figure 3b) is more volatile than the others; the variance of \( \langle \Delta w_t, \zeta_2(3) \rangle \) is about 0.0221, which is bigger than the variances of \( \langle \Delta w_t, \zeta_1(3) \rangle \) (≈0.0198) and that of \( \langle \Delta w_t, \zeta_3(3) \rangle \) (≈0.02). Such information will be lost if we, as in the case \( J = 1 \), average \( \Delta w_t \) over all levels of skill.

We implement an out-of-sample prediction experiment to examine the performance of our estimators; this is done by computing the root mean squared prediction error (RMSPE) using three test sets of which starting points are respectively given by 2013/01, 2015/01, and 2017/01. The RMSPE is computed using a rolling window approach. To be more specific, if we let \( \bar{u}_h \) denote the forecasting error computed from the FIVE, F2SLSE or RIVE, then the RMSPE is given by
obtained by replacing \( \alpha \) and \( \alpha_2 \) are chosen to the number that minimizes their RMSPEs; specifically, (7.1), and \( \alpha_1 \) is chosen from the model obtained by replacing \( \Delta w_t \) (resp. \( \Delta h_t \)) with \( \Delta h_t \) and (resp. \( \Delta z_t \)) in (7.1). A similar choice rule is applied to the regularization parameter of the RIVE.

The model (7.2) is estimated when \( J = 3, 7, \) and 11. Let \( \hat{u}_{h,j(J)} \) denote the \( j \)th element of the forecasting error \( \hat{u}_{h(J)} \in \mathbb{R}^J \) computed from the 2SLSE for each \( J \). The standard RMSPE of the 2SLSE is given by \( (H^{-1} \sum_{h=1}^H \sum_{j=1}^J \hat{u}_{h,j(J)}^2)^{1/2} \), which obviously tends to increase as \( J \) gets larger. Therefore, the standard RMSPE is not directly comparable across different values of \( J \) and to that of the FIVE or F2SLSE. To resolve this issue, we instead consider the normalized RMSPE, given by \( ((JH)^{-1} \sum_{h=1}^H \sum_{j=1}^J \hat{u}_{h,j(J)}^2)^{1/2} \), which can be reasonably compared to the RMSPE of the FIVE or F2SLSE since, for each \( h \geq 1 \), (i) both \( \hat{u}_{h,j(J)} \) and \( \langle u_h, \zeta_j(J) \rangle \) are estimates of the local average of \( u_h \) over the interval \( [(j-1)/J, j/J] \) and (ii) \( \int \hat{u}_h(s)^2 ds \) may be approximated by \( J^{-1} \sum_{j=1}^J \langle u_h, \zeta_j(J) \rangle^2 \).

Estimation results are reported in Table 3. In the table, we first note that the results from our estimators and that from Benatia et al.’s (2017) RIVE are very similar to each other. Moreover, these estimators report smaller RMSPEs than those of the 2SLSE except for the case when \( J = 3 \). Even if the 2SLSE reports smaller RMSPEs when \( J = 3 \), we need to pay attention to the fact that in this case 223 different levels of skill are aggregated into only three groups, which causes a lot of information loss. Note also that the normalized RMSPE of the 2SLSE rapidly increases as we consider more finely defined skill groups. This result may be due to that, as \( J \) gets larger, (i) the number of parameters to be estimated rapidly increases (note that the coefficient matrix in (7.2) is a \( J \times J \) matrix) and (ii) the sample (cross-)covariance matrices needed to compute the 2SLSE tend to be more singular (for example, the minimal eigenvalue of the cross-covariance of \( \Delta z_t(J) \) and \( \Delta h_t(J) \) when \( J = 11 \) is only 1% of that computed when \( J = 3 \)), and thus the 2SLSE is expected to perform poorly. This result also suggests that the pre-classification of workers into a few groups according to their skills can have a significant effect on the estimation results and our interpretation of those. Therefore, the results given by Table 3 imply that our functional IV methodology can be an appealing alternative to practitioners.

Figure 4 reports the estimated effects of immigration computed from the FIVE when an inflow of immigrants are given by \( 1\{0 \leq s < 1/3\}, 1\{1/3 \leq s < 2/3\}, \) and \( 1\{2/3 \leq s < 1\} \); that is, we are here considering the case where immigrants are fully concentrated in any group of occupations.

| Test period | FIVE | F2SLSE | RIVE | 2SLSE |
|-------------|------|--------|------|--------|
|             |      |        |      |        |
| 2013/01~    | 0.1886 | 0.1889 | 0.1883 | 0.1514 |
| 2015/01~    | 0.1828 | 0.1829 | 0.1823 | 0.1432 |
| 2017/01~    | 0.1679 | 0.1678 | 0.1676 | 0.1236 |
|             | 0.1734 | 1.0399 | 2.5877 | 3.4587 |

Notes: Each cell reports the estimated (normalized) RMSPE which is computed using three test sets.
Figure 4: Estimated effects of immigration computed from the FIVE

(a) $\zeta = 1 \{0 \leq s < 1/3\}$  
(b) $\zeta = 1 \{1/3 \leq s < 2/3\}$  
(c) $\zeta = 1 \{2/3 \leq s < 1\}$

Note: The solid line reports $\zeta$. The dashed lines represent the collection of confidence intervals for the local averages of $\bar{A} \bar{H}_K \zeta$ over finely defined interval $[(m-1)/M, m/M]$, where $M = 50$. For each $m$, the interval is constructed as in (3.7) with 95% significance level by noting that the local average is given by $\langle A \bar{H}_K \zeta, \zeta_{m(M)} \rangle$, and this interval, of course, may be viewed as the confidence interval for the local average of $A \zeta$ under certain conditions (see Remark 10).

with low or medium or high communication skill intensity, but they are evenly distributed within the group. The estimation results by the F2SLSE are similar and thus omitted. Overall, Figure 4 suggests that an inflow of immigrants in particular occupations differently affects the wages of native workers depending on skill intensity scores of their occupations. For example, in Figure 4a, it seems that, if the share of immigrants increases in occupations with low communication skill intensity, then the native wages are overall positively affected, and the size of such an effect for each occupation depends on how intensively communication skills are required by the occupation. To be specific, our estimation results given by Figure 4a suggest that the wages of native workers in the occupations of which $s \in [0, 0.4]$ are the most positively affected. This is somewhat consistent with Peri and Sparber’s (2009) finding that native workers in occupations intensive in manual skills take advantages of having better-paid jobs when similarly skilled immigrants enter into the market. On the other hand, in Figure 4b, it seems that the natives in the occupations of which $s \in [0.1, 0.4]$ are negatively affected if the share of immigrants increases in the occupations of which $s \in [1/3, 2/3]$. Overall, our findings in Figure 4 reconfirm the existing evidence that an inflow of immigrants heterogeneously affects the labor market outcomes of native workers according to workers’ skills.

We lastly examine the null hypothesis $H_0 : A^* \psi = 0$ using the significance test given in Section 5. Note that $\psi$ can be any arbitrary element of $\mathcal{H}$, but here we consider only a few cases, where $\psi \in \{\zeta_{j(3)}\}_{j=1}^3$, for the purpose of illustration. Then, by testing $H_0$, we can examine whether the average of changes in the wages of native workers in the occupations of which $s \in [(j-1)/3, j/3]$ is affected by an inflow of immigrants. Similarly to the simulation experiments in Section 6, we set $D$ to $\lceil T^{1/3} \rceil$ and compute the critical values based on 10,000 Monte Carlo simulations. The testing results are reported in Table 4. In the table, we found that an inflow of immigrants significantly affects the wages of native workers who are in occupations intensive in either manual or communication skills.
### Table 4: Significance testing results

|    | $\psi$ | $\zeta_1(3)$ | $\zeta_2(3)$ | $\zeta_3(3)$ |
|----|--------|-------------|-------------|-------------|
| Test statistic | 0.00050* | 0.00038 | 0.00084** |

Notes: We use * and ** to denote rejection at 10% and 5% significance levels, respectively.

### 8 Conclusion

This paper studies the functional linear regression model and pays particular attention to the case where the explanatory variable is correlated with the disturbance term. Specifically, two estimators, which are closely related to the IV and two-stage least square estimators in the standard linear simultaneous equation model, are proposed, and they are called the FIVE and F2SLSE, respectively. We study their asymptotic properties and also propose a significance test based on those. Using our methodology, we reconfirm the existing evidence that an inflow of immigrants heterogeneously affects the labor market outcomes of natives depending on workers’ skills.

### Appendix

#### A1 Random elements of Hilbert spaces

Let $(\mathbb{S}, \mathbb{F}, \mathbb{P})$ denote the underlying probability space and let $\mathcal{H}$ be a separable Hilbert space equipped with the inner product $\langle \cdot, \cdot \rangle$ and the usual Borel $\sigma$-field. A $\mathcal{H}$-valued random variable $X$ is defined by a measurable map from $\mathbb{S}$ to $\mathcal{H}$. We say that such a random variable $X$ is integrable (resp. square-integrable) if $\mathbb{E}[\|X\|] < \infty$ (resp. $\mathbb{E}[\|X\|^2] < \infty$), where $\|\cdot\|$ is the norm induced by the inner product. If $X$ is integrable, there exists a unique element $\mathbb{E}[X] \in \mathcal{H}$ satisfying $\mathbb{E}[\langle X, \zeta \rangle] = \langle \mathbb{E}[X], \zeta \rangle$ for every $\zeta \in \mathcal{H}$. The element $\mathbb{E}[X]$ is called the expectation of $X$.

Let $Y$ be another $\mathcal{H}$-valued random variable. We let $\otimes$ denote the tensor product defined as follows: for all $\zeta_1, \zeta_2 \in \mathcal{H}$,

$$\zeta_1 \otimes \zeta_2(\cdot) = \langle \zeta_1, \cdot \rangle \zeta_2.$$  

(A1.1)

Note that $\zeta_1 \otimes \zeta_2$ is a linear map from $\mathcal{H}$ to $\mathcal{H}$. If $\mathbb{E}[\|X\|\|Y\|] < \infty$, we may well define a linear map $C_{XY}$ from $\mathcal{H}$ to $\mathcal{H}$ as follows: $C_{XY} = \mathbb{E}[(X - \mathbb{E}X) \otimes (Y - \mathbb{E}Y)]$. $C_{XY}$ is called the cross-covariance operator of $X$ and $Y$. If $X = Y$ and $X$ is square-integrable, we then may define $C_{XX}$ similarly, and this is called the covariance operator of $X$. If the cross-covariance operator of two random variables $X$ and $Y$ are a nonzero operator, $X$ is said to be correlated with $Y$.

#### A2 Local likelihood density estimation

In Section 6.2, we compute the local likelihood estimate of $\log p_t^u$ using random samples $\{s_{i,t}\}_{i=1}^n$ drawn from the distribution $p_t^u$. Consider the following log-likelihood:

$$l(s_{i,t})_{i=1}^n = \sum_{i=1}^n \log p_t(s_{i,t}) - n \left( \int p_t(v)dv - 1 \right).$$  

(A2.1)
Under some local smoothness assumptions (Loader, 1996), we can obtain a localized version of (A2.1) by approximating \( \log p_t(s) \) using polynomial functions, as follows.

\[
 l(\{s_{i,t}\}_{i=1}^{n})(s) = \sum_{i=1}^{n} w\left(\frac{s_{i,t} - s}{h_s}\right) H(s_{i,t} - s; \beta_t) - n \int w\left(\frac{v - s}{h_s}\right) \exp(H(v - s; \beta_t))dv, \tag{A2.2}
\]

where \( w(\cdot) \) is a suitable weight function, \( h_s \) is a bandwidth, and \( H(v; \beta_t) \) is polynomial in \( v \) with coefficients \( \beta_t \), i.e., \( H(v; \beta_t) = \sum_{j=0}^{q} \beta_{j,t} v^j \) for some nonnegative integer \( q \). For a fixed \( s \in [0, 1] \), let \( \hat{\beta}_t \) be the maximizer of (A2.2), then the local likelihood log-density estimate is given by \( \hat{\log} p_t(s) = \hat{\beta}_0,t \). By repeating this procedure for a find grid of points and interpolating the results as described by Loader (2006, Chapter 12), we can obtain \( \hat{\log} p_t \). In our simulation experiment in Section 6.2, \( w(\cdot) \) is set to the tricube kernel that is used in many examples given by Loader (2006), \( q = 1 \), and \( h_s \) is set to the nearest neighbor bandwidth covering 33.3% of observations (Loader, 2006, Section 2.2.1).

### A3 Additional tables

**Table A1: Simulation results \((a_1, a_2 \geq 0.6)\)**

| \( n \) | \multicolumn{4}{c}{Sparse Design} | \multicolumn{4}{c}{Exponential Design} |
|---|---:|---:|---:|---:|---:|---:|---:|---:|
| \( T \) | \(100\) | \(150\) | \(200\) | \(500\) | \(200\) | \(500\) | \(200\) | \(500\) |
| Loader’s | FIVE | 0.185 | 0.152 | 0.174 | 0.149 | 0.314 | 0.194 | 0.255 | 0.173 |
| | F2SLSE | 0.188 | 0.151 | 0.174 | 0.148 | 0.313 | 0.188 | 0.256 | 0.170 |
| | RIVE | 0.194 | 0.154 | 0.179 | 0.150 | 0.342 | 0.205 | 0.276 | 0.174 |
| | FLSE | 0.248 | 0.221 | 0.206 | 0.183 | 0.419 | 0.347 | 0.322 | 0.260 |
| Silverman’s | FIVE | 0.273 | 0.231 | 0.222 | 0.194 | 0.339 | 0.242 | 0.271 | 0.207 |
| | F2SLSE | 0.275 | 0.228 | 0.225 | 0.193 | 0.352 | 0.234 | 0.277 | 0.201 |
| | RIVE | 0.271 | 0.226 | 0.224 | 0.192 | 0.342 | 0.237 | 0.278 | 0.199 |
| | FLSE | 0.351 | 0.318 | 0.267 | 0.240 | 0.422 | 0.350 | 0.319 | 0.257 |

Notes: Based on 1,000 replications. Each cell reports the empirical mean squared error (MSE) of the four considered estimators: FIVE, F2SLSE, Benatia et al.’s (2017) RIVE and Park and Qian’s (2012) FLSE.

### References

Autor, D. H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, 103(5):1553–1597.

Babii, A. (2021). High-dimensional mixed-frequency IV regression. *Journal of Business & Economic Statistics*, pages 1–14.

Baker, C. R. (1973). Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186:273–289.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62(3):657–681.
Table A2: Occupations with the lowest and highest communication skill intensity in 2000 (denoted $s$)

| Four occupations with the lowest $s$ | $s$     |
|-------------------------------------|---------|
| Pressing machine operators (clothing) | 0.0010  |
| Construction Trades                 | 0.0035  |
| Machine operators                    | 0.0235  |
| Garbage and recyclable material collectors | 0.0242  |

| Four occupations with the highest $s$ | $s$     |
|--------------------------------------|---------|
| Chief executives and public administrators | 0.9926  |
| Operations and systems researchers and analysts | 0.9954  |
| Management Analysts                  | 0.9985  |
| Economists, market researchers, and survey researchers | 1.0000  |

Benatia, D., Carrasco, M., and Florens, J.-P. (2017). Functional linear regression with functional response. *Journal of Econometrics*, 201(2):269–291.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, 77(4):531–553.

Borjas, G. J., Grogger, J., and Hanson, G. H. (2011). Substitution between immigrants, natives, and skill groups. Working Paper 17461, National Bureau of Economic Research.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.

Card, D. (2009). Immigration and inequality. *American Economic Review*, 99(2):1–21.

Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics*, 170(2):383–398.

Carrasco, M., Florens, J.-P., and Renault, E. (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of Econometrics*, 6:5633–5751.

Chao, J. C. and Swanson, N. R. (2005). Consistent estimation with a large number of weak instruments. *Econometrica*, 73(5):1673–1692.

Chen, C., Guo, S., and Qiao, X. (2022). Functional linear regression: Dependence and error contamination. *Journal of Business & Economic Statistics*, 40(1):444–457.

Conway, J. B. (1994). *A Course in Functional Analysis*. Springer, New York.

Crambes, C. and Mas, A. (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli*, 19(5B):2627–2651.

Delicado, P. (2011). Dimensionality reduction when data are density functions. *Computational Statistics & Data Analysis*, 55(1):401–420.

Dustmann, C., Frattini, T., and Preston, I. P. (2013). The effect of immigration along the distribution of wages. *The Review of Economic Studies*, 80(1):145–173.

Dustmann, C., Schönberg, U., and Stuhler, J. (2016). The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives*, 30(4):31–56.
Egozcue, J. J., Díaz-Barrero, J. L., and Pawlowsky-Glahn, V. (2006). Hilbert space of probability density functions based on Aitchison geometry. Acta Mathematica Sinica, 22(4):1175–1182.
Flood, S., King, M., Rodgers, R., Ruggles, S., and Warren, J. R. (2020). Integrated public use microdata series, current population survey: Version 8.0.
Florens, J.-P. and Van Bellegem, S. (2015). Instrumental variable estimation in functional linear models. Journal of Econometrics, 186(2):465–476.
Glitz, A. (2012). The labor market impact of immigration: A quasi-experiment exploiting immigrant location rules in Germany. Journal of Labor Economics, 30(1):175–213.
Gohberg, I., Goldberg, S., and Kaashoek, M. (2013). Classes of Linear Operators Vol. I. Birkhäuser.
Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. Annals of Statistics, 35(1):70 – 91.
Horváth, L. and Kokoszka, P. (2012). Inference for Functional Data with Applications. Springer, New York.
Hron, K., Menafoglio, A., Templ, M., Hrušová, K., and Filzmoser, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. Computational Statistics & Data Analysis, 94:330–350.
Hu, B. and Park, J. Y. (2016). Econometric analysis of functional dynamics in the presence of persistence. Mimeo, Indiana University.
Ibukiyama, T. and Kaneko, M. (2014). The Euler–Maclaurin summation formula and the Riemann zeta function. In Bernoulli Numbers and Zeta Functions, pages 65–74. Springer, Tokyo.
Imaizumi, M. and Kato, K. (2018). PCA-based estimation for functional linear regression with functional responses. Journal of Multivariate Analysis, 163:15–36.
Kokoszka, P., Miao, H., Petersen, A., and Shang, H. L. (2019). Forecasting of density functions with an application to cross-sectional and intraday returns. International Journal of Forecasting, 35(4):1304–1317.
Llull, J. (2018). Immigration, wages, and education: A labour market equilibrium structural model. The Review of Economic Studies, 85(3):1852–1896.
Loader, C. R. (1996). Local likelihood density estimation. Annals of Statistics, 24(4):1602–1618.
Loader, C. R. (2006). Local Regression and Likelihood. Springer, New York.
Mas, A. (2007). Weak convergence in the functional autoregressive model. Journal of Multivariate Analysis, 98(6):1231–1261.
Newey, W. K. and Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. Econometrica, 77(3):687–719.
Ottaviano, G. I. P. and Peri, G. (2012). Rethinking the effect of immigration on wages. Journal of the European Economic Association, 10(1):152–197.
Park, J. Y. and Qian, J. (2012). Functional regression of continuous state distributions. Journal of Econometrics, 167(2):397–412.
Peri, G. and Sparber, C. (2009). Task specialization, immigration, and wages. American Economic Journal: Applied Economics, 1(3):135–169.
Petersen, A. and Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Annals of Statistics*, 44(1):183–218.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York.

Seo, W.-K. and Beare, B. K. (2019). Cointegrated linear processes in Bayes Hilbert space. *Statistics & Probability Letters*, 147:90 – 95.

Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2):121–142.

Sharpe, J. and Bollinger, C. R. (2020). Who competes with whom? Using occupation characteristics to estimate the impact of immigration on native wages. *Labour Economics*, 66:101902.

Silverman, B. W. (1998). *Density estimation for statistics and data analysis*. Routledge, New York.

Staudenmayer, J. and Buonaccorsi, J. P. (2005). Measurement error in linear autoregressive models. *Journal of the American Statistical Association*, 100(471):841–852.

Sun, X., Du, P., Wang, X., and Ma, P. (2018). Optimal penalized function-on-function regression under a reproducing kernel Hilbert space framework. *Journal of the American Statistical Association*, 113(524):1601–1611.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.

Walker, A. M. (1960). Some consequences of superimposed error in time series analysis. *Biometrika*, 47(1/2):33–43.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, 33(6):2873–2903.

Zhang, C., Kokoszka, P., and Petersen, A. (2021). Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis*, 43(1):30–52.
Supplementary Material on “Functional instrumental variable regression with an application to estimating the impact of immigration on native wages”

Sections S1-S3 of this supplementary material provide mathematical proofs of the main results given in Sections 3-5, respectively. In Section S4, strong consistency results of the FIVE and the F2SLSE, which are mentioned in Sections 3.2 and 4.1, are provided.

S1 Appendix to Section 3 on “Functional IV estimator"

S1.1 Appendix to Section 3.2

Proof of Theorem 1

Note that

$$\hat{A} = \hat{C}_{yy} \hat{C}_{xz} (\hat{C}_{xz}^* \hat{C}_{xz})^{-1} = A \hat{\Pi}_K + \hat{C}_{uz} \hat{C}_{xz} (\hat{C}_{xz}^* \hat{C}_{xz})^{-1},$$

(S1.1)

where \( \hat{\Pi}_K = \sum_{j=1}^{K} \hat{f}_j \otimes \hat{f}_j \). Since \( \| \hat{C}_{xz} (\hat{C}_{xz}^* \hat{C}_{xz})^{-1} \|_{op} \leq \alpha^{-1/2} \) and \( \| \hat{C}_{uz} \|_{HS} = O_p(T^{-1/2}) \), we find that \( \| \hat{A} - A \hat{\Pi}_K \|_{HS} \leq O_p(\alpha^{-1/2}T^{-1/2}) \). Thus the proof becomes complete if \( \| A \hat{\Pi}_K - A \|_{HS} \to 0 \) is shown. From nearly identical arguments used to derive (8.63) of Bosq (2000), it can be shown that

$$\| A \hat{\Pi}_K - A \|_{HS} \leq \sum_{j=K+1}^{\infty} \| A \hat{f}_j \|_{2} \leq \sum_{j=K+1}^{\infty} \| A f_j^* \|_{2}^2 + 2 \| A \|_{op} \| \hat{C}_{xz}^* \hat{C}_{xz} - C_{xz}^* C_{xz} \|_{op} \sum_{j=1}^{K} \tau_j,$$

(S1.2)

where \( f_j^* = \text{sgn}\{ \langle \hat{f}_j, f_j \rangle \} f_j \). Since \( A \) is Hilbert-Schmidt, \( \sum_{j=K+1}^{\infty} \| A f_j^* \|_{2}^2 \) converges in probability to zero as \( T \) gets larger (note that \( K \) diverges almost surely as \( T \to \infty \)). In addition,

$$\| \hat{C}_{xz}^* \hat{C}_{xz} - C_{xz}^* C_{xz} \|_{op} \leq \| \hat{C}_{xz}^* \|_{op} \| \hat{C}_{xz} - C_{xz} \|_{op} + \| \hat{C}_{xz}^* - C_{xz}^* \|_{op} \| C_{xz} \|_{op} = O_p(T^{-1/2}),$$

(S1.3)

which in turn implies that the second term of the right hand side of (S1.2) is \( o_p(1) \). Combining all these results, we find that \( \| A \hat{\Pi}_K - A \|_{HS} = o_p(1) \), which implies the desired result.

Proof of Theorem 2

We first show (i). For some orthonormal bases \( \{ \hat{\xi}_j \}_{j \geq 1} \) and \( \{ \xi_j \}_{j \geq 1} \), note that \( \hat{C}_{xz} \hat{C}_{xz}^* \) and \( C_{xz} C_{xz}^* \) allow the following representations: \( \hat{C}_{xz} \hat{C}_{xz}^* = \sum_{j=1}^{\infty} \hat{\lambda}_j^2 \hat{\xi}_j \otimes \hat{\xi}_j \) and \( C_{xz} C_{xz}^* = \sum_{j=1}^{\infty} \lambda_j^2 \xi_j \otimes \xi_j \). Moreover, the following can be shown: \( \hat{C}_{xz} \hat{\xi}_j = \hat{\lambda}_j \hat{f}_j, \hat{C}_{xz} \hat{f}_j = \hat{\lambda}_j \hat{\xi}_j, C_{xz} \xi_j = \lambda_j f_j, \) and \( C_{xz} f_j = \lambda_j \xi_j \). We first note that \( C_{xz} (C_{xz}^* C_{xz})^{-1} = \sum_{j=1}^{K} (\lambda_j^s)^{-1} f_j^s \otimes \xi_j^s \), where \( f_j^s = \text{sgn}\{ \langle \hat{f}_j, f_j \rangle \} f_j, \xi_j^s = \text{sgn}\{ \langle \hat{\xi}_j, \xi_j \rangle \} \xi_j \) and \( \lambda_j^s = \text{sgn}\{ \langle \hat{f}_j, f_j \rangle \} \cdot \text{sgn}\{ \langle \hat{\xi}_j, \xi_j \rangle \} \cdot \lambda_j \). Observe that

$$\| \hat{C}_{xz} \hat{C}_{xz}^* \hat{C}_{xz} \|_{op} \leq \sum_{j=1}^{K} \| (\lambda_j^s)^{-1} - \hat{\lambda}_j^{-1} \|_{op} \| f_j^s \otimes \xi_j^s \|_{op} + \sum_{j=1}^{K} \| \lambda_j^{-1} (\hat{f}_j \otimes \hat{\xi}_j - f_j^s \otimes \xi_j^s) \|_{op}.$$

(S1.4)

The first term of (S1.4) is bounded above by \( \sup_{1 \leq j \leq K} | \lambda_j^{-1} - (\lambda_j^s)^{-1} | \), where \( | \lambda_j^{-1} - (\lambda_j^s)^{-1} | = | (\lambda_j^s - \lambda_j) \lambda_j^{-1} \lambda_j^{-1} | = | (\lambda_j^s - \lambda_j) \lambda_j^{-1} (\lambda_j^s + \lambda_j)^{-1} | \). Since \( \sup_{1 \leq j \leq K} | \lambda_j^s - \lambda_j^s | \leq \| \hat{C}_{xz} C_{xz} - C_{xz} C_{xz}^* \|_{op} \)
(Bosq, 2000, Lemma 4.2) and $|\lambda_j^*||\hat{\lambda}_j\lambda_j^*| \leq |\lambda_j^*||\hat{\lambda}_j\lambda_j^* + \hat{\lambda}_j^2|$, we find that

$$
\| \sum_{j=1}^{K} ((\lambda_j^*)^{-1} - \hat{\lambda}_j^{-1}) f_j^* \otimes \xi_j^* \|_{op} \leq \sup_{1 \leq j \leq K} \| \lambda_j^* \| |\hat{\lambda}_j\lambda_j^* + \hat{\lambda}_j^2| \leq \sup_{1 \leq j \leq K} \| \lambda_j^* \| |\hat{\lambda}_j\lambda_j^*| \leq \| \hat{C}_{xz} \hat{C}_{zz} - C_{xz}^* C_{zz}^* \|_{op} \cdot \sqrt{\alpha_{K}^2}.
$$

(S1.5)

As $\|f_j - f_j^*\| \leq \tau_j \| \hat{C}_{xz} \hat{C}_{zz} - C_{xz}^* C_{zz}^* \|_{op}$ and $\| \hat{\xi}_j - \xi_j^* \| \leq \tau_j \| \hat{C}_{xz} \hat{C}_{zz} - C_{xz}^* C_{zz}^* \|_{op}$ (Bosq, 2000, Lemma 4.3),

$$
\| \sum_{j=1}^{K} \hat{\lambda}_j^{-1} (f_j \otimes \hat{\xi}_j - f_j^* \otimes \xi_j^*) \|_{op} \leq \alpha^{-1/2} \sum_{j=1}^{K} (\|f_j - f_j^*\| + \| \hat{\xi}_j - \xi_j^* \|) \\
\leq \alpha^{-1/2} \sum_{j=1}^{K} \tau_j (\| \hat{C}_{xz} \hat{C}_{zz} - C_{xz}^* C_{zz}^* \|_{op} + \| \hat{C}_{xz} \hat{C}_{zz} - C_{xz}^* C_{zz}^* \|_{op}).
$$

(S1.6)

From (S1.3), we know that $\| \hat{C}_{xz} \hat{C}_{zz} - C_{xz}^* C_{zz}^* \|_{op} = O_p(T^{-1/2})$ and $\| \hat{C}_{xz} \hat{C}_{zz} - C_{xz}^* C_{zz}^* \|_{op} = O_p(T^{-1/2})$. Moreover, it can be shown that $\lambda_{K}^2 \leq (\lambda_{K}^2 - \lambda_{K+1}^2)^{-1} \leq \tau_K \leq \sum_{j=1}^{K} \tau_j$, so the terms given in the right hand sides of (S1.5) and (S1.6) are $o_p(1)$ under our assumptions. We thus deduce from (S1.4) that

$$
\sqrt{\frac{T}{\theta_K(\zeta)}} (\hat{A} - A\tilde{P}_K) \zeta = \left( \frac{1}{\sqrt{\theta_K(\zeta)}} \sum_{t=1}^{T} z_t \otimes u_t \right) C_{xz}(\hat{C}_{xz} C_{xz})^{-1}_K \zeta + o_p(1).
$$

Let $\zeta_t = (\theta_K(\zeta))^{-1/2} [z_t \otimes u_t] C_{xz}(C_{xz} C_{xz})^{-1}_K \zeta = (\theta_K(\zeta))^{-1/2} (z_t, C_{xz}(C_{xz} C_{xz})^{-1}_K \zeta) u_t$. Then, we have

$$
\mathbb{E}[\zeta_t \otimes \tilde{\zeta}_t] = (\theta_K(\zeta))^{-1/2} (\zeta_t, C_{xz}(C_{xz} C_{xz})^{-1}_K \zeta) C_{uu} (\zeta_t, C_{uu} \psi, \psi). \quad (S1.7)
$$

by Assumption 1.(c). Thus, under Assumption 1, $\{\langle \zeta_t, \psi \rangle \}_{t \geq 1}$ is a real-valued martingale difference sequence for any $\psi \in \mathcal{H}$. By the standard central limit theorem for such a sequence, we have

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \langle \zeta_t, \psi \rangle \overset{d}{\rightarrow} N(0, C_{uu}). \quad (S1.8)
$$

Let $\tilde{\zeta}_T = T^{-1/2} \sum_{t=1}^{T} \zeta_t$. If we show that there exists an orthonormal basis $\{e_j\}_{j \geq 1}$ satisfying

$$
\lim_{n \rightarrow \infty} \sup_{T} \mathbb{P} \left( \sum_{j=n+1}^{\infty} \langle \tilde{\zeta}_T, e_j \rangle^2 > m \right) = 0 \quad (S1.9)
$$

for every $m > 0$, then (S1.8) implies that $\tilde{\zeta}_T \overset{d}{\rightarrow} N(0, C_{uu})$ (van der Vaart and Wellner, 1996, Theorem 1.8.4). To show this, let $\{e_j\}_{j \geq 1}$ be the eigenfunctions of $C_{uu}$ and define $\mathcal{L}_n = \sum_{j=n+1}^{\infty} e_j \otimes e_j$. Then,

$$
\mathbb{E} \left[ \sum_{j=n+1}^{\infty} \langle \tilde{\zeta}_T, e_j \rangle^2 \right] \leq \frac{1}{T \theta_K(\zeta)} \sum_{t=1}^{T} \mathbb{E} \left[ \langle z_t, C_{xz}(C_{xz} C_{xz})^{-1}_K \zeta \rangle^2 \| \mathcal{L}_n u_t \|^2 \right] = \sum_{j=n+1}^{\infty} \langle C_{uu} e_j, e_j \rangle^2, \quad (S1.10)
$$
where the equality follows from that \( \{u_t\}_{t \geq 1} \) is a martingale difference sequence (with respect to \( \mathcal{F}_{t-1} \)). Since \( C_{uu} \) is Hilbert-Schmidt, the right hand side of (S1.10) converges to zero as \( n \to \infty \). Combining this result with Markov’s inequality, we find that for any \( m > 0 \), \( \mathbb{P}(\sum_{j=n+1}^{\infty} (\zeta_j^2, \ell_j)^2 > m) \leq m^{-1} \sum_{j=n+1}^{\infty} (\zeta_j^2, \ell_j)^2 \), from which (S1.9) immediately follows. Thus, the desired result is obtained.

(ii) follows from that \( \|\hat{C}_{zz} - C_{zz}\|_{op} \) and \( \|\hat{C}_{zz}(\hat{C}_{xz}\hat{C}_{xz})^{-1} - C_{zz}(C_{xz}C_{xz})^{-1}\|_{op} \) are all \( o_p(1) \). \( \square \)

### S1.2 Appendix to Section 3.3

We first state a useful lemma and then provide our proofs of the main results given in Section 4.

**Lemma S1.** Suppose that Assumptions 2.(a) and 2.(b) hold and \( \delta > 1 \). Then the following hold.

(i) \( \sum_{\ell \neq j} \frac{\lambda_{2j}^2 \ell^{-\delta}}{(\lambda_j^2 - \lambda_2^2)^2} \leq O(j^{\rho + 2 - \delta}) \), (ii) \( \sum_{\ell \neq j} \frac{\lambda_{2j}^2 \ell^{-\delta}}{(\lambda_j^2 - \lambda_2^2)^2} \leq O(j^{\rho + 2 - \delta}) \).

**Proof of Lemma S1.** We only show (i), because the remaining result can be obtained in a similar manner. As in Imaizumi and Kato (2018), we can choose \( j_0 \geq 1 \) and \( C > 1 \) large enough so that \( \lambda_0^2 / [j/C] \leq 1/2 \) and \( \lambda_2^2 / [j/C] \leq 1/2 \) for all \( j \geq j_0 \), where \( [\cdot] \) denotes the floor function. In addition, because of Assumption 2.(b) we have \( (\lambda_j^2 - \lambda_2^2)^2 \geq O(1) j^{-2\rho - 2}(j - \ell)^2 \) for \( \ell \neq j \) and \( |j/C| < \ell < [Cj] \) (Imaizumi and Kato 2018, p. 29). Using these, for \( \delta > 1 \), we find that

\[
\sum_{\ell = [j/C] + 1}^{\infty} \frac{\lambda_{2\ell}^2 \ell^{-\delta}}{\left(\lambda_j^2 - \lambda_{2j}^2\right)^2} \leq 4 \lambda_j^{-2} \sum_{\ell = [j/C] + 1}^{\infty} \ell^{-\delta} \leq O(j^\rho).
\]

Moreover, by using the inequality \( \lambda_{2j}^2 \leq |\lambda_j^2 - \lambda_2^2| + \lambda_2^2 \) and the property stated at the beginning of the proof, we have

\[
\sum_{\ell = [j/C] + 1, \neq j}^{\infty} \frac{\lambda_{2\ell}^2 \ell^{-\delta}}{\left(\lambda_j^2 - \lambda_{2j}^2\right)^2} \leq \sum_{\ell = [j/C] + 1, \neq j}^{\infty} \frac{\ell^{-\delta}}{\left|\ell - j\right|^2} + \sum_{\ell = [j/C] + 1, \neq j}^{\infty} \frac{\lambda_{2\ell}^2 \ell^{-\delta}}{\left(\lambda_j^2 - \lambda_{2j}^2\right)^2} \leq j^{1+\rho} \sum_{\ell = [j/C] + 1, \neq j}^{\infty} \frac{\ell^{-\delta}}{\left|\ell - j\right|^2} + j^{2+\rho} \sum_{\ell = [j/C] + 1, \neq j}^{\infty} \frac{\ell^{-\delta}}{\left|\ell - j\right|^2} \leq O(j^{2+\rho - \delta}),
\]

where the last two inequalities are obtained by using the fact that \( |\ell - j| < \ell \) and \( \ell^{-\delta + 1} \leq (j/C)^{-\delta + 1} \leq C^{1-\delta} \) for all \( \ell > [j/C] + 1 \). \( \square \)
Proof of Theorem 3

We first note that $\alpha \leq \lambda_K^2 = \tilde{\lambda}_K^2 - \lambda_K^2 + \lambda_K^2 \leq \|\tilde{C}^* \tilde{C} - C^* C\|_{op} + c_o K^{-\rho} \leq O_p(T^{-1/2}) + c_o K^{-\rho}$ and $\alpha^{-1} T^{-1/2} = o(1)$. These imply that

$$K \leq (1 + o_p(1)) \alpha^{-1/\rho}.$$

(S1.11)

Using the fact that $\lambda_K^2 \geq \sum_{l=j}^{\infty} (\lambda_l^2 - \lambda_{l+1}^2) \geq \rho^{-1} c_o^{-1} j^{-\rho} \text{and} \text{ } \alpha^{-1} T^{-1/2} = o(1)$, we also find that

$$(c_o \rho)^{-1} (K + 1)^{-\rho} \leq \lambda_{K+1}^2 + \frac{\lambda_{K+1}^2}{\lambda_K^2} \leq O_p(T^{-1/2}) + \alpha \leq (o_p(1) + 1) \alpha.$$

(S1.12)

We will now obtain the stochastic orders of $\|\hat{A} - \hat{A} \Pi K\|_{HS}$, $\|\hat{A} \Pi K - A \Pi K\|_{HS}$, and $\|\hat{A}(I - \Pi K)\|_{HS}$. From (S1.1), we know that $\|\hat{A} - \hat{A} \Pi K\|_{HS} \leq \|\hat{C} u z \|_{HS} \|\hat{C} u z (\hat{C}^* \hat{C} - K^2)\|_{op} \leq O_p(\alpha^{-1/2} T^{-1/2})$. Moreover, we have

$$\|\hat{A}(I - \Pi K)\|_{HS}^2 = \sum_{\ell=K+1}^{\infty} \sum_{j=1}^{\infty} \|\langle A f_\ell, \xi_j \rangle \|_{2}^2 \leq c_o \sum_{\ell=K+1}^{\infty} \sum_{j=1}^{\infty} \|\ell^{-2\gamma} j^{-2\gamma} \leq O((K + 1)^{-2\gamma}) \leq O_p(\alpha^{-1/2} T^{-1/2})$$

(S1.13)

where the first inequality immediately follows from Assumption 2(c), and the second and third inequalities follow from (S1.12), Assumption 2, and the Euler-Maclaurin summation formula for the Riemann zeta-function (see e.g., (5.6) of Ibukiyama and Kaneko, 2014). We then focus on the remaining term $\|\hat{A} \Pi K - A \Pi K\|_{HS}$. Note that

$$\|\hat{A} \Pi K - A \Pi K\|_{HS}^2 \leq 2 \sum_{j=1}^{K} \|\hat{f}_j \otimes A(\hat{f}_j - f_j^*)\|_{HS}^2 + 2 \sum_{j=1}^{K} \|\hat{f}_j - f_j^* \otimes A f_j^*\|_{HS}^2.$$

(S1.14)

For the moment, we note that

$$\|\hat{f}_j - f_j^*\|_{2}^2 = O_p(j^2 T^{-\gamma}),$$

(S1.15)

$$\|A(\hat{f}_j - f_j^*)\|_{2}^2 = O_p(T^{-1})(j^2 - 2\gamma + j^{\rho+2-2\gamma}).$$

(S1.16)

These will be proved below after discussing the main result of interest. Specifically, from (S1.16), we can show the second term in (S1.14) satisfies that

$$\|\sum_{j=1}^{K} \|\hat{f}_j - f_j^* \otimes A f_j^*\|_{HS}^2 = \sum_{\ell=1}^{\infty} \sum_{j=1}^{K} \|\langle A f_j, f_\ell \rangle (\hat{f}_j - f_j^*)\|_{2}^2 \leq \sum_{\ell=1}^{\infty} \left( \sum_{j=1}^{K} \|\langle A f_j, f_\ell \rangle\|_{op} \right) \|\hat{f}_j - f_j^*\|_{2}^2$$

$$\leq \sum_{\ell=1}^{\infty} \ell^{-2\gamma} \left( \sum_{j=1}^{K} j^{-\gamma} \|\hat{f}_j - f_j^*\|_{2}^2 \right) = O_p(T^{-1}) \left( \sum_{j=1}^{K} j^{-1} \right)^2$$

$$= \begin{cases} O_p(T^{-1}) & \text{if } \gamma > 2, \\ O_p(T^{-1} \max \{\log^2 \alpha^{-1}, \alpha^{(2\gamma-4)/\rho} \}) & \text{if } \gamma \leq 2, \end{cases}$$

(S1.17)

where the first equality follows from the properties of the Hilbert-Schmidt norm and the remaining
relationships follow from (S1.15), Assumption 2, and the fact that $\sum_{j=1}^{K} j^{1-\varsigma} = O_p(1)$ if $\varsigma > 2$ and $\sum_{j=1}^{K} j^{1-\varsigma} = O_p(\max\{\log \alpha^{-1}, \alpha^{(-2)/\rho}\})$ otherwise. Similarly, the first term in (S1.14) satisfies that

$$
\|\sum_{j=1}^{K} \hat{f}_j \otimes A(\hat{f}_j - f_j)\|_{HS}^2 = \sum_{j=1}^{K} \|A(\hat{f}_j - f_j)\|_{2}^2 = O_p(T^{-1}) \sum_{j=1}^{K} (j^{2-2\varsigma} + j^{\rho+2-2\varsigma})
$$

$$
= \begin{cases} 
O_p(T^{-1}) & \text{if } \varsigma > \rho/2 + 3/2, \\
O_p(T^{-1} \max\{\log \alpha^{-1}, \alpha^{(2\varsigma-\rho-3)/\rho}\}) & \text{if } \varsigma \leq \rho/2 + 3/2,
\end{cases} 
$$

(S1.18)

where (S1.16) is used to establish the second equality. Since $2\varsigma - \rho - 3 < 2\varsigma - 4$, $\alpha \log \alpha^{-1} = o(1)$, and $\alpha \log^2 \alpha^{-1} = o(1)$, (3.8) may be deduced from (S1.13), (S1.17) and (S1.18).

**Proofs of (S1.15) and (S1.16):** We first show (S1.15). Note that for each $j$,

$$
\hat{f}_j - f_j^s = \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_\ell^2)^{-1} \langle (\hat{C}_{xz}^* \hat{C}_{xz} - C_{xz}^* C_{xz}) \hat{f}_j, f_\ell^s \rangle f_\ell^s + \langle \hat{f}_j - f_j^s, f_\ell^s \rangle f_\ell^s. 
$$

(S1.19)

Then, using the arguments used to derive (4.48) of Bosq (2000) and the expansion of $\langle \hat{f}_j - f_j^s, f_\ell^s \rangle$ that was used to derive (S1.19), it can be shown that

$$
\|\hat{f}_j - f_j^s\|^2 \leq 4 \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\hat{C}_{xz}^* \hat{C}_{xz} - C_{xz}^* C_{xz}) \hat{f}_j, f_\ell \rangle^2. 
$$

(S1.20)

Since $\hat{C}_{xz}^* \hat{C}_{xz} - C_{xz}^* C_{xz} = (\hat{C}_{xz}^* - C_{xz}^*) \hat{C}_{xz} + C_{xz}^* (\hat{C}_{xz} - C_{xz})$, the sum given in (S1.20) satisfies that

$$
\sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\hat{C}_{xz}^* \hat{C}_{xz} - C_{xz}^* C_{xz}) \hat{f}_j, f_\ell \rangle^2 
\leq 2 \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\hat{C}_{xz}^* - C_{xz}^*) \hat{\lambda}_j \hat{\xi}_j, f_\ell \rangle^2 + 2 \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\hat{C}_{xz} - C_{xz}) \hat{f}_j, \lambda_\ell \xi_\ell \rangle^2. 
$$

(S1.21)

The second term of (S1.21) satisfies that

$$
\sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \langle (\hat{C}_{xz} - C_{xz}) \hat{f}_j, \lambda_\ell \xi_\ell \rangle^2 
\leq 2 \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \langle (\hat{C}_{xz} - C_{xz}) (\hat{f}_j - f_j^s), \xi_\ell \rangle^2 + 2 \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \langle (\hat{C}_{xz} - C_{xz}) f_j^s, \xi_\ell \rangle^2 
\leq 2\Delta_1 ||f_j - f_j^s||^2 + 2 \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \langle (\hat{C}_{xz} - C_{xz}) f_j^s, \xi_\ell \rangle^2, 
$$

(S1.22)
where $\Delta_{1j} = \sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-2} \hat{\lambda}^2_\ell \| \hat{C}_{xz} - C_{xz} \|^2_{op}$. Similarly, for the first term of (S1.21), we have

$$
\sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-2} \hat{\lambda}^2_\ell \langle (\hat{C}_{xz}^* - C_{xz}^*), \xi_j, f_\ell \rangle^2
\leq \sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-2} \hat{\lambda}^2_\ell \| (\hat{C}_{xz}^* - C_{xz}^*) \xi_j, f_\ell \|^2 + (\hat{\lambda}^2_j - \lambda^2_\ell)^{-1} \langle (\hat{C}_{xz}^* - C_{xz}^*) \xi_j, f_\ell \rangle^2
\leq 2 \sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-2} \hat{\lambda}^2_\ell \langle (\hat{C}_{xz}^* - C_{xz}^*) (\xi_j - \xi^*_j), f_\ell \rangle^2 + 2 \sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-2} \hat{\lambda}^2_\ell \langle (\hat{C}_{xz}^* - C_{xz}^*) \xi^*_j, f_\ell \rangle^2
+ 2 \sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-1} \langle (\hat{C}_{xz}^* - C_{xz}^*) (\xi_j - \xi^*_j), f_\ell \rangle^2 + 2 \sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-1} \langle (\hat{C}_{xz}^* - C_{xz}^*) \xi^*_j, f_\ell \rangle^2
\leq 2(\Delta_{1j} + \Delta_{2j}) \| \xi_j - \xi^*_j \|^2 + 2 \sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-2} \hat{\lambda}^2_\ell \langle (\hat{C}_{xz}^* - C_{xz}^*) \xi^*_j, f_\ell \rangle^2, \quad (S1.23)
$$

where $\Delta_{2j} = \max_{\ell \neq j, 1 \leq k \leq K} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-1} \| \hat{C}_{xz} - C_{xz} \|^2_{op}$, and the second inequality simply follows from the decomposition $\xi_j = (\xi_j - \xi^*_j) + \xi^*_j$ and the fact that $(a + b)^2 \leq 2a^2 + 2b^2$ for $a, b \in \mathbb{R}$. Let

$$
\Delta_{3j} = \Delta_{3j,1} + \Delta_{3j,2}, \quad (S1.24)
$$

where

$$
\Delta_{3j,1} = \sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-2} \hat{\lambda}^2_\ell \langle (\hat{C}_{xz} - C_{xz}), f_\ell^\ast, \xi_j \rangle^2, \quad \Delta_{3j,2} = \sum_{\ell \neq j} (\hat{\lambda}^2_j - \lambda^2_\ell)^{-2} \hat{\lambda}^2_\ell \langle (\hat{C}_{xz}^* - C_{xz}^*) \xi^*_j, f_\ell \rangle^2. \quad (S1.25)
$$

We then deduce from (S1.20)-(S1.25) that

$$
\| \hat{f}_j - f_j^\ast \|^2 \leq 16 \Delta_{1j} \| \hat{f}_j - f_j^\ast \|^2 + 16(\Delta_{1j} + \Delta_{2j}) \| \xi_j - \xi^*_j \|^2 + 16 \Delta_{3j}. \quad (S1.26)
$$

A similar bound of $\| \hat{\xi}_j - \xi_j^\ast \|^2$ can be obtained from nearly identical arguments to derive (S1.26), from which the following can be deduced with a little algebra:

$$
\| \hat{f}_j - f_j^\ast \|^2 \leq \frac{16(1 + 16 \Delta_{2j})}{(1 - 16 \Delta_{1j})^2 - 16^2(\Delta_{1j} + \Delta_{2j})^2} \Delta_{3j}. \quad (S1.27)
$$

From (S1.12), the condition $\alpha^{-1} = o(T^{\rho/(2\rho+2)})$, and similar arguments used to derive (A.3) of Inaizumi and Kato (2018), we find that

$$
P\{ |\hat{\lambda}^2_j - \lambda^2_\ell | \geq |\lambda^2_j - \lambda^2_\ell | / \sqrt{2}, \text{for} \ j = 1, \ldots, K \text{ and} \ \ell \neq j \} \to 1. \quad (S1.27)
$$

Because of (S1.11), (S1.27), and the condition $\alpha^{-1} = o(T^{\rho/(2\rho+2)})$, we first find that $\Delta_{1j} \leq O_p(T^{-1}) \max_{\ell \neq j, 1 \leq k \leq K} |\hat{\lambda}^2_j - \lambda^2_\ell |^{-2} \leq O_p(T^{-1} K^{2\rho+2}) \leq O_p(T^{-1} \alpha^{-2(\rho+2)/\rho}) = o_p(1)$. Similarly, from (S1.11), (S1.27) and Assumption 2.(b), we also deduce that $\Delta_{2j} \leq O_p(K^{\rho+1} \| \hat{C}_{xz}^* - C_{xz} \|^2_{op}) \leq O_p(\alpha^{-(\rho+1)/\rho} T^{-1}) = o_p(1)$. We thus find that

$$
\| \hat{f}_j - f_j^\ast \|^2 \leq 16(1 + o_p(1)) \Delta_{3j}. \quad (S1.28)
$$
We now focus on $\Delta_{3j}$. The following can be deduced:

$$|\Delta_{3j,1}| \leq O_p(1) \sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \langle (\hat{C}_{xz} - C_{xz}) f_{\ell s}, \xi_\ell \rangle^2 \leq O_p(j^2 T^{-1});$$

this may be deduced from (S1.27), Assumption 2, Lemma S1(i) and the fact that

$$T \mathbb{E}[\langle (\hat{C}_{xz} - C_{xz}) f_{j s}, \xi_\ell \rangle^2] \leq \sum_{s=0}^T \mathbb{E}[v_t(j, \ell)v_{t-s}(j, \ell)] \leq O(1)\mathbb{E}[\langle x_t, f_j \rangle^2(\xi_t)] \leq O(1)\mathbb{E}[\|x_t\|_2^2] \leq O(1)\lambda_j \lambda_\ell,$$

where the second inequality follows from Assumption 2.(d). To obtain the third inequality, note that $\mathbb{E}[\langle x_t, f_j \rangle^2(\xi_t)] = \mathbb{E}[\langle (\hat{C}_{xz} - C_{xz}) f_{j s}, \xi_\ell \rangle^2] \langle \langle x_t, f_j \rangle^2\| (\xi_t) x_t \rangle$ and then apply the Cauchy-Schwarz inequality and Assumption 2.(d). In a similar manner, we also find that $|\Delta_{3j,2}| \leq O_p(j^2 T^{-1})$, and thus $|\Delta_{3j}| \leq O_p(j^2 T^{-1})$. Combining this result with (S1.28), we find that the desired result (S1.15) holds.

We next show (S1.16). Note that

$$\mathcal{A}(\hat{f}_j - f^s_j) = \sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-1} \langle (\hat{C}_{xz}^s \hat{C}_{xz} - C_{xz}^s C_{xz}) \hat{f}_j, f^s_\ell \rangle \mathcal{A} f^s_\ell + \langle \hat{f}_j - f^s_j, f^s_\ell \rangle \mathcal{A} f^s_\ell,$$  \hspace{1cm} (S1.29)

where $\|\langle \hat{f}_j - f^s_j, f^s_\ell \rangle \mathcal{A} f^s_\ell \|^2 \leq O_p(T^{-1}) j^{2-2k}$. For each $j = 1, \ldots, K$, the first term in (S1.29) is bounded above as follows:

\begin{align*}
    (\sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-1} \langle (\hat{C}_{xz}^s \hat{C}_{xz} - C_{xz}^s C_{xz}) \hat{f}_j, f^s_\ell \rangle \mathcal{A} f^s_\ell)^2 &\leq O(1)(\sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-1} \langle (\hat{C}_{xz}^s \hat{C}_{xz} - C_{xz}^s C_{xz}) \hat{f}_j, f^s_\ell \rangle)^2 \\
    &\leq O(1)(\sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-1} \langle (\hat{C}_{xz} - C_{xz}) \hat{f}_j, \xi_\ell \rangle)^2 + O(1)(\sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-1} \langle (\hat{C}_{xz}^s \hat{C}_{xz} - C_{xz}^s C_{xz}) \hat{f}_j, \xi_\ell \rangle)^2 \\
    &\leq O(1)\|\hat{C}_{xz} - C_{xz}\|_{op}^2 \left(\sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \ell^{-2k} + \sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-2} \lambda_j^2 \ell^{-2k}\right) \\
    &\leq O_p(T^{-1}) \left(\sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-2} \lambda_\ell^2 \ell^{-2k} + \sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-2} \lambda_j^2 \ell^{-2k} + \sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-2} \lambda_j^2 \ell^{-2k}\right) \\
    &\leq O_p(T^{-1}) \left(j^{6-2k+2} + O_p(T^{-1/2})\lambda_j^{2} + 1\right) \left(\sum_{\ell \neq j} (\lambda_j^2 - \lambda_\ell^2)^{-2} \lambda_j^2 \ell^{-2k}\right) \\
    &\leq (1 + O_p(T^{-1/2}a^{-1}))O_p(T^{-1} j^{6-2k+2}) \leq (1 + o_p(1))O_p(T^{-1} j^{6-2k+2}), \hspace{1cm} (S1.30)
\end{align*}

where the second inequality holds since Assumption 2, and the remaining inequalities follow from Hölder’s inequality, Lemma S1(i)-(ii), and the fact that $\|\hat{C}_{xz} - C_{xz}\|_{op} = O_p(T^{-1/2})$. From (S1.29) and (S1.30), we find that (S1.16) holds. \hfill \Box
Proof of Theorem 4

The whole proof is divided into two parts.

1. **Proof of the convergence results:** For the subsequent discussion, we first need to obtain an upper bound of \( \langle \hat{f}_j - f_j^s, \zeta \rangle \). From the expansion given in (S1.19), we have

\[
\langle \hat{f}_j - f_j^s, \zeta \rangle = \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_j^2)^{-1} (\langle \hat{C}_{xx}^s \hat{C}_{xx} - C_{xx}^s C_{xx} \rangle \hat{f}_j, f_\ell) (f_\ell, \zeta) + \langle \hat{f}_j - f_j^s, f_j^s \rangle (f_j^s, \zeta). \tag{S1.31}
\]

Note that the second term in (S1.31) satisfies that \( (\langle \hat{f}_j - f_j^s, f_j^s \rangle (f_j^s, \zeta))^2 \leq O_p(T^{-1}j^{-2\delta_\zeta+2}) \), for all \( j = 1, \ldots, K \). Moreover, the first term in (S1.31) satisfies the following:

\[
\sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_j^2)^{-1} (\langle \hat{C}_{xx}^s \hat{C}_{xx} - C_{xx}^s C_{xx} \rangle \hat{f}_j, f_\ell) (f_\ell, \zeta))^2 \leq O(1) |\langle \hat{\lambda}_j^2 - \lambda_j^2 |^{-1} \ell^{-\delta_\zeta} (\langle \hat{C}_{xx}^s \hat{C}_{xx} - C_{xx}^s C_{xx} \rangle \hat{f}_j, f_\ell) \rangle|^2
\]

\[
\leq O(1) |\langle \hat{\lambda}_j^2 - \lambda_j^2 |^{-1} \ell^{-\delta_\zeta} (\langle \hat{C}_{xx}^s \hat{C}_{xx} - C_{xx}^s C_{xx} \rangle \hat{f}_j, f_\ell) \rangle|^2 + O(1) |\langle \hat{\lambda}_j^2 - \lambda_j^2 |^{-1} \ell^{-\delta_\zeta} \lambda_j (\langle \hat{C}_{xx}^s \hat{C}_{xx} - C_{xx}^s C_{xx} \rangle \hat{f}_j, f_\ell) \rangle|^2
\]

\[
\leq O(1) \| \hat{C}_{xx} - C_{xx} \|_{op}^2 \left( \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_j^2)^{-2} \lambda_j^2 \ell^{-2\delta_\zeta} + \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_j^2)^{-2} \lambda_j^2 \ell^{-2\delta_\zeta} \right)
\]

\[
\leq O_p(T^{-1}) \left( j^{\rho - 2\delta_\zeta + 2} \right) + O_p(T^{-1/2}) \lambda_j^{-2} + 1 \sum_{\ell \neq j} (\hat{\lambda}_j^2 - \lambda_j^2)^{-2} \lambda_j^2 \ell^{-2\delta_\zeta} \right)
\]

\[
\leq O_p(T^{-1}) j^{\rho - 2\delta_\zeta + 2} \left( 1 + O_p(T^{-1/2}) j^{\rho} \right), \tag{S1.32}
\]

where the first inequality follows from the assumption on \( \langle f_\ell, \zeta \rangle \) and the remaining inequalities are deduced from similar arguments that are used to obtain (S1.30). Because \( T^{-1/2} j^{\rho} \leq O_p(T^{-1/2} \alpha^{-1}) = o_p(1) \) uniformly in \( j = 1, \ldots, K \), we conclude that, for each \( j = 1, \ldots, K \),

\[
\langle \hat{f}_j - f_j^s, \zeta \rangle^2 = O_p(T^{-1}) j^{-2\delta_\zeta+2} + O_p(T^{-1}) j^{-2\delta_\zeta+2+\rho} (1 + o_p(1)). \tag{S1.33}
\]

We next show the following:

\[
\| \hat{C}_{xx} (\hat{C}_{xx}^s \hat{C}_{xx})_K^{-1} \zeta - C_{xx} (C_{xx}^s C_{xx})_K^{-1} \zeta \|_{op} = o_p(1). \tag{S1.34}
\]

Given that \( \| \hat{C}_{xx} - C_{xx} \|_{op} = o_p(1) \), the asymptotic results given in Theorem 2.(i) and 2.(ii) are deduced without difficulty from (S1.34) and similar arguments used in our proofs given in Section S1. From the decomposition given in (S1.4), it can be deduced that the desired result (S1.34) is established if the following terms are all \( o_p(1) \): 

\[
\| \sum_{j=1}^{K} (\lambda_j^s)^{-1} - \hat{\lambda}_j^{-1} \| (f_j^s, \zeta) (f_j^s, \zeta), \| \sum_{j=1}^{K} (\lambda_j^s)^{-1} - \hat{\lambda}_j^{-1} \| (f_j^s, \zeta) (\hat{\xi}_j, \zeta), \| \sum_{j=1}^{K} (\lambda_j^s)^{-1} (f_j^s, \zeta) (\xi_j - \xi_j^s), \| \sum_{j=1}^{K} (\lambda_j^s)^{-1} (f_j^s, \zeta) (\hat{\xi}_j, \zeta) \|, \| \sum_{j=1}^{K} (\lambda_j^s)^{-1} (f_j^s, \zeta) (\hat{f}_j - f_j^s, \zeta) \|.
\]
First, note that
\[
\| \sum_{j=1}^{K} ((\lambda_j^p)^{-1} - \hat{\lambda}_j^{-1}) (f_j^p, \zeta) \|_2^2 = \sum_{j=1}^{K} ((\lambda_j^p)^{-1} - \hat{\lambda}_j^{-1})^2 (f_j^p, \zeta)^2 \leq \sum_{j=1}^{K} \frac{(\lambda_j^2 - \hat{\lambda}_j^2)^2}{\lambda_j^2 (\lambda_j^2 + \lambda_j^p \lambda_j^s)} c_j j^{-2\delta_c}
\]
where the last inequality follows from (S1.11). In addition, using (S1.33) and the arguments used to derive (S1.35), it can be shown that
\[
\| \sum_{j=1}^{K} ((\lambda_j^p)^{-1} - \hat{\lambda}_j^{-1}) (f_j^p, \zeta) \|_2^2 = \sum_{j=1}^{K} ((\lambda_j^p)^{-1} - \hat{\lambda}_j^{-1})^2 (f_j^p, \zeta)^2 \leq O_p(T^{-1} \alpha^{-1} \sum_{j=1}^{K} j^{2p-2\delta_c}) \leq O_p(T^{-1} \max\{\alpha^{-3p(2\delta_c+1)}/\rho, \alpha^{-1}(1+p)/\rho\})
\]
(S1.35)

Note that \( T^{-1} \alpha^{-1}(1+p)/\rho = o(1) \) and \( T^{-2} \alpha^{-4p-2\delta_c+3}/\rho = T^{-1} \alpha^{-3p-2\delta_c+1}/\rho \) by (3.10). These imply that the terms given in (S1.35) and (S1.36) are all \( o_p(1) \). We also find that
\[
\| \sum_{j=1}^{K} (\lambda_j^p)^{-1} (f_j^p, \zeta) (\hat{\xi}_j - \xi_j^p) \| \leq \sum_{j=1}^{K} (\lambda_j^p)^{-1} (f_j^p, \zeta) (\hat{\xi}_j - \xi_j^p) \| \leq O_p(T^{-1/2} \max\{\alpha^{-1}/\rho, \alpha^{(\delta_c+2)/2}/\rho\}) = o_p(1)
\]
where the first inequality follows from the triangular inequality and the second inequality is deduced from Assumption 2.(b) and the fact that \( \| \hat{\xi}_j - \xi_j^p \|^2 \leq O_p(T^{-1} j^2) \) for \( j = 1, \ldots, K \). This can be obtained from nearly identical arguments used to derive (S1.15). The remaining inequalities are deduced since \( 2\delta_c > 1 \) and (S1.11) holds. It only remains to show that \( \| \sum_{j=1}^{K} (\lambda_j^p)^{-1} (f_j^p, \zeta) (\hat{\xi}_j - \xi_j^p) \|^2 \leq \alpha(2p-2\delta_c+3)/\rho, \alpha^{-1}/\rho = o_p(1). \)
This can be obtained from Assumption 2.(a) and (S1.33); specifically, we observe that
\[
\| \sum_{j=1}^{K} (\lambda_j^p)^{-1} (f_j^p, \zeta) (\hat{\xi}_j - \xi_j^p) \|^2 \leq O_p(T^{-1}) \max\{\alpha^{-2p-2\delta_c+3}/\rho, \alpha^{-1}/\rho\} = o_p(1)
\]
(S1.38)

Hence, from the results given in (S1.35)-(S1.38), (S1.34) is established.

2. Analysis on the regularization bias: Next, we focus on the regularization bias term, \( \| \mathcal{A}(\hat{\Pi}_K - \Pi_K) \| \). For convenience, we let
\[
\mathcal{A}(\hat{\Pi}_K - \Pi_K) = F_1 + F_2 + F_3 + F_4,
\]
S9
where $F_4 = \mathcal{A}(\Pi_K - I)\zeta$,

$$F_1 = \sum_{j=1}^{K} (\hat{f}_j - f_j^\circ, \zeta)\mathcal{A}(\hat{f}_j - f_j^\circ), \quad F_2 = \sum_{j=1}^{K} (f_j^\circ, \zeta)\mathcal{A}(\hat{f}_j - f_j^\circ), \quad F_3 = \sum_{j=1}^{K} (\hat{f}_j - f_j^\circ, \zeta)\mathcal{A}f_j^\circ,$$

and thus $F_1 + F_2 + F_3 = \mathcal{A}(\tilde{\Pi}_K - \Pi_K)\zeta$. Then, by using (S1.16) and (S1.33), we find that

$$\|F_1\| \leq \sum_{j=1}^{K} |\langle \hat{f}_j - f_j^\circ, \zeta \rangle| \|\mathcal{A}(\hat{f}_j - f_j^\circ)\| \leq O_p(T^{-1/2}) \sum_{j=1}^{K} j^\rho - \zeta - \delta_\zeta + 2 \leq O_p(T^{-1/2}) \sum_{j=1}^{K} j^\rho - \zeta - \delta_\zeta + 1,$$

where the last bound is obtained because $\alpha^{-1} = o(T^{\rho/(2\rho+2)})$. In a similar manner, it can be shown that

$$\|F_2\| \leq \sum_{j=1}^{K} |\langle f_j^\circ, \zeta \rangle| \|\mathcal{A}(\hat{f}_j - f_j^\circ)\| \leq O_p(T^{-1/2}) \sum_{j=1}^{K} j^\rho - \zeta - \delta_\zeta + 1,$$

and

$$\|F_3\| \leq \sum_{j=1}^{K} |\langle \hat{f}_j - f_j^\circ, \zeta \rangle| \|\mathcal{A}f_j^\circ\| \leq O_p(T^{-1/2}) \sum_{j=1}^{K} j^\rho - \zeta - \delta_\zeta + 1.$$

Therefore, $\|F_1\|$, $\|F_2\|$ and $\|F_3\|$ are bounded by the following quantity:

$$O_p(T^{-1/2}) \sum_{j=1}^{K} j^\rho - \zeta - \delta_\zeta + 1 \leq \begin{cases} O_p(T^{-1/2}) & \text{if } \rho/2 + 2 < \zeta + \delta_\zeta, \\ O_p(T^{-1/2} \max\{\log \alpha^{-1}, \alpha^{-(\rho/2 - \zeta - \delta_\zeta + 2)/\rho}\}) & \text{if } \rho/2 + 2 \geq \zeta + \delta_\zeta. \end{cases}$$

Lastly, the following can be shown:

$$\|F_4\|^2 \leq \sum_{j=K+1}^{\infty} \|\langle f_j, \zeta \rangle\mathcal{A}f_j\|^2 \leq \sum_{j=K+1}^{\infty} j^{-2\delta_\zeta} \|\mathcal{A}f_j\|^2 = O_p\left( \sum_{j=K+1}^{\infty} j^{-2\delta_\zeta - 2\zeta} \right) \leq O_p((2\zeta + 2\delta_\zeta - 1)/\rho).$$

This concludes the proof. \(\square\)

## S2 Appendix to Section 4 on “Extension: the functional two-stage least square estimator"

We first provide a useful lemma that is related to our discussion on the F2SLSE in Section 4.

**Lemma S2.** There exist unique bounded linear operators $\mathcal{R}_{xz}$ and $\mathcal{R}_{yz}^*$ satisfying the following:

$$\mathcal{C}_{zz}^{1/2} \mathcal{R}_{xz} \mathcal{C}_{zz}^{1/2} = C_{xz}, \quad \mathcal{R}_{xz}[\text{ran } \mathcal{C}_{zz}^{1/2}]^\perp = \{0\}, \quad \mathcal{R}_{xz}^* [\text{ran } \mathcal{C}_{zz}^{1/2}]^\perp = \{0\},$$

$$\mathcal{C}_{yy}^{1/2} \mathcal{R}_{yz} \mathcal{C}_{zz}^{1/2} = C_{yz}, \quad \mathcal{R}_{yz}[\text{ran } \mathcal{C}_{zz}^{1/2}]^\perp = \{0\}, \quad \mathcal{R}_{yz}^* [\text{ran } \mathcal{C}_{zz}^{1/2}]^\perp = \{0\},$$

where $V^\perp$ denotes the orthogonal complement of $V \subset \mathcal{H}$.

Lemma S2 directly follows from Theorem 1 of Baker (1973). From the properties of $\mathcal{R}_{xz}$ (resp. $\mathcal{R}_{yz}$) given above, it can be understood as the cross-correlation operator of $x_t$ and $z_t$ (resp. $y_t$ and $z_t$). Let $C_{zz}^{1/2}$ be defined by $\sum_{j=1}^{\infty} \mu_j^{-1/2} g_j \otimes g_j$, which is not a bounded linear operator since $\mu_j \to 0$.
as \( j \to \infty \). However, even with this property, we know as a direct consequence of Lemma S2 that 
\( C_{zz}^{-1/2}C_{xz} \) and \( C_{zy}C_{zz}^{-1/2} \) are well defined bounded linear operators and they are respectively given by

\[
C_{zz}^{-1/2}C_{xz} = R_{xz}C_{xz}^{1/2} \quad \text{and} \quad C_{zy}C_{zz}^{-1/2} = C_{zy}^{1/2}R_{yz}^*.
\]

We thus find that

\[
C_{yz}C_{zz}^{-1}C_{xz} = C_{yz}C_{zz}^{-1/2}C_{xz} = C_{yz}^{1/2}R_{yz}^* R_{yz}C_{zz}^{1/2} =: \mathcal{P},
\]

\[
C_{xx}^{-1}C_{xz} = C_{xx}^{-1/2}C_{xz} = C_{xx}^{1/2}R_{xx}^* R_{xx}C_{xx}^{1/2} =: \mathcal{Q}.
\]

As desired, \( \mathcal{P} \) and \( \mathcal{Q} \) are uniquely defined elements of \( \mathcal{L}_H \), and moreover, they are compact since \( C_{xx}^{1/2} \) and \( C_{yy}^{1/2} \) are compact.

### S2.1 Appendix to Section 4.1

#### Proof of Theorem 5

Since \( \|\hat{\mathcal{C}}_{zu}\|_{\text{HS}} = O_p(T^{-1/2}) \), we find that

\[
\|\hat{\mathcal{A}} - \mathcal{A}\Pi_{K_2}\|_{\text{HS}} \leq \|\hat{\mathcal{C}}_{uz}\|_{\text{HS}} \|\hat{\mathcal{C}}_{zz}\|_{K_2}^{-1/2}\|_{\text{op}} \|\hat{\mathcal{C}}_{zz}\|_{K_2}^{-1/2}\|_{\text{op}} \leq O_p(\alpha_1^{-1/4} \alpha_2^{-1/4} T^{-1/2}). \tag{S2.1}
\]

Thus, \( \|\hat{\mathcal{A}} - \mathcal{A}\Pi_{K_2}\|_{\text{HS}} = o_p(1) \), and hence it suffices to show that \( \|\mathcal{A}\Pi_{K_2} - \mathcal{A}\|_{\text{HS}}^2 = o_p(1) \). Note that

\[
\|\mathcal{A}\Pi_{K_2} - \mathcal{A}\|_{\text{HS}}^2 \leq \sum_{j=K_2+1}^{\infty} \|\mathcal{A}h_j^s\|_{\text{op}} + |\mathcal{R}|, \tag{S2.2}
\]

where \( h_j^s = \text{sgn}\{\langle \hat{h}_j, h_j \rangle\} h_j \) and \( \mathcal{R} = \sum_{j=K_2+1}^{\infty} (\|\mathcal{A}\hat{h}_j\|_{\text{HS}}^2 - \|\mathcal{A}h_j^s\|_{\text{HS}}^2) \). Since \( \mathcal{A} \) is Hilbert-Schmidt, the first term of (S2.2) is \( o_p(1) \). It thus only remains to verify that \( |\mathcal{R}| = o_p(1) \). To show this, we first deduce the following inequality from similar arguments used to derive the equation between (8.62) and (8.63) of Bosq (2000):

\[
|\mathcal{R}| \leq 2\|\mathcal{A}\|_{\text{op}}^2 \sum_{j=1}^{K_2} \|\hat{h}_j - h_j^s\|. \tag{S2.3}
\]

We find that \( \mathcal{Q}\hat{h}_j - \nu_j\hat{h}_j = (\mathcal{Q} - \hat{\mathcal{Q}})\hat{h}_j + (\nu_j - \nu_j)\hat{h}_j \). Hence, by Lemma 4.2 of Bosq (2000),

\[
\|\mathcal{Q}\hat{h}_j - \nu_j\hat{h}_j\| \leq 2\| \hat{\mathcal{Q}} - \mathcal{Q} \|_{\text{op}}. \tag{S2.4}
\]

Moreover, it can be shown from similar arguments used in the proof of Lemma 4.3 of Bosq (2000) that \( \|\hat{h}_j - h_j^s\| \leq \frac{\tau_{2,j}}{2} \|\mathcal{Q}\hat{h}_j - \nu_j\hat{h}_j\| \), which, combined with (S2.4), implies that

\[
\|\hat{h}_j - h_j^s\| \leq \frac{\tau_{2,j}}{2} \| \hat{\mathcal{Q}} - \mathcal{Q} \|_{\text{op}}. \tag{S2.5}
\]

We then deduce from (S2.3) and (S2.5) that

\[
|\mathcal{R}| \leq 2\|\mathcal{A}\|_{\text{op}}^2 \left( \sum_{j=1}^{K_2} \tau_{2,j} \right) \| \hat{\mathcal{Q}} - \mathcal{Q} \|_{\text{op}}. \tag{S2.6}
\]
Then the following can be shown:

\[ \| \hat{\mathcal{Q}} - \mathcal{Q} \|_{\text{op}} \leq \| \mathcal{B} \hat{\Pi}_{K_1} \hat{C}_{zz} \hat{\Pi}_{K_1} B^* - B \mathcal{C}_{zz} B^* \|_{\text{op}} + \| S \|_{\text{op}}, \]  

(S2.7)

where \( S = \hat{C}_{vz} \hat{\Pi}_{K_1} B^* + \mathcal{B} \hat{\Pi}_{K_1} \hat{C}_{vz} + \hat{C}_{vz}(\hat{C}_{zz})^{-1} \hat{C}_{vz} \). Let \( \Pi_{K_1} = \sum_{j=1}^{K_1} g_j \otimes g_j \) and let \( \mathcal{T} = \mathcal{B}(\mathcal{I} - \Pi_{K_1}) \mathcal{C}_{zz}(\mathcal{I} - \Pi_{K_1}) B^* \).

We further find that

\[
\| \mathcal{B} \hat{\Pi}_{K_1} \hat{C}_{zz} \hat{\Pi}_{K_1} B^* - B \mathcal{C}_{zz} B^* \|_{\text{op}} \leq \| \mathcal{B} \hat{\Pi}_{K_1} \hat{C}_{zz} \hat{\Pi}_{K_1} B^* - \mathcal{B} \Pi_{K_1} \mathcal{C}_{zz} \Pi_{K_1} B^* \|_{\text{op}} + \| \mathcal{T} \|_{\text{op}}
\]

\[
\leq \left( \| \mathcal{B} \|_{\text{op}}^2 \sum_{j=1}^{K_1} (\hat{\mu}_j - \mu_j) \hat{g}_j \otimes \hat{g}_j \|_{\text{op}} + 2 \| \mathcal{B} \|_{\text{op}}^2 \sum_{j=1}^{K_1} \mu_j \| \hat{g}_j - g^*_j \| + \| \mathcal{T} \|_{\text{op}} \right)
\]

\[
\leq \left( \| \mathcal{B} \|_{\text{op}}^2 + 2 \| \mathcal{B} \|_{\text{op}}^2 \sum_{j=1}^{K_1} \mu_j \tau_1,j \right) \| \hat{C}_{zz} - \mathcal{C}_{zz} \|_{\text{op}} + \| \mathcal{T} \|_{\text{op}} \leq O_p \left( \sum_{j=1}^{K_1} \mu_j \tau_1,j \right) \| \hat{C}_{zz} - \mathcal{C}_{zz} \|_{\text{op}} + \| \mathcal{T} \|_{\text{op}}.
\]

(S2.8)

where \( g^*_j = \text{sgn}\{ (\hat{g}_j, g_j) \} g_j \). From (S2.6), (S2.7) and (S2.8), the following is established:

\[
| \mathcal{R} | \leq \left\{ O_p \left( \sum_{j=1}^{K_1} \mu_j \tau_1,j \right) \| \hat{C}_{zz} - \mathcal{C}_{zz} \|_{\text{op}} + \| S \|_{\text{op}} + \| \mathcal{T} \|_{\text{op}} \right\} O_p \left( \sum_{j=1}^{K_2} \tau_2,j \right).
\]

(S2.9)

Since \( \| \hat{C}_{vz} \|_{\text{HS}} = O_p(T^{-1/2}) \), \( \| \hat{\Pi}_{K_1} \|_{\text{op}} \leq 1 \) and \( \| (\hat{\mathcal{C}}_{zz})^{-1} \|_{\text{op}} \leq \alpha_1^{-1/2} \), we have

\[
\| S \|_{\text{op}} \leq O_p(T^{-1/2}) + O_p(\alpha_1^{-1/2} T^{-1}).
\]

(S2.10)

Note that \( \| \hat{C}_{zz} - \mathcal{C}_{zz} \|_{\text{op}} = O_p(T^{-1/2}) \) and \( \| \mathcal{T} \|_{\text{op}} \sum_{j=1}^{K_2} \tau_2,j = o_p(1) \) (which follows from the fact that \( \| \mathcal{T} \|_{\text{op}} \leq \| \mathcal{B} \|_{\text{op}}^2 \sum_{j=1}^{\infty} \mu_j \)). Combining these results with (S2.9) and (S2.10), we find that \( | \mathcal{R} | \leq O_p(T^{-1/2} (\sum_{j=1}^{K_1} \mu_j \tau_1,j)(\sum_{j=1}^{K_2} \tau_2,j)) + O_p((T^{-1/2} + \alpha_1^{-1/2} T^{-1}) \sum_{j=1}^{K_2} \tau_2,j) \). Given that \( \alpha_1^{-1} T^{-1} \rightarrow 0 \) and \( \sum_{j=1}^{K_2} \tau_2,j \leq o_p(\sum_{j=1}^{K_1} \mu_j \tau_1,j)(\sum_{j=1}^{K_2} \tau_2,j) \), it may be easily deduced that \( | \mathcal{R} | = o_p(1) \) as desired.

Proof of Theorem 6

To show (i), we will first verify that

\[
\| (\hat{\mathcal{C}}_{zz})^{-1} \hat{\mathcal{C}}_{zz} \hat{\mathcal{Q}}_{K_2}^{-1} - (\mathcal{C}_{zz})^{-1} \mathcal{C}_{zz} \mathcal{Q}_{K_2}^{-1} \|_{\text{op}} \leq E_1 + E_2 + E_3 = o_p(1),
\]

(S2.11)

where \( E_1, E_2 \) and \( E_3 \) are defined as follows:

\[
E_1 = \| ((\hat{\mathcal{C}}_{zz})^{-1} - (\mathcal{C}_{zz})^{-1}) \hat{\mathcal{C}}_{zz} \hat{\mathcal{Q}}_{K_2}^{-1} \|_{\text{op}}, \quad E_2 = \| (\mathcal{C}_{zz})^{-1} (\hat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}) \hat{\mathcal{Q}}_{K_2}^{-1} \|_{\text{op}}, \quad E_3 = \| (\mathcal{C}_{zz})^{-1} \mathcal{C}_{zz} (\hat{\mathcal{Q}}_{K_2}^{-1} - \mathcal{Q}_{K_2}^{-1}) \|_{\text{op}}.
\]
Note first that \( \| \hat{C}_{zz} \hat{Q}_{K2}^{-1} \|_{op} = O_p(\alpha_2^{-1/2}) \), and thus

\[
E_1 \leq O_p(\alpha_2^{-1/2}) \left( \| \sum_{j=1}^{K_1} (\mu_j^{-1} - \bar{\mu}_j^{-1}) g_j^s \otimes g_j^s \|_{op} + \| \sum_{j=1}^{K_1} \bar{\mu}_j^{-1} (\bar{g}_j \otimes \bar{g}_j - g_j^s \otimes g_j^s) \|_{op} \right),
\]

where \( g_j^s = \text{sgn} \{ \langle \bar{g}_j, g_j \rangle \} g_j \). We then find that

\[
\| \sum_{j=1}^{K_1} (\mu_j^{-1} - \bar{\mu}_j^{-1}) g_j^s \otimes g_j^s \|_{op} \leq \sup_{1 \leq j \leq K_1} |\bar{\mu}_j^{-1} - \mu_j^{-1}| = \sup_{1 \leq j \leq K_1} \left| \frac{\bar{\mu}_j - \mu_j}{\mu_j \bar{\mu}_j} \right| \leq \alpha_1^{-1/2} \mu_{K_1}^{-1} \| \hat{C}_{zz} - C_{zz} \|_{op}
\]

and

\[
\| \sum_{j=1}^{K_1} \bar{\mu}_j^{-1} (\bar{g}_j \otimes \bar{g}_j - g_j^s \otimes g_j^s) \|_{op} \leq 2 \alpha_1^{-1/2} \sum_{j=1}^{K_1} \| \bar{\mu}_j - g_j^s \| \leq 2 \alpha_1^{-1/2} \| \hat{C}_{zz} - C_{zz} \|_{op} \sum_{j=1}^{K_1} \tau_{1,j}.
\]

Since \( \mu_{K_1}^{-1} \leq \sum_{j=1}^{K_1} \tau_{1,j} = O_p(\alpha_1^{1/2} T^{1/2}) \) and \( \| \hat{C}_{zz} - C_{zz} \|_{op} = O_p(T^{-1/2}) \), the right hand sides of (S2.12) and (S2.13) are \( o_p(1) \), and hence \( E_1 = o_p(1) \). Since \( \| (C_{zz})_{K1}^{-1} \|_{op} \leq \mu_{K_1}^{-1} \), \( \| \hat{Q}_{K2}^{-1} \|_{op} \leq \alpha_2^{-1/2} \), and \( \| \hat{C}_{zz} - C_{zz} \|_{op} = O_p(T^{-1/2}) \), we also find that

\[
E_2 \leq \mu_{K_1}^{-1} O_p(\alpha_2^{-1/2} T^{-1/2}) \leq O_p(\alpha_2^{-1/2} T^{-1/2}) \sum_{j=1}^{K_1} \tau_{1,j} = o_p(1).
\]

Given that \( \| (C_{zz})_{K1}^{-1} C_{zz} \|_{op} \leq \| B^* \|_{op} = O_p(1) \), it remains to show that \( \| \hat{Q}_{K2}^{-1} - Q_{K2}^{-1} \|_{op} = o_p(1) \) since this implies \( E_3 = o_p(1) \) (and thus the desired result (S2.11) is obtained). To show this, we first note that

\[
\| \hat{Q}_{K2}^{-1} - Q_{K2}^{-1} \|_{op} \leq \| \sum_{j=1}^{K_2} (\nu_j^{-1} - \tilde{\nu}_j^{-1}) h_j \otimes h_j \|_{op} + \| \sum_{j=1}^{K_2} \tilde{\nu}_j^{-1} (\hat{h}_j \otimes \hat{h}_j - h_j \otimes h_j) \|_{op}.
\]

Let \( S = \hat{C}_{zz} \hat{\Pi}_K B^* + B \hat{\Pi}_K \hat{C}_{zz} + \hat{C}_{zz} (\hat{C}_{zz})_{K1}^{-1} \hat{C}_{zz} \). We then deduce from our proof of Theorem 5 that

\[
\| \hat{Q} - Q \|_{op} \leq O_p \left( \sum_{j=1}^{K_1} \mu_j \tau_{1,j} \right) \| \hat{C}_{zz} - C_{zz} \|_{op} + \| S \|_{op} + \| T \|_{op}.
\]

As in (S2.12), it can be shown that \( \| \sum_{j=1}^{K_2} (\nu_j^{-1} - \tilde{\nu}_j^{-1}) h_j \otimes h_j \|_{op} \leq \alpha_2^{-1/2} \nu_{K_2}^{-1} \| \hat{Q} - Q \|_{op} \), hence

\[
\| \sum_{j=1}^{K_2} (\nu_j^{-1} - \tilde{\nu}_j^{-1}) h_j \otimes h_j \|_{op} \leq \alpha_2^{-1/2} \nu_{K_2}^{-1} \left( O_p \left( \sum_{j=1}^{K_1} \mu_j \tau_{1,j} \right) \| \hat{C}_{zz} - C_{zz} \|_{op} + \| S \|_{op} + \| T \|_{op} \right).
\]
We find that
which is repeatedly used in our proofs of the main results.

Define

From these results, (S2.10) and the fact that
We also deduce the following from (S2.5), (S2.13) and (S2.14):

\[ \| \sum_{j=1}^{K_2} \hat{h}_j \otimes \hat{h}_j - h_j \otimes h_j \|_{op} \leq 2 \alpha_2^{-1/2} \sum_{j=1}^{K_2} \tau_{2,j} \left( O_p \left( \sum_{j=1}^{K_1} \mu_j \tau_{1,j} \right) \right) \| \hat{C}_{zz} - C_{zz} \|_{op} + \| S \|_{op} + \| T \|_{op}. \]  

(S2.16)

We find that \( \nu_{K_2}^{-1} \leq \sum_{j=1}^{K_2} \tau_{2,j} \leq O_p(\sum_{j=1}^{K_1} \mu_j \tau_{1,j})(\sum_{j=1}^{K_2} \tau_{2,j}) = o_p(\alpha_2^{1/2} T^{1/2}) \), which follows from that \( \mu_1^{-1} \tau_{1,1}(\sum_{j=1}^{K_1} \mu_j \tau_{1,j}) \geq 1 \). Moreover, note that

\[ \sum_{j=1}^{K_1} \mu_j \tau_{1,j} \leq \nu_1 \nu_{K_2}^{-1} \sum_{j=1}^{K_2} \mu_j \tau_{1,j} \leq \nu_1 \left( \sum_{j=1}^{K_1} \mu_j \tau_{1,j} \right) \left( \sum_{j=1}^{K_2} \tau_{2,j} \right) = o_p(\alpha_2^{1/2} T^{1/2}), \]

and

\[ \nu_{K_2}^{-1} \| T \|_{op} \leq \sum_{j=1}^{K_2} \tau_{2,j} \| T \|_{op} \leq \| B \|_{op}^2 \left( \sum_{j=K_1+1}^{\infty} \mu_j \right) \left( \sum_{j=1}^{K_2} \tau_{2,j} \right) = o_p(\alpha_2^{1/2}). \]

From these results, (S2.10) and the fact that \( \| \hat{C}_{zz} - C_{zz} \|_{op} = O_p(T^{-1/2}) \), we may deduce that the right hand sides of (S2.15) and (S2.16) are all \( o_p(1) \), and thus \( E_3 = o_p(1) \) and (S2.11) holds.

We thus know that

\[ \sqrt{T}(\hat{A} - A \hat{\Pi}_{K_2}) \xi = \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T z_t \otimes u_t \right) (C_{zz})_{K_1}^{-1} C_{xx} Q_{K_2}^{-1} \xi + o_p(1). \]

Define \( \xi_t = (\phi_{K_2}(\xi))^{-1/2}[z_t \otimes u_t](C_{zz})_{K_1}^{-1} C_{xx} Q_{K_2}^{-1} \xi \) and let \( \tilde{\xi}_T = T^{-1/2} \sum_{t=1}^T \xi_t \). Then from nearly identical arguments used to derive (S1.7), (S1.8) and (S1.9), we find that, for any \( \psi \in \mathcal{H} \) and \( m > 0 \),

\[ T^{-1/2} \sum_{t=1}^T \langle \xi_t, \xi_t \rangle \xrightarrow{d} N(0, \langle C_{uu} \psi, \psi \rangle) \text{ and } \lim_{n \to \infty} \limsup_{T} P(\sum_{j=n+1}^\infty (\tilde{\xi}_T, \ell_j)^2 > m) = 0, \]

where \( \{\ell_j\}_{j \geq 1} \) denote the eigenfunctions of \( C_{uu} \). Hence (i) is established.

Given that \( \| Q_{K_2}^{-1} - Q_{K_2}^{-1} \|_{op} = o_p(1) \), (ii) is immediately deduced.

\[ \square \]

**S2.2 Appendix to Section 4.2**

Throughout this section, we define

\[ Q_{K_1} = C_{xx}^*(C_{zz})_{K_1}^{-1} C_{xx}, \]

which is repeatedly used in our proofs of the main results.

**Proof of Theorem 7**

We will show the following:

\[ K_2 \leq (1 + o_p(1))\alpha_2^{-1/\rho_2}, \]  

(S2.17)

\[ (c_0 p)^{-1}(K_2 + 1)^{-\rho_2} \leq (1 + o_p(1))\alpha_2, \]  

(S2.18)

\[ \| \hat{h}_j - h_j \|_2^2 \leq O_p(\alpha_2) j^2, \]  

(S2.19)

\[ \| A(\hat{h}_j - h_j) \|_2^2 \leq O_p(\alpha_2) j^{2-2\omega} + O_p(d_T) j^{\rho_2 - 2\omega + 2}, \]  

(S2.20)
where $h_j^*$ is defined as in our proof of Theorem 5 and $d_T$ is defined by

\[ d_T = \alpha_1(4\nu + \rho_\nu - 2)/\rho_\nu + T^{-1} \max \{ \alpha_1^{-1/\rho_\nu}, \alpha_1^{-1}(\rho_\nu - 2\nu + 3)/\rho_\nu \}. \tag{S2.21} \]

Note that

\[ \| \tilde{\Pi} - A\|_{HS} \leq \| \tilde{\Pi} - A\tilde{\Pi}_{K_2}\|_{HS} + \| A\tilde{\Pi}_{K_2} - A\Pi_{K_2}\|_{HS} + \| (I - \Pi_{K_2})A\|_{HS}, \]

where $\| \tilde{\Pi} - A\tilde{\Pi}_{K_2}\|_{HS} = O_p(\alpha_1^{-1/4}\alpha_2^{-1/4}T^{-1/2})$ as is shown in (S2.1). Using (S2.18), we also find that

\[ \| (I - \Pi_{K_2})A\|_{HS}^2 = \sum_{\ell = K_2 + 1}^\infty \| Ah_\ell\|^2 \leq O(1) \sum_{\ell = K_2 + 1}^\infty \sum_{j = 1}^\infty \| h_{\ell j} - h_{j}^* \|^2 \leq O_p(\alpha_2(2\nu - 1)/\rho_\nu). \tag{S2.22} \]

We next focus on the remaining term $\| A\tilde{\Pi}_{K_2} - A\Pi_{K_2}\|_{HS}$. Note that

\[ 2^{-1}\| A\tilde{\Pi}_{K_2} - A\Pi_{K_2}\|_{HS}^2 \leq \| \sum_{j = 1}^{K_2} \tilde{h}_j \otimes A(\tilde{h}_j - h_j^*)\|_{HS}^2 + \| \sum_{j = 1}^{K_2} (\tilde{h}_j - h_j^*) \otimes Ah_j\|_{HS}^2. \tag{S2.23} \]

We know from (S2.19) and (S2.23) that

\[ \| \sum_{j = 1}^{K_2} (\tilde{h}_j - h_j^*) \otimes Ah_j\|_{HS}^2 \leq \sum_{\ell = 1}^\infty \| \sum_{j = 1}^{K_2} \langle Ah_j, h_\ell \rangle (\tilde{h}_j - h_j^*) \|^2 \leq \sum_{\ell = 1}^\infty \left( \sum_{j = 1}^{K_2} \| \langle Ah_j, h_\ell \rangle \| \| \tilde{h}_j - h_j^* \| \right)^2 \]

\[ \leq O_p(\alpha_1) \left( \sum_{j = 1}^{K_2} j^{1/\omega} \right)^2 = \begin{cases} O_p(\alpha_1) & \text{if } \nu > 2, \\ O_p(\alpha_1 \max \{ \log^2 \alpha_2^{-1}, \alpha_2(2\nu - 4)/\rho_\nu \}) & \text{if } \nu \leq 2. \end{cases} \tag{S2.24} \]

Moreover, from (S2.20) and the fact that $d_T\alpha_1^{-1} = o(1)$, the following may be deduced:

\[ \| \sum_{j = 1}^{K_2} \tilde{h}_j \otimes A(\tilde{h}_j - h_j^*)\|_{HS}^2 \leq \sum_{j = 1}^{K_2} \| A(\tilde{h}_j - h_j^*)\|_{HS}^2 \leq O_p(\alpha_1) \sum_{j = 1}^{K_2} j^{2 - 2\nu} + O_p(d_T) \sum_{j = 1}^{K_2} j^{\nu - 2\nu + 2} \]

\[ \leq \begin{cases} O_p(\alpha_1) & \text{if } \nu/2 + 3/2 < \nu, \\ O_p(\alpha_1 \max \{ \log \alpha_2^{-1}, \alpha_2(2\nu - 3)/\rho_\nu \}) & \text{if } \nu/2 + 3/2 \geq \nu. \end{cases} \tag{S2.25} \]

Since $2\nu - \nu - 3 < 2\nu - 4$, $\alpha_2 \log \alpha_2^{-1} = o(1)$, and $\alpha_2 \log^2 \alpha_2^{-1} = o(1)$, (4.4) may be deduced from (S2.22), (S2.24) and (S2.25).

**Proofs of (S2.17)-(S2.20):** To obtain the desired results, we first need to discuss on $\| (\hat{Q} - Q)h_\ell \|$ and $\| \hat{Q} - Q\|_{HS}$. Note that, for any $\ell$,

\[ 2^{-1}\| (\hat{Q} - Q)h_\ell \|^2 \leq \| (\hat{Q} - Q_{K_1})h_\ell \|^2 + \| (Q_{K_1} - Q)h_\ell \|^2. \tag{S2.26} \]
The second term in (S2.26) is bounded above as follows,

\[
\| (\mathcal{Q}_{K_1} - \mathcal{Q}) h_{\ell} \|^2 = \| \sum_{j = K_1 + 1}^{\infty} \mu_j (B g_j, h_{\ell}) B g_j \|^2 \leq O(1) \sum_{j = K_1 + 1}^{\infty} \mu_j^2 \| B g_j \|^2 \sum_{j = K_1 + 1}^{\infty} j^{-2\nu} \ell^{-\rho_\nu/2} \\
\leq O(1) \mu_{K_1}^2 (K_1 + 1)^{-4\nu + 2} \ell^{-\rho_\nu/2} \leq O_p(1) \alpha_1^{(4\nu + \rho_\nu - 2)/\rho_\nu} \ell^{-\rho_\nu/2},
\]

(S2.27)

where the first inequality follows from the Hölder’s inequality and the second is obtained because \( \rho_\nu / 2 > 1 \), \( \sum_{j = K_1 + 1}^{\infty} j^{-2\nu} \leq (K_1 + 1)^{-2\nu + 1} \), and \( \mu_j^2 \leq \mu_{K_1}^2 \) for \( j > K_1 \). The last inequality is obtained using the arguments that are used to derive (S1.12). We now focus on the first term in (S2.26). Note that

\[
4^{-1} \| (\mathcal{Q} - \mathcal{Q}_{K_1}) h_{\ell} \|^2 \leq \| \tilde{C}_{vz} (\tilde{C}_{zz})_{K_1}^{-1} \tilde{C}_{vz} h_{\ell} \|^2 + \| \tilde{C}_{vz} h_{\ell} \|^2 + \| B \Pi_{K_1} \tilde{C}_{vz} h_{\ell} \|^2 \\
+ \| (B (\Pi_{K_1} \tilde{C}_{zz} \Pi_{K_1} - \Pi_{K_1} C_{zz} \Pi_{K_1}) B^*) h_{\ell} \|^2,
\]

where \( \| \tilde{C}_{vz} \Pi_{K_1} B^* h_{\ell} \|^2 \leq \ell^{-\rho_\nu/2} O_p(T^{-1}) \). Moreover, we have \( \| \tilde{C}_{vz} h_{\ell} \|^2 \leq \ell^{-\rho_\nu/2} O_p(T^{-1}) \) and \( \| \tilde{C}_{vz} (\tilde{C}_{zz})_{K_1}^{-1} \tilde{C}_{vz} h_{\ell} \|^2 \leq O_p(\alpha_1^{-1} T^{-1}) \| \tilde{C}_{vz} h_{\ell} \|^2 \leq O_p(T^{-2} \alpha_1^{-1}) \ell^{-\rho_\nu/2} \) since

\[
T \mathbb{E}[\| \tilde{C}_{vz} h_{\ell} \|^2] = T^{-1} \mathbb{E}[\| \sum_{t = 1}^{T} (v_t, h_j) z_t \|^2] \leq O(1) \mathbb{E}[\| (v_t, h_j) z_t \|^2] \leq O(1) \nu_j,
\]

(S2.28)

where the inequalities are obtained from Assumption 4; specifically, under the assumption, we have that \( \mathbb{E}[\| (v_t, h_j) z_t \|^2] \leq \mathbb{E}[\| (x_t, h_j) z_t \|^2] \leq c_\circ \| C_{zz} h_j \|^2 \leq \| \Pi_{K_1} C_{zz} \|_{\text{op}} \| (\tilde{C}_{zz})_{K_1}^{-1/2} C_{zz} h_j \|^2 \leq O(1) \nu_j \). Lastly, using the arguments used to obtain (S1.16), we can show that \( \| B (\tilde{g}_j - g_j^*) \|^2 \leq O_p(T^{-1}) (j^{2 - 2\nu} + j^{\rho_\nu + 2 - 2\nu}) \). Using this bound, we find that

\[
4^{-1} \| (B (\Pi_{K_1} \tilde{C}_{zz} \Pi_{K_1} - \Pi_{K_1} C_{zz} \Pi_{K_1}) B^*) h_{\ell} \|^2 \\
\leq \| \sum_{j = 1}^{K_1} (\tilde{\mu}_j - \mu_j) (\tilde{g}_j, B^* h_{\ell}) B g_j \|^2 + \| \sum_{j = 1}^{K_1} \mu_j (\tilde{g}_j - g_j^*, B^* h_{\ell}) B g_j \|^2 \\
+ \| \sum_{j = 1}^{K_1} \mu_j (g_j^*, B^* h_{\ell}) B (\tilde{g}_j - g_j^*) \|^2 + \| \sum_{j = 1}^{K_1} \mu_j (\tilde{g}_j - g_j^*, B^* h_{\ell}) B (\tilde{g}_j - g_j^*) \|^2 \\
\leq K_1 \| B \|_{\text{op}}^2 \| C_{zz} - C_{zz} \|_{\text{op}} \| B^* h_{\ell} \|^2 + \| \tilde{C}_{zz} \|_{\text{HS}}^2 \sum_{j = 1}^{K_1} \| \tilde{g}_j - g_j^* \|^2 \| B g_j \|^2 \\
+ \| C_{zz} \|_{\text{HS}}^2 \sum_{j = 1}^{K_1} \| B^* h_{\ell} \|^2 \| B (\tilde{g}_j - g_j^*) \|^2 + \| \tilde{C}_{zz} \|_{\text{HS}}^2 \| B^* h_{\ell} \|^2 \sum_{j = 1}^{K_1} \| \tilde{g}_j - g_j^* \|^4 \\
\leq \ell^{-\rho_\nu/2} \alpha_1^{-1/\rho_\nu} O_p(T^{-1}) + \ell^{-\rho_\nu/2} O_p(T^{-1}) \sum_{j = 1}^{K_1} j^{-2\nu + 4} \\
+ \ell^{-\rho_\nu/2} O_p(T^{-1}) \sum_{j = 1}^{K_1} (j^{-2\nu + 2} + j^{\rho_\nu - 2\nu + 2}) + \ell^{-\rho_\nu/2} O_p(T^{-2}) \sum_{j = 1}^{K_1} j^4 \\
\leq \ell^{-\rho_\nu/2} O_p(T^{-1} \max\{\alpha_1^{-1/\rho_\nu}, \alpha_1^{-1(\rho_\nu - 2\nu + 3)/\rho_\nu}\}),
\]

(S2.29)
where the second inequality is obtained by using Lemma 4.2 in Bosq (2000) and noting that
\[ \| \sum_{j=1}^{K_1} \mu_j (g_j - g_j, B^* h_t) B g_j \|^2 \leq \left( \sum_{j=1}^{K_1} \mu_j^2 \right) \left( \sum_{j=1}^{K_1} \| B g_j \|^2 (g_j - g_j, B^* h_t)^2 \right) \] holds by the Hölder's inequality. The last two inequalities are deduced from Assumption 1, Assumption 4, and the the fact that \( \| B^* h_t \|^2 \leq O(1) \ell^{-\rho_\nu/2} \) and \( \alpha_1^{-1} = o(T^{\rho_\nu/(2\rho_\nu+2)}). \) Then, from the results given in (S2.26) to (S2.29) and the definition of \( d_T \) given in (S2.21), we conclude the following: for any \( t \leq K_2, \)
\[ \| (\hat{Q} - Q) h_t \|^2 \leq \ell^{-\rho_\nu/2} O_p(d_T), \] (S2.30)
from which we also find that
\[ \| \hat{Q} - Q \|^2_{HS} = O_p(d_T). \] (S2.31)

We now verify (S2.17) and (S2.18). It can be shown without difficulty that \( d_T = O(\alpha_1). \) Given that \( \alpha_1^{-1} = o(T^{1/2}) \) and \( \alpha_2^{-1} \alpha_1^{1/2} = o(1), \) we find that \( \alpha_2^{-1} d_T^{1/2} = o(1), \) from which the following is deduced:
\[ \alpha_2 = \hat{\alpha}_2 - \nu_2 K_2 + \nu_2^2 \leq \| \hat{Q} - Q \|_{op} + c_0 K_2^{-\rho_\nu} \leq o(1) \alpha_2 + c_0 K_2^{-\rho_\nu}. \]
Then (S2.17) follows from the above. Similarly as in (S1.12), it can be shown that
\[ (c_0 \rho)^{-1} (K_2 + 1)^{-\rho_\nu} \leq \nu_{K_2+1}^2 = \nu_{K_2+1}^2 - \nu_{K_2+1}^2 + \nu_{K_2+1}^2 \leq (1 + o_p(1)) \alpha_2, \]
and thus we find that (S2.18) holds.

We then show (S2.19). To this end, it should first be noted that, under the employed assumptions, (S1.27) holds if \( \hat{\lambda}_j \) (resp. \( \lambda_j \)) is replaced by \( \tilde{\nu}_j \) (resp. \( \nu_\ell \)). Moreover, note that the eigenfunctions of \( Q^2 \) and \( \tilde{Q}^2 \) are equivalent to those of \( Q \) and \( \hat{Q} \). Therefore, by applying the arguments that are used in our proof of Theorem 3, we can show that
\[ 8^{-1} \| \hat{h}_j - h_j^\alpha \|^2 \leq \sum_{\ell \neq j} (\hat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_\ell^2 \| (\hat{Q} - Q) \hat{h}_j, h_\ell \|^2 + \sum_{\ell \neq j} (\hat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_\ell^2 \| (\hat{Q} - Q) h_j, h_\ell \|^2. \] (S2.32)
From similar arguments used to derive (S1.22), the first term in (S2.32) is bounded above as follows:
\[ \sum_{\ell \neq j} (\hat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_\ell^2 \| (\hat{Q} - Q) h_j, h_\ell \|^2 \leq 2 \Delta_{1j} \| \hat{h}_j - h_j^\alpha \|^2 + 2 \sum_{\ell \neq j} (\hat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_\ell^2 \| (\hat{Q} - Q) h_j, (\hat{Q} - Q) h_\ell \|^2, \] (S2.33)
where \( \Delta_{1j} = \sum_{\ell \neq j} (\hat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_\ell^2 \| (\hat{Q} - Q) h_\ell \|^2. \) Moreover, by using similar arguments that are used to obtain (S1.23), we can show that
\[ \sum_{\ell \neq j} (\hat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_\ell^2 \| (\hat{Q} - Q) h_j, h_\ell \|^2 \leq 2 \sum_{\ell \neq j} (\hat{\nu}_j^2 - \nu_\ell^2)^{-2} \nu_\ell^2 \| (\hat{Q} - Q) h_j, h_\ell \|^2 + 2 (\Delta_{1j} + \Delta_{2j}) \| \hat{h}_j - h_j^\alpha \|^2, \] (S2.34)
where \( \Delta_{2j} = \sum_{\ell \neq j} (\hat{\nu}_j^2 - \nu_\ell^2)^{-1} \| (\hat{Q} - Q) h_\ell \|^2. \) We then use the results given in (S2.30) and (S2.31) to obtain the following bounds of \( \Delta_{1j} \) and \( \Delta_{2j}: \) for \( j = 1, \ldots, K_2, \)
\[ \Delta_{1j} \leq O_p(d_T) \sum_{\ell \neq j} (\nu_j^2 - \nu_\ell^2)^{-2} \nu_\ell^2 \ell^{-\rho_\nu/2} \leq j^{1+\rho_\nu/2} O_p(d_T) \leq O_p(\alpha_2^{-1+\rho_\nu/(2\rho_\nu)} d_T) = o_p(1), \] (S2.35)
where the second inequality follows from Lemma S1(i). Similarly, for \( j = 1, \ldots, K_2 \),

\[
\tilde{\Delta}_{2j} \leq \max_{1 \leq \ell \leq K_2} (\nu_j^2 - \nu_{\ell}^2)^{-1} \sum_{\ell \neq j} \|(\hat{Q} - Q) h_{\ell}\|^2 \leq j^{1 + \rho_\nu} O_p(d_T) \leq O(\alpha_2^{-1} \eta \rho_\nu d_T) = o_p(1). \tag{S2.36}
\]

From (S2.32)-(S2.36), we have

\[
\|\hat{h}_j - h_j^\ast\|^2 \leq O(1)(1 + o_p(1)) \tilde{\Delta}_{3j}, \tag{S2.37}
\]

for \( j = 1, \ldots, K_2 \), where

\[
\tilde{\Delta}_{3j} = \sum_{\ell \neq j} (\hat{v}_j^2 - \nu_{\ell}^2)^{-2} \hat{v}_j^2 \langle (\hat{Q} - Q) h_j, h_\ell \rangle^2 + \sum_{\ell \neq j} (\hat{v}_j^2 - \nu_{\ell}^2)^{-2} \nu_{\ell}^2 \langle (\hat{Q} - Q) h_j, h_\ell \rangle^2. \tag{S2.38}
\]

We will analyze the above term using the decomposition \( \hat{Q} - Q = \hat{Q} - Q_{K_1} + Q_{K_1} - Q \). Note that

\[
\langle (Q_{K_1} - Q) h_j, h_\ell \rangle^2 \leq \left( \sum_{i = K_1 + 1}^\infty \mu_i^2 \langle g_i, B^* h_j \rangle^2 \right) \left( \sum_{i = K_1 + 1}^\infty \langle g_i, B^* h_\ell \rangle^2 \right) \leq O_p(\alpha_1) \ell^{1 - \rho_\nu / 2} \ell^{1 - \rho_\nu / 2}, \tag{S2.39}
\]

where the first inequality follows from the Hölder’s inequality, and the second is deduced from Assumption 4 and similar arguments used to derive (S1.12). From (S2.39) and Lemma S1(i), we find that

\[
\sum_{\ell \neq j} (\hat{v}_j^2 - \nu_{\ell}^2)^{-2} \nu_{\ell}^2 \langle (Q_{K_1} - Q) h_j, h_\ell \rangle^2 \leq O_p(\alpha_1) j^2. \tag{S2.40}
\]

Similarly, for \( j = 1, \ldots, K_2 \), we have

\[
\sum_{\ell \neq j} (\hat{v}_j^2 - \nu_{\ell}^2)^{-2} \hat{v}_j^2 \langle (Q_{K_1} - Q) h_j, h_\ell \rangle^2 \leq (O_p(d_T^{1/2})^\rho_\nu + 1)O_p(\alpha_1) j^2, \tag{S2.41}
\]

where the second inequality follows from \( |\hat{v}_j^2 - \nu_{\ell}^2| \leq \|\hat{Q}^2 - Q^2\|_{op} \leq O_p(1)\|\hat{Q} - Q\|_{op} \). Given that \( \alpha_2^{-1} d_T^{1/2} = o(1) \), (S2.41) is bounded above by \( O_p(\alpha_1) j^2 \). Next, we will obtain an upper bound of \( \langle (\hat{Q} - Q_{K_1}) h_j, h_\ell \rangle^2 \). Note that

\[
4^{-1} \langle (\hat{Q} - Q_{K_1}) h_\ell, h_j \rangle^2 \leq \langle \hat{C}_{vz}(\hat{C}_{zz})^{-1} \hat{C}_{vz} h_\ell, h_j \rangle^2 + \langle \hat{C}_{vz} \hat{\Pi}_{K_1} B^* h_\ell, h_j \rangle^2 + \langle B \hat{\Pi}_{K_1} \hat{C}_{vz} h_\ell, h_j \rangle^2 \leq \alpha_1^{-1} \nu_j \nu_{\ell} O_p(T^{-2}) + \nu_j \ell^{1 - \rho_\nu / 2} O_p(T^{-1}) + \nu_{\ell} \ell^{1 - \rho_\nu / 2} O_p(T^{-1}) + \langle B (\hat{\Pi}_{K_1} \hat{C}_{zz} \hat{\Pi}_{K_1} - \Pi_{K_1} C_{zz} \Pi_{K_1}) B^* \rangle h_\ell, h_j \rangle^2, \tag{S2.42}
\]

where the last inequality is obtained by using (S2.28). The last term in (S2.42) satisfies the
following:

\[
4^{-1} \langle (\mathcal{B}(\tilde{\Pi}_K \tilde{C}_{zz} \Pi K_1 - \Pi K_1 C_{zz} \Pi K_1) \mathcal{B}^*) h_\ell, h_j \rangle^2 \\
\leq \left( \sum_{i=1}^{K_1} (\tilde{\mu}_i - \mu_i) \langle \tilde{g}_i, \mathcal{B}^* h_\ell \rangle \langle \mathcal{B} \tilde{g}_i, h_j \rangle \right)^2 + \left( \sum_{i=1}^{K_1} \mu_i \langle \tilde{g}_i, \mathcal{B}^* h_\ell \rangle \langle \mathcal{B} \tilde{g}_i, h_j \rangle \right)^2 \\
+ \left( \sum_{i=1}^{K_1} \mu_i \langle g_i, \mathcal{B}^* h_\ell \rangle \langle \mathcal{B} (\tilde{g}_i - g_i), h_j \rangle \right)^2 \\
\leq \max_{1 \leq i \leq K_1} |\tilde{\mu}_i - \mu_i|^2 \| \tilde{\Pi}_K \mathcal{B}^* h_\ell \|^2 \| \tilde{\Pi}_K \mathcal{B}^* h_j \|^2 + \sum_{i=1}^{K_1} \mu_i^2 \| \tilde{g}_i - g_i \|^2 \| \mathcal{B}^* h_\ell \|^2 \| \tilde{\Pi}_K \mathcal{B}^* h_j \|^2 \\
+ \sum_{i=1}^{K_1} \mu_i^2 \| \tilde{g}_i - g_i \|^2 \| \mathcal{B}^* h_\ell \|^2 \| \Pi_K \mathcal{B}^* h_j \|^2 \\
\leq \ell^{-\rho_\nu/2} j^{-\rho_\nu/2} O_p(T^{-1} \alpha_1^{-1/\rho_\nu}), \tag{S2.43}
\]

where the last inequality follows from Assumptions 3 and 4. Combining the results given in (S2.42) and (S2.43), we find that

\[
4^{-1} \langle (\tilde{\mathcal{Q}} - Q_{\mathcal{K}_1}) h_\ell, h_j \rangle^2 \leq O_p(T^{-1})(\alpha_1^{-1} T^{-1} j^{-\rho_\nu/2} \ell^{-\rho_\nu/2} + 2j^{-\rho_\nu/2} \ell^{-\rho_\nu/2} + \alpha_1^{-1/\rho_\nu} j^{-\rho_\nu/2} \ell^{-\rho_\nu/2}) \\
\leq O_p(T^{-1}) \alpha_1^{-1/\rho_\nu} j^{-\rho_\nu/2} \ell^{-\rho_\nu/2}.
\]

Together with Lemma S1(i), this implies that

\[
\sum_{\ell \neq j} (\tilde{\nu}_j^2 - \nu_j^2)^{-2} \nu_j^2 \langle (\tilde{\mathcal{Q}} - Q_{\mathcal{K}_1}) h_\ell, h_j \rangle^2 \leq O_p(T^{-1} \alpha_1^{-1/\rho_\nu}) j^2, \tag{S2.44}
\]

where the last inequality follows from Lemma S1(i). In addition, from the arguments that are used to derive (S2.41) and the fact that \( \alpha_2^{-1} d_{T}^{1/2} = o(1) \), the following may be deduced: for \( j = 1, \ldots, K_2 \),

\[
\sum_{\ell \neq j} (\tilde{\nu}_j^2 - \nu_j^2)^{-2} \tilde{\nu}_j^2 \langle (\tilde{\mathcal{Q}} - Q_{\mathcal{K}_1}) h_\ell, h_j \rangle^2 \leq (O_p(\alpha_2^{-1} d_{T}^{1/2}) + 1)O_p(T^{-1} \alpha_1^{-1/\rho_\nu}) j^2 = O_p(T^{-1} \alpha_1^{-1/\rho_\nu}) j^2. \tag{S2.45}
\]

From (S2.37), (S2.38), (S2.40), (S2.41), (S2.44), (S2.45), and the decomposition \( \tilde{\mathcal{Q}} - \mathcal{Q} = \tilde{\mathcal{Q}} - Q_{\mathcal{K}_1} + Q_{\mathcal{K}_1} - \mathcal{Q} \), we conclude that

\[
\| \tilde{h}_j - h_j^* \|^2 \leq (1 + o_p(1)) \tilde{\Delta}_{3j} \leq O_p(\alpha_1) j^2 + O_p(T^{-1} \alpha_1^{-1/\rho_\nu}) j^2 \leq O_p(\alpha_1) j^2,
\]

where the last inequality follows from \( \alpha_1^{-1} = o(T^{\rho_\nu/(2\rho_\nu+2)}) \). This completes our proof of (S2.19).

Lastly, to show (S2.20), we note that

\[
A(\tilde{h}_j - h_j^*) = \sum_{\ell \neq j} (\tilde{\nu}_j^2 - \nu_j^2)^{-1} \langle (\tilde{\mathcal{Q}}^2 - Q^2) \tilde{h}_j, h_\ell \rangle A \tilde{h}_\ell + \langle \tilde{h}_j - h_j^*, h_j \rangle A h_j, \tag{S2.46}
\]
where \( \| (\hat{h}_j - h^*_j, h_j) A h_j \| \leq O_p(\alpha_1) j^{2 - 2\nu}. \) The first term in (S2.46) is bounded above as follows,

\[
\left( \sum_{\ell \neq j} (\tilde{\nu}^2_j - \nu^2_\ell) \right)^{-1} \langle (\tilde{Q} - Q) \hat{h}_j, h_\ell \rangle A h_\ell^2 \leq \left( \sum_{\ell \neq j} |\tilde{\nu}^2_j - \nu^2_\ell|^{-1} \right) \langle (\tilde{Q} - Q) \hat{h}_j, h_\ell \rangle A \| A h_\ell \| ^2
\]

\[
\leq O(1) \| \tilde{Q} - Q \| _{op}^2 \left( \sum_{\ell \neq j} (\tilde{\nu}^2_j - \nu^2_\ell)^{-2} \nu^2_\ell - 2^\nu + \sum_{\ell \neq j} (\tilde{\nu}^2_j - \nu^2_\ell)^{-2} \nu^2_\ell - 2^\nu \right)
\]

\[
\leq O_p(d_T) \left( \sum_{\ell \neq j} (\tilde{\nu}^2_j - \nu^2_\ell)^{-2} \nu^2_\ell - 2^\nu + \sum_{\ell \neq j} (\tilde{\nu}^2_j - \nu^2_\ell)^{-2} (\tilde{\nu}^2_j - \nu^2_\ell) \nu^2_\ell - 2^\nu + \sum_{\ell \neq j} (\tilde{\nu}^2_j - \nu^2_\ell)^{-2} \nu^2_\ell - 2^\nu \right)
\]

\[
\leq O_p(d_T) \left( j^{\rho_\nu - 2\nu + 2} + (O_p(d_T)^{1/2} \nu^2_\ell - 2^\nu + 1) \sum_{\ell \neq j} (\tilde{\nu}^2_j - \nu^2_\ell)^{-2} \nu^2_\ell - 2^\nu \right)
\]

\[
\leq (1 + o_p(1)) O_p(d_T) j^{\rho_\nu - 2\nu + 2}.
\]  

(S2.47)

Combining (S2.46) and (S2.47), we obtain (S2.20) as desired. □

**Proof of Theorem 8**

The whole proof is divided into two parts.

1. **Proof of the convergence results:** We need an upper bound of \( \langle \hat{h}_j - h^*_j, \zeta \rangle \), which is importantly used in the following discussion. Using the expansion in (S1.19), we find that

\[
\langle \hat{h}_j - h^*_j, \zeta \rangle = \sum_{\ell \neq j} (\tilde{\nu}^2_j - \nu^2_\ell)^{-1} \langle (\tilde{Q} - Q) \hat{h}_j, h_\ell \rangle \langle h_\ell, \zeta \rangle + \langle \hat{h}_j - h^*_j, h^*_j \rangle \langle h^*_j, \zeta \rangle.
\]

Note that \( \langle \langle \hat{h}_j - h^*_j, h^*_j \rangle \langle h^*_j, \zeta \rangle \rangle \leq O_p(\alpha_1) j^{2 - 2\delta \nu} \) for \( j = 1, \ldots, K_2 \), because of (S2.19). Moreover, using similar arguments that are used to derive (S1.32) and (S2.47), we find that

\[
\left( \sum_{\ell \neq j} (\tilde{\nu}^2_j - \nu^2_\ell)^{-1} \langle (\tilde{Q} - Q) \hat{h}_j, h_\ell \rangle \langle h_\ell, \zeta \rangle \right)^2 \leq (1 + o_p(1)) O_p(d_T) j^{\rho_\nu + 2 - 2\nu + 2}\delta \nu.
\]

(S2.48)

Hence, we conclude that, for \( j = 1, \ldots, K_2 \),

\[
\langle \hat{h}_j - h^*_j, \zeta \rangle ^2 \leq O_p(\alpha_1) j^{2 - 2\delta \nu} + O_p(d_T) j^{\rho_\nu + 2 - 2\nu + 2}\delta \nu \leq O_p(\alpha_1) j^{\rho_\nu + 2 - 2\nu + 2}\delta \nu,
\]

(S2.49)

where the last inequality follows from the fact that \( d_T \alpha_1^{-1} = o(1) \) and \( j^{-\rho_\nu} \leq 1 \).

Using the result given in (S2.49), we will show that

\[
\| (\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{xz} \hat{Q}_{K_2}^{-1} \zeta - (C_{zz})^{-1}_{K_1} C_{xz} Q_{K_2}^{-1} \zeta \| = o_p(1),
\]

(S2.50)

\[
\| \hat{Q}_{K_2}^{-1} \zeta - Q_{K_2}^{-1} \zeta \| = o_p(1).
\]

(S2.51)

To show (S2.50), note that

\[
\| (\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{xz} \hat{Q}_{K_2}^{-1} \zeta - (C_{zz})^{-1}_{K_1} C_{xz} Q_{K_2}^{-1} \zeta \| \leq \| (\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{xz} (\hat{Q}_{K_2}^{-1} - Q_{K_2}^{-1}) \zeta \| + \| (C_{zz})^{-1}_{K_1} C_{xz} - (C_{zz})^{-1}_{K_1} C_{xz} \| \| Q_{K_2}^{-1} \zeta \|.
\]

(S2.52)
Because \((\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{zz} = \hat{\Pi}_{K_1} B^* + (\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{vz}\) and \(\| (\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{vz} \|_{op} = O_p(\alpha_1^{-1/2} T^{-1/2})\), the first term in (S2.52) satisfies that
\[
\| (\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{zz} (\hat{Q}^{-1}_{K_2} - Q^{-1}_{K_2}) \| \leq O_p(1) \| (\hat{Q}^{-1}_{K_2} - Q^{-1}_{K_2}) \|.
\tag{S2.53}
\]
Moreover, under the employed assumptions, the following holds:
\[
\| Q^{-1}_{K_2} \|^2 = \| \sum_{j=1}^{K_2} \nu_j^{-1} \langle h_j, \zeta \rangle h_j \|^2 \leq O(1) \sum_{j=1}^{K_2} j^{\rho_{\nu} - 2\delta} \leq O(\max\{\alpha_2^{-1/\rho_{\nu}}, \alpha_2^{-(\rho_{\nu} - 2\delta + 1)/\rho_{\nu}}\}).
\tag{S2.54}
\]
Given that \(\alpha_1^{1/2} \alpha_2^{-1} = o(1)\), we have \(\alpha_1 \alpha_2^{-1/\rho_{\nu}} \leq \alpha_1 \alpha_2^{-1} \alpha_2^{-(\rho_{\nu} - 2\delta + 1)/\rho_{\nu}} = o(1)\) and \(\alpha_1 \alpha_2^{-(\rho_{\nu} - 2\delta + 1)/\rho_{\nu}} = o(1)\). This implies that \(\| Q^{-1}_{K_2} \|^2 \leq o(\alpha_1^{-1})\). Note also that
\[
((\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{zz} - (\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{zz} Q^{-1}_{K_2}) \zeta = (\hat{\Pi}_{K_1} - \Pi_{K_1}) B^* Q^{-1}_{K_2} \zeta + (\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{vz} Q^{-1}_{K_2} \zeta,
\tag{S2.55}
\]
where the second term on the right hand side is \(o(1)\) since \(\| (\hat{C}_{zz})^{-1}_{K_1} \hat{C}_{vz} \|_{op} = O_p(\alpha_1^{1/2} T^{-1/2}), \| Q^{-1}_{K_2} \| \leq o(\alpha_1^{-1/2}), \) and \(\alpha_1^{-1} T^{-1/2} = o(1)\). Furthermore, since \(\| \hat{\Pi}_{K_1} - \Pi_{K_1} \|_{op} \leq O(1) \sum_{j=1}^{K_1} \| \hat{g}_j - g_j^* \| \leq O_p(T^{-1/2} K_1^2) = O_p(T^{-1/2} \alpha_1^{-2/\rho_{\nu}})\) and \(\alpha_1^{-1} = o(T^{\rho_{\nu}/(2\rho_{\nu} + 2)})\), we find that
\[
\| (\hat{\Pi}_{K_1} - \Pi_{K_1}) B^* Q^{-1}_{K_2} \zeta \| \leq \| \hat{\Pi}_{K_1} - \Pi_{K_1} \|_{op} \| B \|_{op} \| Q^{-1}_{K_2} \| \zeta = O_p(T^{-1/2} \alpha_1^{-2/\rho_{\nu}}) = O_p(1).
\tag{S2.56}
\]
From (S2.53)-(S2.56), it is deduced that (S2.51) implies (S2.50) and thus we only need to show (S2.51) for the desired results. From a similar decomposition to that given in (S1.4), it can be shown that (S2.51) holds if the following terms are all \(o_p(1)\): \(\| \sum_{j=1}^{K_2} \nu_j^{-1} (\hat{h}_j - \hat{h}_j^*) \| \), \(\| \sum_{j=1}^{K_2} (\nu_j^{-1} - \hat{\nu}_j^{-1}) \hat{h}_j \| \), \(\| \sum_{j=1}^{K_2} \nu_j^{-1} \zeta \| \), \(\| \sum_{j=1}^{K} \nu_j^{-1} (\hat{h}_j - \hat{h}_j^*) \| \), \(\| \sum_{j=1}^{K} j^{2\rho_{\nu}} \| \), \(\| \sum_{j=1}^{K} j^{2\rho_{\nu} - 2\delta} \| \), \(\| \sum_{j=1}^{K} j^{2\rho_{\nu} - 2\delta + 1} \| \), and \(\| \sum_{j=1}^{K} j^{2\rho_{\nu} - 2\delta + 3} \| \).

As in (S1.35) and (S1.36), we obtain the following:
\[
\| \sum_{j=1}^{K_2} (\nu_j^{-1} - \hat{\nu}_j^{-1}) \langle h_j, \zeta \rangle \hat{h}_j \| = \sum_{j=1}^{K_2} (\nu_j^{-1} - \hat{\nu}_j^{-1})^2 \langle h_j, \zeta \rangle \| \leq \sum_{j=1}^{K_2} \frac{(\nu_j^{-2} - \hat{\nu}_j^{-2})^2}{\nu_j^{-2} \hat{\nu}_j^{-2}} \zeta \| \leq O_p(d_T \max\{\alpha_2^{-(3\rho_{\nu} - 2\delta + 1)/\rho_{\nu}}, \alpha_2^{-(1+\rho_{\nu})/\rho_{\nu}}\}),
\tag{S2.57}
\]
\[
\| \sum_{j=1}^{K} (\nu_j^{-1} - \hat{\nu}_j^{-1}) \langle h_j - h_j^*, \zeta \rangle \hat{h}_j \| = O_p(d_T \alpha_2) \sum_{j=1}^{K} j^{2\rho_{\nu}} \langle h_j - h_j^*, \zeta \rangle \| \leq O_p(d_T \alpha_1 \alpha_2) \sum_{j=1}^{K} j^{2+3\rho_{\nu} - 2\delta} \| \leq O_p(d_T \alpha_1) \max\{\alpha_2^{-(4\rho_{\nu} - 2\delta + 3)/\rho_{\nu}}, \alpha_2^{-(1+\rho_{\nu})/\rho_{\nu}}\}. \tag{S2.58}
\]
Under the conditions on \(\alpha_1\) and \(\alpha_2\) given in (4.5), we have \(d_T = O(\alpha_1)\) and \(\alpha_1^{1/2} \alpha_2^{-1} = o(1)\). These imply that the right hand sides of (S2.57) and (S2.58) are \(o_p(1)\). We then use the arguments that
are used to show (S1.37) and find that
\[
\| \sum_{j=1}^{K_2} \nu_j^{-1} (h_j, \zeta) (\hat{h}_j - h_j^s) \| \leq O_p(\alpha_1^{1/2}) \sum_{j=1}^{K_2} j^{-\delta_\xi + \rho_\nu/2 + 1} \leq O_p(\alpha_1^{1/2} \max \{ \alpha_2^{-1/\rho_\nu}, \alpha_2^{(2\delta_\xi - 2\rho_\nu - 3)/\rho_\nu} \}).
\]
(S2.59)

Due to \( \alpha_1^{1/2} \alpha_2^{-1} = o(1) \) and (4.5), we can show that the above term is \( o_p(1) \). We then note that
\[
\| \sum_{j=1}^{K_2} \nu_j^{-1} (\hat{h}_j - h_j^s, \zeta) \| \leq O_p(\alpha_1) \sum_{j=1}^{K_2} j^{-2\delta_\xi + 2\rho_\nu + 2} \leq O_p(\alpha_1 \max \{ \alpha_2^{-1/\rho_\nu}, \alpha_2^{(2\delta_\xi - 2\rho_\nu - 3)/\rho_\nu} \}),
\]
(S2.60)
where the inequalities follow from (S2.49) and the fact that \( \sum_{j=1}^{K_2} j^{-2\delta_\xi} \leq K_2 \sum_{j=1}^{K_2} j^{-2\delta_\xi} \leq K_2^2 \) due to \( \delta_\xi > 1/2 \). Because of (4.5), the right hand side of (S2.60) is also \( o_p(1) \). Thus, by combining the results given in (S2.57)-(S2.60), we conclude that (S2.51) holds. Then the results given in Theorem 6.(i) and 6.(ii) immediately follow from (S2.50) and (S2.51), and hence the details are omitted.

2. Analysis on the regularization bias: We next focus on the regularization bias term, \( \| A(\Pi_{K_2} - \Pi_{K_2}) \| \). Note that \( \| A(\Pi_{K_2} - \Pi_{K_2}) \| \leq G_1 + G_2 + G_3 + \| A(\Pi_{K_2} - I) \| \), where
\[
G_1 = \| \sum_{j=1}^{K_2} (h_j - h_j^s, \zeta) A(\hat{h}_j - h_j^s) \|, \quad G_2 = \| \sum_{j=1}^{K_2} (h_j^s, \zeta) A(\hat{h}_j - h_j^s) \|, \quad G_3 = \| \sum_{j=1}^{K_2} (\hat{h}_j - h_j^s, \zeta) A h_j^s \|.
\]
Then, by using (S2.20) and (S2.49), we find that
\[
\| G_1 \| \leq \sum_{j=1}^{K_2} \| (h_j - h_j^s, \zeta) \| A(\hat{h}_j - h_j^s) \| \leq O_p(\alpha_1) \sum_{j=1}^{K_2} j^{\rho_\nu - \delta_\xi - \omega + 2} \leq O_p(\alpha_1^{1/2}) \sum_{j=1}^{K_2} j^{\rho_\nu/2 - \omega - \delta_\xi + 1},
\]
\[
\| G_2 \| \leq \sum_{j=1}^{K_2} \| (h_j^s, \zeta) \| A(\hat{h}_j - h_j^s) \| \leq O_p(\alpha_1^{1/2}) \sum_{j=1}^{K_2} j^{\rho_\nu/2 - \omega - \delta_\xi + 1},
\]
and
\[
\| G_3 \| \leq \sum_{j=1}^{K_2} \| (\hat{h}_j - h_j^s, \zeta) \| A h_j^s \| \leq O_p(\alpha_1^{1/2}) \sum_{j=1}^{K_2} j^{\rho_\nu/2 - \omega - \delta_\xi + 1}.
\]
Hence, \( \| G_1 \| \), \( \| G_2 \| \) and \( \| G_3 \| \) are bounded by the following.
\[
O_p(\alpha_1^{1/2}) \sum_{j=1}^{K_2} j^{\rho_\nu/2 - \omega - \delta_\xi + 1} \leq \begin{cases} O_p(\alpha_1^{1/2}) & \text{if } \rho_\nu/2 + 2 < \omega + \delta_\xi, \\ O_p(\alpha_1^{1/2}) \max \{ \log \alpha_2^{-1}, \alpha_2^{-(\rho_\nu/2 - \omega - \delta_\xi + 2)/\rho_\nu} \} & \text{if } \rho_\nu/2 + 2 \geq \omega + \delta_\xi. \end{cases}
\]
Lastly, we have
\[
\| A(\Pi_{K_2} - I) \| \leq \sum_{j=K_2+1}^{\infty} \| (h_j, \zeta) A h_j \| \leq O(\sum_{j=K_2+1}^{\infty} j^{-2\delta_\xi - 2\omega}) \leq O_p(\alpha_2^{(2\omega + 2\delta_\xi - 1)/\rho_\nu}).
\]
Hence, the desired result follows. □
S3  Appendix to Section 5 on “Significance testing in functional endogenous linear model"

Proof of Theorem 9

For notational convenience, let \( c_\psi = \langle C_{uu}\psi, \psi \rangle^{1/2} \). To show (i), first note that

\[
\sqrt{T}\hat{C}_{yz}\psi = \sqrt{T}\hat{C}_{xz}A^*\psi + \sqrt{T}\hat{C}_{uz}\psi. \tag{S3.1}
\]

Under \( H_0 \), the first term in (S3.1) is equal to zero, and thus \( \sqrt{T}\hat{C}_{yz}\psi = T^{-1/2}\sum_{t=1}^T c_\psi \psi_t \), where \( \psi_t = c_\psi^{-1}(u_t, \psi)z_t \). Then, note that \( \mathbb{E}[\psi_t] = 0, \mathbb{E}[\psi_t \otimes \psi_t] = C_{zz} \) and \( \{\langle \psi_t, \zeta \rangle\}_{t \geq 1} \) is a real-valued martingale difference sequence for any \( \zeta \in \mathcal{H} \). Thus, by applying nearly identical arguments that are used to show (S1.8) and (S1.9), it can be shown that, for any \( \zeta \in \mathcal{H} \) and \( m \geq 0, T^{-1/2}\sum_{t=1}^T \langle \psi_t, \zeta \rangle \overset{d}{\to} N(0, \langle C_{zz} \zeta, \zeta \rangle) \) and \( \limsup_{n \to \infty} \limsup_{T \to \infty} \mathbb{P}(\sum_{j=m+1}^\infty T^{-1/2}\sum_{t=1}^T \psi_t, g_j)^2 > m) = 0 \) since \( C_{zz} \) is Hilbert-Schmidt. Therefore, \( T^{-1/2}\sum_{t=1}^T \psi_t \overset{d}{\to} N(0, C_{zz}) \), and we note that \( N(0, C_{zz}) \overset{d}{\to} \sum_{j=1}^\infty \sqrt{\mu_j} \epsilon_j g_j \), where \( \epsilon_j \sim_{iid} N(0, 1) \) across \( j \). The rest of the proof follows from the consistency of \( \hat{c}_\psi \) (Corollaries 1 and 2), continuous mapping theorem, and orthonormality of \( \{g_j\}_{j \geq 1} \).

Under \( H_1 \), the first term in (S3.1) is not equal to zero, and, by combining the results given above, we find that \( \|\sqrt{T}\hat{C}_{yz}\psi\|^2 = T\|C_{zz}A^*\psi\|^2 + O_p(1) \) holds in this case. Since \( \ker C_{zz} = \{0\} \) under either of Assumptions 1 or 3, we find that \( \|C_{zz}A^*\psi\|^2 > 0 \) under \( H_1 \), and therefore \( J \overset{p}{\to} \infty \).

To show (ii), note that \( \sum_{j=1}^\infty \mu_j x_j^2 - \sum_{j=1}^D \mu_j |x_j|^2 \leq \sum_{j=1}^\infty \mu_j - \sum_{j=1}^D \mu_j |x_j|^2 \leq \sum_{j=1}^\infty \mu_j |x_j|^2 = O_p(1) \) by Markov’s inequality, and \( \sup_{j \geq 1} \mu_j - \mu_j \leq \|C_{zz} - C_{zz}\|_{op} \leq O_p(T^{-1/2}) \) under Assumption 1.(g) and Lemma 4.2 in Bosq (2000). Since \( C_{zz} \) is nonnegative, we also have \( \mathbb{E}[\sum_{j=1}^\infty \mu_j |x_j|^2] \leq \sum_{j=1}^\infty \mu_j |x_j|^2 \to 0 \) as \( D \to \infty \), which implies that \( \sum_{j=1}^\infty \mu_j |x_j|^2 = o_p(1) \). Since \( D \to \infty \) and \( D/\sqrt{T} \to 0 \) as \( T \to \infty \), we find that \( \sum_{j=1}^\infty \mu_j x_j^2 - \sum_{j=1}^D \mu_j x_j^2 = o_p(1) \), as desired. \( \Box \)

S4  Strong consistency results

S4.1  Strong consistency of the FIVE

We first review some essential mathematics to establish the strong consistency of our estimators. The space of Hilbert-Schmidt operators, denoted \( \mathcal{S}_H \), is a separable Hilbert space with respect to the inner product given by \( \langle T_1, T_2 \rangle_{\mathcal{S}_H} = \sum_{j,k \geq 1} \langle T_1 \zeta_{ij}, \zeta_{2k} \rangle \langle T_2 \zeta_{ij}, \zeta_{2k} \rangle \) for two arbitrary orthonormal bases \( \{\zeta_{ij}\}_{j \geq 1} \) and \( \{\zeta_{2j}\}_{j \geq 1} \) of \( \mathcal{H} \); this inner product does not depend on the choice of orthonormal bases (Bosq, 2000, Chapter 1). We then note that \( \{x_t \otimes z_t - C_{xz}\}_{t \geq 1} \) is zero-mean stationary and geometrically strongly mixing in \( \mathcal{S}_H \), and, in the sequel, employ the following additional assumption: below, \( \{A_j\}_{j \geq 1} \) is the sequence of eigenvalues of the covariance operator of \( d_t = x_t \otimes z_t - C_{xz} \).

Assumption S5.  (a) \( \sup_{t \geq 1} \|x_t\| \leq m_x, \sup_{t \geq 1} \|z_t\| \leq m_z \), and \( \sup_{t \geq 1} \|u_t\| \leq m_u \) a.s., (b) \( A_j \leq ab^j \) for some \( a > 0 \) and \( 0 < b < 1 \).

As shown by Corollaries 2.4 and 4.2 of Bosq (2000), Assumption S5 combined with Assumption 1.(b) helps us obtain the stochastic order of \( \|\hat{C}_{xz} - C_{xz}\|_{op} \), which is given as follows:
Lemma S3. Under Assumptions 1.(b) and S5, the following holds almost surely:

\[ \| \hat{C}_{zz} - C_{zz} \|_{op} = O(T^{-1/2} \log^{3/2} T). \]

We omit the proof of Lemma S3 since it is a direct consequence of Theorem 2.12 and Corollary 2.4 of Bosq (2000), and the fact that \( \sup_{t \geq 1} \| d_t \| \leq m_d \) holds for some \( m_d > 0 \) under the employed assumptions. Based upon this result, we can establish the strong consistency of the FIVE as follows:

Theorem 1 (continued). If Assumption S5 is additionally satisfied, \( \tau(\alpha) = o(T^{1/2} \log^{-3/2} T) \) a.s., and \( \alpha^{-1} T^{-1} \log T \to 0 \), then \( \| \hat{A} - A \|_{op} \to 0 \) a.s.

Proof. It can be easily shown from our proof of Theorem 1 that \( \| \hat{A} - \hat{A}_{\Pi} \|_{op} \leq \alpha^{-1/2} \| T^{-1} \sum_{t=1}^{T} z_t \otimes u_t \|_{op} \) holds a.s. Under Assumptions 1 and S5, the sequence of \( z_t \otimes u_t \) is a martingale difference, and \( \| z_t \otimes u_t \|_{HS} \) and \( \mathbb{E} \| z_t \otimes u_t \|_{HS}^2 \) are uniformly bounded, and we thus know from Theorem 2.14 of Bosq (2000) that \( \| T^{-1} \sum_{t=1}^{T} z_t \otimes u_t \|_{op} = O(T^{-1/2} \log^{1/2} T), \) a.s. This implies that \( \| \hat{A} - \hat{A}_{\Pi} \|_{op} = O(\alpha^{-1/2} T^{-1/2} \log^{1/2} T) \) a.s. Moreover, we note that \( \| A_{\Pi} - A \|_{op}^2 \) is bounded above by the term given in the right hand side of (S1.2), and deduce from Lemma S3 that \( \| \hat{C}_{xx} \hat{C}_{zz} - C_{xx} C_{zz} \|_{op} = O(T^{-1/2} \log^{3/2} T) \) a.s. These results imply that \( \| A_{\Pi} - A \|_{op}^2 = o(1) \) a.s.

S4.2 Strong consistency of the F2SLSE

As in the case of the FIVE, we need some additional assumptions: below, as we did for the sequence of \( d_t \) in Section S4.1, we let \( \{ M_j \}_{j \geq 1} \) be the sequence of eigenvalues of the covariance of \( z_t \otimes z_t - C_{zz} \).

Assumption S6. (a) \( \sup_{t \geq 1} \| x_t \| \leq m_x, \sup_{t \geq 1} \| z_t \| \leq m_z \), and \( \sup_{t \geq 1} \| u_t \| \leq m_u \) a.s., (b) \( M_j \leq ab^j \) for some \( a > 0 \) and \( 0 < b < 1 \), (c) the sequence of \( v_t \) is a martingale difference with respect to \( \mathcal{G}_t = \sigma(\{ z_s \}_{s \leq t+1}, \{ u_s \}_{s \leq t}) \).

We may establish the following preliminary results:

Lemma S4. Under (4.1), Assumptions 1.(b) and S6, the following hold almost surely:

\[ \| \hat{C}_{zz} - C_{zz} \|_{op} = O(T^{-1/2} \log^{3/2} T) \quad \text{and} \quad \| \hat{C}_{vz} \|_{op} = O(T^{-1/2} \log^{1/2} T). \]

Proof. The first result follows from Theorem 2.12 and Corollary 2.4 of Bosq (2000). We then note that under (4.1) and Assumption S6, \( \sup_{t \geq 1} \| v_t \| \leq \sup_{t \geq 1} \| x_t \| + \| B \|_{op} \sup_{t \geq 1} \| z_t \| < \infty \), and apply Theorem 2.14 of Bosq (2000) to find that \( \| \hat{C}_{vz} \|_{op} = O(T^{-1/2} \log^{1/2} T) \) a.s.

In our proof of the strong consistency of the F2SLSE, what we want to have by employing Assumption S6.(c) is the asymptotic order of \( \| \hat{C}_{vz} \|_{op} \) given in Lemma S4; in fact, our proof does not require any change once the following weaker condition holds:

\[ \| \hat{C}_{vz} \|_{op} = O(T^{-1/2} \log T), \quad a.s. \quad \text{(S4.1)} \]

Hence, in the sequel, (S4.1) may replace Assumption S6.(c). We now establish the strong consistency.
**Theorem 5** (continued). If Assumption S6 is additionally satisfied, $(\sum_{j=1}^{K_1} \mu_j \tau_{1,j}) (\sum_{j=1}^{K_2} \tau_{2,j}) = o(T^{1/2} \log^{-3/2} T)$ a.s., $(\sum_{j=K_1+1}^{\infty} \mu_j) (\sum_{j=1}^{K_2} \tau_{2,j}) = o(1)$ a.s., $\alpha_1 \alpha_2^{-1} \rightarrow 0$, and $\alpha_1^{-1} T^{-1} \log T \rightarrow 0$, then $\|\hat{A} - A\|_{op} \rightarrow 0$ a.s.

**Proof.** From (S2.1), we know that $\|\hat{A} - \tilde{A} \Pi K_2\|_{op} \leq \alpha_1^{-1/4} \alpha_2^{-1/4} \|T^{-1} \sum_{t=1}^{T} z_t \otimes u_t\|_{op}$. Moreover, $\{z_t \otimes u_t\}_{t\geq 1}$ is a martingale difference sequence satisfying that $\sup_{t} \mathbb{E}\|z_t \otimes u_t\|_{HS} < \infty$ a.s. and $\sup_{t} \mathbb{E}\|z_t \otimes u_t\|_{HS} < \infty$. We therefore deduce from Theorem 2.14 of Bosq (2000) that $\|\hat{A} - \tilde{A} \Pi K_2\|_{op} = O(\alpha_1^{-1/4} \alpha_2^{-1/4} T^{-1/2} \log^{1/2} T)$ a.s., and hence, $\|\hat{A} - \tilde{A} \Pi K_2\|_{op} = o(1)$ a.s. Note also that $\|A \Pi K_2 - A\|_{op} \leq \sum_{j=K_2+1}^{\infty} \|Ah_j\|^2 + |\mathcal{R}|$, where $h_j$ and $\mathcal{R}$ are defined as in our proof of Theorem 5. Since $A$ is a Hilbert-Schmidt operator, we have $\sum_{j=K_2+1}^{\infty} \|Ah_j\|^2 = o(1)$ a.s. It thus only remains to show that $|\mathcal{R}| = o(1)$ a.s. We know from Lemma S4 that $\|\hat{C}_{zz} - C_{zz}\|_{op} = O(T^{-1/2} \log^{3/2} T)$ a.s., and hence

$$O \left( \sum_{j=1}^{K_1} \mu_j \tau_{1,j} \right) \sum_{j=1}^{K_2} \tau_{2,j} \|\hat{C}_{zz} - C_{zz}\|_{op} = o(1), \quad \text{a.s.} \quad (S4.2)$$

As shown in Lemma S4, we have $\|\hat{C}_{zz}\|_{op} = O(T^{-1/2} \log^{1/2} T)$, a.s., and also find that

$$\|S\|_{op} \sum_{j=1}^{K_2} \tau_{2,j} \leq (O(T^{-1/2} \log^{1/2} T) + O(\alpha_1^{-1/2} T^{-1} \log T)) \sum_{j=1}^{K_2} \tau_{2,j}$$

$$= o(\log^{-1} T) + o(\alpha_1^{-1/2} T^{-1/2} \log^{-1/2} T) = o(1), \quad \text{a.s.} \quad (S4.3)$$

by the definition of $S$. Moreover,

$$\|T\|_{op} \sum_{j=1}^{K_2} \tau_{2,j} \leq O \left( \sum_{j=K_1+1}^{\infty} \mu_j \right) \sum_{j=1}^{K_2} \tau_{2,j} = o(1), \quad \text{a.s.} \quad (S4.4)$$

From (S2.9), (S4.2), (S4.3), and (S4.4), it immediately follows that $|\mathcal{R}| = o(1)$ a.s.

**References**

Baker, C. R. (1973). Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society* 186, 273–289.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.

Ibukiyama, T. and M. Kaneko (2014). The Euler–Maclaurin summation formula and the Riemann zeta function. In *Bernoulli Numbers and Zeta Functions*, pp. 65–74. Springer, Tokyo.

Imaizumi, M. and K. Kato (2018). PCA-based estimation for functional linear regression with functional responses. *Journal of Multivariate Analysis* 163, 15–36.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.