Econometric Theory, **0**, 2024, 1–36.
doi:10.1017/S0266466624000252

FUNCTIONAL INSTRUMENTAL VARIABLE REGRESSION WITH AN APPLICATION TO ESTIMATING THE IMPACT OF IMMIGRATION ON NATIVE WAGES

DAKYUNG SEONG
University of Sydney

WON-KI SEO
University of Sydney

Functional linear regression has gained popularity as a statistical tool for studying the relationship between function-valued variables. However, in practice, it is hard to expect that the explanatory variables of interest are strictly exogenous, due to, for example, the presence of omitted variables and measurement error. This issue of endogeneity remains insufficiently explored, in spite of its empirical importance. To fill this gap, this article proposes new consistent FPCA-based instrumental variable estimators and develops their asymptotic properties in detail. Simulation experiments under a wide range of settings show that the proposed estimators perform considerably well. We apply our methodology to estimate the impact of immigration on native labor market outcomes in the US.

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, including economics, have 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 topic include Bosq (2000), Yao, Müller, and Wang (2005), Mas (2007), Hall and Horowitz (2007), Park and Qian (2012), Crambes and Mas (2013), Benatia,

The authors express deep appreciation to the editor, the co-editor, and the three anonymous referees for their invaluable and insightful suggestions. We are also thankful to Morten Ø. Nielsen and seminar participants at the University of Sydney, University of Queensland, and SETA 2022 for their helpful comments. Data and R code to replicate the empirical results in Table 3 are available on the authors' websites. Address correspondence to Dakyung Seong, School of Economics, University of Sydney, Camperdown, NSW, Australia, e-mail: dakyung.seong@sydney.edu.au.

Carrasco, and Florens (2017), Imaizumi and Kato (2018), Sun, Du, Wang and Ma (2018), and Chen, Guo, and Qiao (2022), to name only a few.

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 is particularly relevant in the context of functional linear models because functional observations used in the analysis are typically constructed by smoothing their discrete, and often sparsely observed, realizations (see, e.g., Yao et al., 2005). If this being the case, the functional observations may inevitably involve small or large measurement errors, which leads to the violation of the exogeneity condition at least to some degree (see, e.g., Sects. 2 and 5.3). This issue may hinder practitioners from applying the functional linear model.

If the volume of the literature on functional linear regression under the exogeneity condition provides any indication, the developments made so far to deal with endogeneity do not seem to sufficiently meet practitioners' needs. 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 the 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 their asymptotic properties. This is a crucial point where the present article is differentiated from the existing ones concerning the issue of endogeneity in the functional linear model.

Specifically, this article 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 (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, 2022). 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 most of the papers mentioned above, we do not require the variables of interest to be independent and identically distributed (i.i.d.), but allow those to exhibit some weak dependence so that our methodology can be applied to various empirical examples. Given that many functional observations considered in the literature for applications in the fields of energy, environmental and financial economics tend to involve time dependence, this extension may be attractive to practitioners.

Among the aforementioned papers, the study by Chen et al. (2022) is most closely related to the present article 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). 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 considered, and are expected to have their own merits (see Remark 3). It is worth summarizing some crucial differences between our estimators and the alternative estimator proposed by Benatia et al. (2017) based on ridge regularization. To the best of our knowledge, there has not been exploration of asymptotic inference on specific characteristics of the regression (coefficient) operator for their estimator; this seems to be because of a nontrivial challenge associated with a particular asymptotic bias (see Benatia et al., 2017, Sect. 4). In contrast, in this article, we tackle the issue by employing the FPCA augmented with a proper extension of the asymptotic approach introduced by Hall and Horowitz (2007). As a result, this article provides mathematical conditions that support a valid asymptotic inference on the regression operator. Moreover, Benatia et al.'s (2017) methodology does not take into account for the more general presence of weak dependence, although we believe their results can be extended to such a setting.

This article studies in depth the asymptotic properties of the proposed estimators. It is first shown that, under some mild conditions on the data generating process (DGP) 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 estimator, 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,

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

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 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 article 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 estimator is defined. This result implies that the proposed estimator does not weakly converge to any elements in the usual operator topology, which generalizes 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.

To see how the asymptotic properties of our estimators 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.

As an empirical illustration, we study the impact of immigration on wages of native workers in the US. Specifically, we employ a model that is similar to those considered by Dustmann, Frattini, and Preston (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 skill levels (captured by, e.g., years of education and experience) of both natives and immigrants. We, in this article, 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 workers' skill (will be detailed in Sect. 5.4). This approach has a couple of advantages compared to that taken in the earlier literature. For example, in the previous literature, workers of various skill levels are often classified into a few skill groups before analysis, which is necessitated to reduce the dimensionality of the considered model (see Example 1 and Sect. 5.4). However, such a preclassification, which may affect estimation results and their interpretation, is not required in our approach. Moreover, our methodology allows for studying if an inflow of immigrants in a particular skill group heterogeneously affects workers equipped with different skill levels. Using the methodology developed in this article, we find evidence supporting the presence of heterogeneous effects of immigration.

This article is organized as follows: Section 2 introduces a functional endogenous linear model and provides motivating examples. In Section 3, we define the FIVE and discuss its asymptotic properties. Section 4 introduces the F2SLSE and discusses its asymptotic properties. Section 5 reports simulation results and details our empirical example. Section 6 concludes. The mathematical proofs of the theoretical results can be found in the Supplementary Material.

2. FUNCTIONAL ENDOGENOUS LINEAR MODEL

2.1. Endogeneity and Motivating Examples

We suppose that a stationary sequence of random functions $\{y_t, x_t, u_t\}_{t\geq 1}$ satisfies the following:

$$y_t = c_y + \mathcal{A}x_t + u_t, \tag{2.1}$$

where c_y is the intercept function and $\mathcal{A}$ is a linear operator satisfying certain conditions to be clarified. In (2.1), y_t, x_t and u_t will be technically understood as random variables taking values in separable Hilbert spaces. Section S1.1 of the Supplementary Material briefly introduces the definitions of a Hilbert-valued random variable X, its expectation (denoted $\mathbb{E}[X]$), covariance operator (denoted $\mathcal{C}_{XX} := \mathbb{E}[(X-\mathbb{E}[X])\otimes(X-\mathbb{E}[X])]$), 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 by $X\otimes Y(\cdot) = \langle X,\cdot\rangle Y$ for any random or nonrandom X and Y taking values in $\mathcal{H}$ (see (S1.1) in the Supplementary Material).

We say that the explanatory variable x_t is endogenous if 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 article focuses on estimation and inference of the functional linear model in the presence of endogeneity. Below we provide specific empirical examples that motivate this model of interest.

Example 1 (Effects of immigration on the native labor market). In Section 5.4, we will explore a functional version of a well-known linear regression model that examines the skill-dependent effects of immigrant inflows on native workers' wages. In this example, the dependent variable y_t and the explanatory variable x_t are functions representing skill-specific changes in wage and the share of immigrants, respectively, at time t. As will be detailed in Section 5.4, this model suitably extends existing approaches that typically require pre-classifying workers into only a few groups (e.g., low, mid, and high skilled groups) and/or may not allow for spillover effects across different skill groups. As is known in the literature (see, e.g., Llull, 2018), estimating this model involves challenging issues, including the endogenous occupational adjustment of workers.

Example 2 (Functional AR model with measurement errors). The functional AR model has been used in many applications involving functional data. We,

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

in this example, consider the functional AR model where each observation is contaminated by a measurement error; this may be understood as a special case of the model considered in Chen et al. (2022). An example can be found in the recent literature on forecasting of probability densities; see, for example, Kokoszka, Miao, Peterson and Shang (2019). Since true probability density functions are not observable in practice, they need to be replaced by appropriate nonparametric estimates that inherently involve estimation errors. Beyond this specific case, it seems to be quite common in practice that the true functional realization y_t° 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, that is, $y_t = y_t^\circ + e_t$. If $\{y_t^\circ\}_{t\geq 1}$ satisfies the stationary AR law of motion given by $y_t^\circ = \mathcal{A}y_{t-1}^\circ + \epsilon_t$ for $t \geq 1$, with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[y_{t-1}^\circ \otimes \varepsilon_t] = 0$, we have

$$y_t = \mathcal{A}y_{t-1} + u_t, \quad \text{where} \quad u_t = e_t - \mathcal{A}e_{t-1} + \epsilon_t.$$

In this case, $\mathbb{E}[y_{t-1} \otimes u_t] \neq 0$ in general, and hence y_{t-1} is endogenous.

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 to be violated. In particular, due to advancements in data collection techniques, it is now possible to construct density- or curve-valued economic variables from large datasets. As a result, the analysis of these variables has gained popularity, as evidenced by various empirical examples in the literature (e.g., Benatia et al., 2017; Babii, 2022; Chen et al., 2022; Nielsen, Seo, and Seong, 2023; Seo, 2024). In economic functional data, the observations are often incomplete, and the discrete and finite realizations used to construct functional observations may not be enough to fully capture the entire function. Therefore, practitioners may remain cautious about potential endogeneity, even if the considered regressor x_t is presumed to be exogenous. This limitation could hinder the practical use of the functional linear model. As 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 will be pursued in our Hilbert space setting.

2.2. Model, Assumptions, and Notation

To facilitate the subsequent discussions, it may be helpful to introduce some additional notation. We first let $\mathcal{H}$ denote 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 $\mathcal{H}$. $\mathcal{L}_\mathcal{H}$ denotes the space of bounded linear operators acting on $\mathcal{H}$, equipped with the operator norm $\|\mathcal{T}\|_{\text{op}} = \sup_{\|\zeta\|\leq 1}\|\mathcal{T}\zeta\|$. For any $\mathcal{T} \in \mathcal{L}_\mathcal{H}$, we let $\mathcal{T}^*$, $\operatorname{ran}\mathcal{T}$, $\ker\mathcal{T}$, and $\|\mathcal{T}\|_{\text{HS}}$ denote the adjoint, range, kernel, and Hilbert–Schmidt norm of $\mathcal{T}$, which are briefly reviewed in Section S1 of the Supplementary Material; in that section, various properties of $\mathcal{T} \in \mathcal{L}_{\mathcal{H}}$, such as nonnegativity, positivity, self-adjointness, compactness and Hilbert-Schmidtness, are also reviewed. For any nonnegative, self-adjoint and compact $\mathcal{T}$, we may write $\mathcal{T} = \sum_{j=1}^{\infty} a_j \zeta_j \otimes \zeta_j$ for some nonnegative sequence $\{a_j\}_{j\geq 1}$ and some orthonormal basis $\{\zeta_j\}_{j\geq 1}$. Then $\mathcal{T}^{1/2}$ can be well defined by replacing a_j with $\sqrt{a_j}$.

This article concerns the case where the response variable y_t and the endogenous explanatory variable x_t are infinite-dimensional random variables taking values in separable Hilbert spaces. We hereafter conveniently assume that all of such variables take values in $\mathcal{H}$. This setup in fact encompasses an apparently more general scenario where y_t and x_t take values in different separable Hilbert spaces of infinite dimension, say $\mathcal{H}_y$ and $\mathcal{H}_x$. This is because these spaces are all isomorphic to $\mathcal{H}$ (see, e.g., Conway, 2007, Cor. 5.5, p. 21), and thus there is no loss of generality by assuming $\mathcal{H}_y = \mathcal{H}_x = \mathcal{H}$. We further assume for convenience that y_t and x_t have zero means, that is, $\mathbb{E}[y_t] = \mathbb{E}[x_t] = 0$; 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 $\mathcal{A} : \mathcal{H} \mapsto \mathcal{H}$,

$$y_t = \mathcal{A}x_t + u_t, \quad \text{where} \quad \mathbb{E}[x_t \otimes u_t] \neq 0 \quad \text{and} \quad \mathbb{E}[u_t] = 0. \tag{2.2}$$

We then let z_t (to be called the IV) be another zero-mean $\mathcal{H}$-valued random variable satisfying $\mathbb{E}[z_t \otimes u_t] = 0$. For notational convenience, we use $\mathcal{C}_{zz}$, $\mathcal{C}_{xz}$, $\mathcal{C}_{yz}$ and $\mathcal{C}_{uz}$ to denote the following operators:

$$\mathcal{C}_{zz} = \mathbb{E}[z_t \otimes z_t], \quad \mathcal{C}_{xz} = \mathbb{E}[x_t \otimes z_t], \quad \mathcal{C}_{yz} = \mathbb{E}[y_t \otimes z_t], \quad \text{and} \quad \mathcal{C}_{uz} = \mathbb{E}[u_t \otimes z_t].$$

Similarly, let $\widehat{\mathcal{C}}_{zz}$, $\widehat{\mathcal{C}}_{xz}$, $\widehat{\mathcal{C}}_{yz}$ and $\widehat{\mathcal{C}}_{uz}$ denote their sample counterparts that are computed as follows:

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

We will employ the following assumptions throughout the article: below, $\mathfrak{F}_t$ denotes the filtration given by $\mathfrak{F}_t = \sigma\left(\{z_s\}_{s\leq t+1}, \{u_s\}_{s\leq t}\right)$, and $\widehat{\mathcal{C}}_{uu} = T^{-1}\sum_{t=1}^{T} u_t \otimes u_t$.

Assumption M. (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) $\mathbb{E}[u_t \otimes u_t|\mathfrak{F}_{t-1}] = \mathcal{C}_{uu}$, and $\sup_{1\leq t\leq T} \mathbb{E}[\|u_t\|^{2+\delta}|\mathfrak{F}_{t-1}] < \infty$ for $\delta > 0$, (e) $\mathcal{A}$ is Hilbert–Schmidt, (f) $\|\widehat{\mathcal{C}}_{xz} - \mathcal{C}_{xz}\|_{\text{HS}}$, $\|\widehat{\mathcal{C}}_{zz} - \mathcal{C}_{zz}\|_{\text{HS}}$ and $\|\widehat{\mathcal{C}}_{uz} - \mathcal{C}_{uz}\|_{\text{HS}}$ are $O_p(T^{-1/2})$, (g) $\|\widehat{\mathcal{C}}_{uu} - \mathcal{C}_{uu}\|_{\text{HS}} = o_p(1)$, (h) $\ker \mathcal{C}_{xz} = \{0\}$.

By Assumption M(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, Chaps. 13–16), by not restricting our attention to the i.i.d. case and (ii) the variables to be considered in our

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

empirical application (Sect. 5.4) naturally exhibit time series dependence. In Assumptions M(c) and M(d), the error term u_t is assumed to be a homoscedastic martingale difference sequence. Assumption M(c) states the conditions required for the IV z_t in this setting (see Example 3 for a possible IV presented for the model in Example 2), and this condition implies that u_t is uncorrelated with z_t. In Assumption M(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 i.i.d. sequence, as often assumed in the literature, Assumptions M(c) and M(d) reduce to the following:

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

The Hilbert–Schmidt condition of $\mathcal{A}$ given in Assumption M(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 Assumptions M(f) and M(g), high-level conditions on limiting behaviors of some sample operators are given, and these are for mathematical convenience. We first note that $\{x_t \otimes z_t - \mathcal{C}_{xz}\}_{t\geq 1}$, $\{z_t \otimes z_t - \mathcal{C}_{zz}\}_{t\geq 1}$, and $\{u_t \otimes z_t - \mathcal{C}_{uz}\}_{t\geq 1}$ are sequences in the Hilbert space of Hilbert–Schmidt operators, denoted by $\mathcal{S}_{\mathcal{H}}$ (see Section S2.3.1 of the Supplementary Material). If those sequences are i.i.d. (resp. geometrically strongly mixing), then Assumption M(f) holds once $\mathbb{E}[(\|x_t\|\|z_t\|)^{\upsilon}]$, $\mathbb{E}[\|z_t\|^{2\upsilon}]$, and $\mathbb{E}[(\|u_t\|\|z_t\|)^{\upsilon}]$ are finite for some $\upsilon \geq 2$ (resp. $\upsilon \geq 2+\delta$ for some $\delta > 0$) (Bosq, 2000, Thms. 2.7 and 2.17); such primitive sufficient conditions can also be found for martingale differences (Bosq, 2000, Thm. 2.16) and weakly stationary sequences (Bosq, 2000, Thm. 2.18). We also observe that $\{u_t \otimes u_t - \mathcal{C}_{uu}\}_{t\geq 1}$ is a martingale difference sequence in $\mathcal{S}_{\mathcal{H}}$, and some primitive sufficient conditions for Assumption M(g) can be found in, for example, Theorems 2.11 and 2.14 of Bosq (2000). Lastly, Assumption M(h) enables us to identify the unique bounded linear operator $\mathcal{A}$ satisfying (2.2) using the IV z_t, which will be discussed in more detail in Section 3.

Example 3. Consider Example 2 in Section 2.1. Suppose that the sequence of measurement errors $\{e_t\}_{t\geq 1}$ satisfies that $\mathbb{E}[e_t|\mathcal{G}_{t-1}] = 0$ where $\mathcal{G}_{t-1} = \sigma\left(\{y_s\}_{s\leq t-1}, \{e_s\}_{s\leq t-1}, \{\varepsilon_s\}_{s\leq t}\right)$. Then $y_{t-\ell}$ for $\ell > 1$ satisfies the exogeneity condition implied by Assumption M(c). Note that y_{t-2} satisfies

$$\mathbb{E}[y_{t-2} \otimes y_t] = \mathcal{A}\mathbb{E}[y_{t-2} \otimes y_{t-1}] + \mathbb{E}[y_{t-2} \otimes u_t] = \mathcal{A}\mathbb{E}[y_{t-2} \otimes y_{t-1}]. \tag{2.3}$$

This 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 the univariate AR(1) case, the modified Yule–Walker estimator can be obtained by replacing the population moments in (2.3) by their sample counterparts; see Walker (1960) and Staudenmayer and Buonaccorsi (2005, Sect. 4.4). As may be expected from this example, 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 $\mathcal{A}$. In this regard, our model

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

is related to that of Chen et al. (2022), but the theory and methodology that are subsequently pursued in this article move in an apparently different direction.

Remark 1. Consider the standard Euclidean space setting where $\mathcal{C}_{xz}$ is invertible. From (2.2) and Assumption M(c), the standard IV estimator, $\widehat{\mathcal{C}}_{xz}^{-1}\widehat{\mathcal{C}}_{yz}$ can be defined and it can be understood as a sample analogue of $\mathcal{A}^*$. Alternatively, (2.2) and Assumption M(c) imply that $\mathcal{C}_{xz}^*\mathcal{C}_{yz} = \mathcal{C}_{xz}^*\mathcal{C}_{xz}\mathcal{A}^*$. Thus, based on this equation, we may define another IV estimator, $(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})^{-1}\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{yz}$, which can be understood as a GMM estimator with an identity weight. Our FIVE is defined as in the latter case. This choice offers a theoretical advantage especially in our functional setup because we can utilize well-established mathematical results concerning the spectral properties of self-adjoint operators ($(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})^{-1}$) to study the asymptotic properties of the FIVE.

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

We find from (2.2) that $\mathcal{C}_{yz}^* = \mathbb{E}[z_t \otimes y_t] = \mathcal{A}\mathbb{E}[z_t \otimes x_t] = \mathcal{A}\mathcal{C}_{xz}^*$ and hence,

$$\mathcal{C}_{yz}^*\mathcal{C}_{xz} = \mathcal{A}\mathcal{C}_{xz}^*\mathcal{C}_{xz}. \tag{3.1}$$

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

$$\mathcal{C}_{xz}^*\mathcal{C}_{xz} = \sum_{j=1}^{\infty} \lambda_j^2 f_j \otimes f_j,$$

where f_j is the eigenfunction corresponding to λ_j^2. Given (3.1), it may be natural to consider an estimator $\bar{\mathcal{A}}$ that satisfies the equation $\widehat{\mathcal{C}}_{yz}^*\widehat{\mathcal{C}}_{xz} = \bar{\mathcal{A}}\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}$, obtained by replacing $\mathcal{C}_{yz}$ and $\mathcal{C}_{xz}$ with their sample counterparts. However, it is generally impossible to directly compute the estimator $\bar{\mathcal{A}}$ from this equation since $\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}$ is not invertible over the entire Hilbert space $\mathcal{H}$. We circumvent this issue by employing a regularized inverse of $\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}$ which may be understood as the well-defined inverse on a strict subspace of $\mathcal{H}$.

To this end, we first note that $\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}$ is nonnegative, self-adjoint, and compact, and hence allows the following representation:

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

$$\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz} = \sum_{j=1}^{\infty}\widehat{\lambda}_j^2\widehat{f}_j \otimes \widehat{f}_j,$$

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

$$\mathrm{K} = \#\{j : \widehat{\lambda}_j^2 > 1/\alpha\}. \tag{3.2}$$

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

$$\widehat{\mathcal{A}} = \widehat{\mathcal{C}}_{yz}^*\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}, \quad \text{where} \quad (\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1} = \sum_{j=1}^{\mathrm{K}}\widehat{\lambda}_j^{-2}\widehat{f}_j \otimes \widehat{f}_j. \tag{3.3}$$

The largest eigenvalue of the regularized inverse $(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}$ is bounded above by α 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 2), so our estimator may be understood as an extension of their estimator. We also note that the FIVE $\widehat{\mathcal{A}}$ may be viewed as a sample-analogue of $\mathcal{A}$ satisfying (3.1) in the sense that $\widehat{\mathcal{A}}$ is the solution to $\widehat{\mathcal{C}}_{yz}^*\widehat{\mathcal{C}}_{xz} = \widehat{\mathcal{A}}\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}$ on the restricted domain given by $\widehat{\mathcal{H}}_{\mathrm{K}} = \mathrm{span}\{\widehat{f}_j\}_{j=1}^{\mathrm{K}}$. Section S5 of the Supplementary Material discusses how $\widehat{\mathcal{A}}$ can be computed from the data using the FPCA.

Remark 2. It should be noted that K in $(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}$ is by construction a random variable associated with the choice of α. In the literature where a similar regularized inverse is discussed, K is chosen by practitioners and hence regarded as deterministic (e.g., Mas, 2007; Park and Qian, 2012). However, even in this case, it is generally recommended to choose K taking the eigenvalues of $\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}$ into account, and thus treating K as a random variable appears to be natural. This alternative perspective on K helps practitioners more directly control the degree of instability of the regularized inverse, measured by its largest eigenvalue, by means of the parameter α that they choose. Moreover, this approach distinguishes our asymptotic approach from those in the aforementioned papers.

3.2. General Asymptotic Properties

As may be deduced from our construction of the FIVE, $\widehat{\mathcal{A}} = 0$ is imposed outside a subspace whose dimension increases as α gets larger. Thus, for $\widehat{\mathcal{A}}$ to be a consistent estimator of $\mathcal{A}$ defined on the entire space $\mathcal{H}$, the regularization parameter α given in (3.2) needs to diverge to infinity. Taking this into consideration, we investigate the asymptotic properties of the FIVE when $T \to \infty$ and $\alpha \to \infty$ jointly. We will employ the following assumption:

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

Assumption E1. $\lambda_1^2 > \lambda_2^2 > \cdots > 0$.

That is, the eigenvalues of $\mathcal{C}_{xz}^*\mathcal{C}_{xz}$ are required to be distinct. This is employed to see asymptotic properties of the FIVE in detail and does not seem to be restrictive in practice; in fact, similar assumptions have been employed in the literature on functional linear models, see, for example, Bosq (2000, Sect. 8.3), Mas (2007), Hall and Horowitz (2007), and Park and Qian (2012), to name only a few.

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

$$\widehat{\mathcal{A}} - \mathcal{A} = (\widehat{\mathcal{A}} - \mathcal{A}\widehat{\Pi}_{\mathrm{K}}) - \mathcal{A}(\mathcal{I} - \widehat{\Pi}_{\mathrm{K}}), \tag{3.4}$$

where $\widehat{\Pi}_{\mathrm{K}}$ denotes the orthogonal projection defined by $\widehat{\Pi}_{\mathrm{K}} = \sum_{j=1}^{\mathrm{K}} \widehat{f}_j \otimes \widehat{f}_j$ and $\mathcal{I}$ is the identity operator acting on $\mathcal{H}$. Given that the FIVE is computed on the restricted domain ran $\widehat{\Pi}_{\mathrm{K}}$ (note that $\widehat{\mathcal{A}} = 0$ on $\mathrm{ran}(\mathcal{I} - \widehat{\Pi}_{\mathrm{K}})$ by construction) the first term of (3.4) may be understood as the deviation of $\widehat{\mathcal{A}}$ from $\mathcal{A}$ on ran $\widehat{\Pi}_{\mathrm{K}}$. Thus, this term is hereafter called the deviation component on the restricted domain (the DR component). On the other hand, the second term $\mathcal{A}(\mathcal{I} - \widehat{\Pi}_{\mathrm{K}})$ may be understood as the bias induced by the fact that $\widehat{\mathcal{A}}$ is enforced to zero on $\mathrm{ran}(\mathcal{I} - \widehat{\Pi}_{\mathrm{K}})$. We thus call this term the regularization bias component (the RB component). Our first result below shows that both the DR and RB components are asymptotically negligible and thus $\widehat{\mathcal{A}}$ becomes weakly consistent once the regularization parameter α diverges to infinity at an appropriate rate; in the next theorem, we let $\tau(\alpha)$ be a random function that increases without bound as $\alpha \to \infty$, which is defined by $\tau(\alpha) = \sum_{j=1}^{\mathrm{K}} \tau_j$, where $\tau_j = 2\sqrt{2}\max\{(\lambda_{j-1}^2 - \lambda_j^2)^{-1}, (\lambda_j^2 - \lambda_{j+1}^2)^{-1}\}$.

THEOREM 1. *Suppose that Assumptions M and E1 are satisfied,* $T^{-1/2}\tau(\alpha) \xrightarrow{p} 0$ *and* $T^{-1}\alpha \to 0$ *as* $\alpha \to \infty$ *and* $T \to \infty$. *Then*

$$\|\widehat{\mathcal{A}} - \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\|_{\mathrm{op}}^2 = O_p(T^{-1}\alpha) \quad \textit{and} \quad \|\mathcal{A}(\mathcal{I} - \widehat{\Pi}_{\mathrm{K}})\|_{\mathrm{op}}^2 = o_p(1).$$

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

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

The condition imposed on $\tau(\alpha)$ in Theorem 1 does not place any essential restrictions on the eigenvalues of $\mathcal{C}_{xz}^*\mathcal{C}_{xz}$. Given that $\tau(\alpha)$ increases as α (and thus K) gets larger, the condition, together with the requirement that $T^{-1}\alpha \to 0$, merely tells us that α needs to grow at 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 S2.3.1 of 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 $\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{xz}$, while we use a rank-regularized inverse of $\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{xz}$. 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 $\mathcal{C}_{xz}^{*}\mathcal{C}_{xz}$ to be distinct, which is not required in Benatia et al. (2017). In addition, Benatia et al.'s (2017) approach restricts the range of $\mathcal{A}$ to a certain subspace of $\mathcal{H}$, called the β-regularity space, while we need to restrict the increasing rate of α or K depending on $\tau(\alpha)$.

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

Theorem 2. *Suppose that Assumptions M and E1 are satisfied,* $\alpha^{1/2}T^{-1/2}\tau(\alpha) \xrightarrow{p} 0$ *and* $T^{-1}\alpha \to 0$ *as* $\alpha \to \infty$ *and* $T \to \infty$. *Then the following hold for any* $\zeta \in \mathcal{H}$.

(i) $\sqrt{T/\theta_{\mathrm{K}}(\zeta)}(\widehat{\mathcal{A}}-\mathcal{A}\widehat{\Pi}_{\mathrm{K}})\zeta \xrightarrow{d} \mathcal{N}(0,\mathcal{C}_{uu})$, *where* $\theta_{\mathrm{K}}(\zeta) = \langle \zeta, (\mathcal{C}_{xz}^{*}\mathcal{C}_{xz})_{\mathrm{K}}^{-1}\mathcal{C}_{xz}^{*}\mathcal{C}_{zz}\mathcal{C}_{xz}(\mathcal{C}_{xz}^{*}\mathcal{C}_{xz})_{\mathrm{K}}^{-1}\zeta\rangle$.

(ii) *If* $\widehat{\theta}_{\mathrm{K}}(\zeta) := \langle \zeta, (\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{zz}\widehat{\mathcal{C}}_{xz}(\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}\zeta\rangle$, *then* $|\widehat{\theta}_{\mathrm{K}}(\zeta)-\theta_{\mathrm{K}}(\zeta)| \xrightarrow{p} 0$.

Depending on the choice of ζ, $\theta_{\mathrm{K}}(\zeta)$ may be convergent or divergent in probability (see Remark 4), and thus the convergence rate of the DR component $(\widehat{\mathcal{A}}-\mathcal{A}\widehat{\Pi}_{\mathrm{K}})\zeta$ depends on ζ. 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 $\theta_{\mathrm{K}}(\zeta)$ is convergent in probability, then $(\widehat{\mathcal{A}}-\mathcal{A}\widehat{\Pi}_{\mathrm{K}})\zeta$ converges at $\sqrt{T}$-rate, otherwise it converges at a slower rate given by $\sqrt{T/\theta_{\mathrm{K}}(\zeta)}$ which is random (because of the randomness of K). As noted by Mas (2007), this discrepancy in convergence rates implies that (i) there exists no sequence of normalizing constants c_T such that $c_T(\widehat{\mathcal{A}}-\mathcal{A}\widehat{\Pi}_{\mathrm{K}})\zeta$ weakly converges to a well-defined limiting distribution uniformly in $\zeta \in \mathcal{H}$, and therefore it is impossible that $\widehat{\mathcal{A}}-\mathcal{A}\widehat{\Pi}_{\mathrm{K}}$ 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 $\mathcal{A}\widehat{\Pi}_{\mathrm{K}}$ is replaced by $\mathcal{A}$ (see Thm. 3.1 of Mas, 2007). Moreover, it can be deduced from Theorem 2 that as K increases, the regularization parameter α induces a trade-off between the decay rates of the DR and RB components when $\theta_{\mathrm{K}}(\zeta)$ is not convergent. If K relative to

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

T increases by a larger choice of α, then the operator norm of the RB component tends to shrink to zero at a faster rate. On the other hand, this change results in a faster divergence of $\theta_{\mathrm{K}}(\zeta)$, and Theorem 2 shows that the DR component will decay at a slower rate in such a case.

Remark 4. To see if $\theta_{\mathrm{K}}(\zeta)$ can be convergent or divergent depending on the choice of ζ, it is useful to assume that $x_t = z_t + v_t$, where $\{v_t\}_{t\geq 1}$ satisfies that $\mathbb{E}[z_t \otimes v_t] = 0$. For $j \geq 1$, let $\{\mu_j, g_j\}_{j\geq 1}$ be the pairs of eigenvalues and eigenfunctions of $\mathcal{C}_{zz}$. Then it can be shown that $\operatorname{plim}_{T\to\infty}\theta_{\mathrm{K}}(\zeta)$ is simply given by $\sum_{j=1}^{\infty}\mu_j^{-1}\langle g_j,\zeta\rangle^2$, and this quantity is bounded only when ζ belongs to a certain strict subspace of $\mathcal{H}$; see Carrasco, Florens, and Renault (2007, Sect. 3.2).

In applications involving economic or statistical time series, practitioners are often interested in the marginal effect of some additive and hypothetical perturbation, say ζ, in x_t on y_t. In the considered linear model, this marginal effect is simply given by $\mathcal{A}\zeta$, which can be consistently estimated by $\widehat{\mathcal{A}}\zeta$. Let $\{\zeta_{\mathrm{K}}\}$ be a sequence of random elements given by $\zeta_{\mathrm{K}} = \widehat{\Pi}_{\mathrm{K}}\zeta$. Then ζ_{K} is given by the orthogonal projection of the new perturbation ζ 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, ζ_{K} is the best linear approximation of ζ based on the covariation of the explanatory and instrumental variables. Therefore, ζ_{K} may be interpreted as a nice approximation showing how a hypothetical perturbation ζ can be revealed given the dataset, and we thus call ζ_{K} a data-supporting approximation of ζ. The following is an immediate consequence of Theorem 2: under the assumptions employed in Theorem 2,

$$|\widehat{\theta}_{\mathrm{K}}(\zeta_{\mathrm{K}}) - \theta_{\mathrm{K}}(\zeta_{\mathrm{K}})| \xrightarrow{p} 0 \quad \text{and} \quad \sqrt{T/\theta_{\mathrm{K}}(\zeta_{\mathrm{K}})}(\widehat{\mathcal{A}} - \mathcal{A})\zeta_{\mathrm{K}} \xrightarrow{d} \mathcal{N}(0, \mathcal{C}_{uu}).$$

Based on this result, we may implement standard statistical inference on various characteristics of $\mathcal{A}\zeta_{\mathrm{K}}$, which may provide a practical and interpretable insight for practitioners. We illustrate this by constructing a confidence interval for the random variable given by $\langle \mathcal{A}\zeta_{\mathrm{K}}, \psi\rangle$ for some $\psi \in \mathcal{H}$. In fact, various characteristics of $\mathcal{A}\zeta_{\mathrm{K}}$ may be written in this form; for example, if $\psi(s) = 1\{s_1 \leq s \leq s_2\}$ then $\langle \mathcal{A}\zeta_{\mathrm{K}}, \psi\rangle = \int_{s_1}^{s_2}\mathcal{A}\zeta_{\mathrm{K}}(s)ds$ 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 on y_t. We then consider the interval whose endpoints are given as follows:

$$\langle \widehat{\mathcal{A}}\zeta_{\mathrm{K}}, \psi\rangle \pm \Phi^{-1}(1-\varpi/2)\sqrt{\widehat{\theta}_{\mathrm{K}}(\zeta_{\mathrm{K}})\langle \widehat{\mathcal{C}}_{\widehat{u}\widehat{u}}\psi, \psi\rangle/T}, \tag{3.5}$$

where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution and $\widehat{\mathcal{C}}_{\widehat{u}\widehat{u}} = T^{-1}\sum_{t=1}^{T}\widehat{u}_t \otimes \widehat{u}_t$. Based on Theorem 2 and Corollary 1, the intervals that are repeatedly constructed as in (3.5) are expected to include $\langle \mathcal{A}\zeta_{\mathrm{K}}, \psi\rangle$ with $100(1-\varpi)\%$ of probability for a large T. Of course, (3.5) may not be quite satisfactory to practitioners who want to consider a purely hypothetical perturbation ζ without a data-supporting approximation. However, (i) the discrepancy between ζ and ζ_{K}

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

caused by the noninvertibility of $\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{xz}$ inevitably arises in our functional setting and (ii) the magnitude of discrepancy is expected to be small since it is, anyhow, asymptotically negligible. Therefore, this small bias caused by the data-supporting approximation may be understood as a cost of implementing 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 even if $\mathcal{A}\widehat{\Pi}_{\mathrm{K}}$ is replaced by $\mathcal{A}$ (see Remark 11); this, of course, implies that (3.5) can be understood as a confidence interval for $\langle \mathcal{A}\zeta, \psi\rangle$ with no data-supporting approximation.

Remark 5. Note that $\|\widehat{\mathcal{C}}_{xz}\|_{\mathrm{HS}}^2 = \sum_{j=1}^{\infty}\widehat{\lambda}_j^2 = \sum_{j=1}^{\infty}\|\widehat{\mathcal{C}}_{xz}\zeta_j\|^2$ 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

$$\sum_{j=1}^{\mathrm{K}}\|\widehat{\mathcal{C}}_{xz}\zeta_j\|^2 \Big/ \sum_{j=1}^{\infty}\widehat{\lambda}_j^2 = \sum_{k=1}^{\mathrm{K}}\sum_{j=1}^{\infty}\widehat{\lambda}_j^2\langle\widehat{f}_j,\zeta_k\rangle^2 \Big/ \sum_{j=1}^{\infty}\widehat{\lambda}_j^2.$$

Provided that $\{\widehat{\lambda}_j^2, \widehat{f}_j\}_{j\geq 1}$ is the sequence of the eigenelements of $\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{xz}$, it is deduced from the results given in Horváth and Kokoszka (2012, Thm. 3.2 and Sect. 3.2) that the above quantity is bounded above by $\sum_{j=1}^{\mathrm{K}}\widehat{\lambda}_j^2 \Big/ \sum_{j=1}^{\infty}\widehat{\lambda}_j^2$, and this upper bound is attained if and only if $\zeta_j = \pm\widehat{f}_j$ for $j = 1,\ldots,\mathrm{K}$. This shows that, among all the subspaces of dimension K, ran $\widehat{\Pi}_{\mathrm{K}}$ is the unique subspace that explains the most proportion of the squared Hilbert-Schmidt norm of $\widehat{\mathcal{C}}_{xz}$.

Remark 6. In Section S7 of the Supplementary Material, we develop a significance test to examine whether various characteristics of y_t, expressed as $\langle y_t, \psi\rangle$ for some $\psi \in \mathcal{H}$, depend on x_t. A crucial input to this test is a consistent estimator of $\mathcal{A}$, and the FIVE (and also the F2SLSE to be developed in Section 4) can be used.

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 (Thm. 1) and its DR component is asymptotically normal in a pointwise sense (Thm. 2) if α diverges to infinity at a sufficiently slow rate. We in this section investigate the asymptotic behavior of the FIVE in more detail under a set of assumptions which is stronger than Assumption E1 but 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 note that $\mathcal{C}_{xz}\mathcal{C}_{xz}^{*}$ allows the spectral

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

decomposition $\mathcal{C}_{xz}\mathcal{C}_{xz}^* = \sum_{j=1}^{\infty} \lambda_j^2 \xi_j \otimes \xi_j$ for some orthonormal basis $\{\xi_j\}_{j\geq 1}$, and let $\upsilon_t(j,\ell) = \langle x_t, f_j\rangle\langle z_t, \xi_\ell\rangle - \mathbb{E}[\langle x_t, f_j\rangle\langle z_t, \xi_\ell\rangle]$ for $j,\ell \geq 1$.

Assumption E2. There exist constants $c_\circ > 0$, $\rho > 2$, $\varsigma > 1/2$, $\gamma > 1/2$ and $m > 1$ satisfying the following: (a) $\lambda_j^2 \leq c_\circ j^{-\rho}$, (b) $\lambda_j^2 - \lambda_{j+1}^2 \geq c_\circ^{-1} j^{-\rho-1}$, (c) $|\langle \mathcal{A} f_j, \xi_\ell\rangle| \leq c_\circ j^{-\varsigma}\ell^{-\gamma}$, (d) $\mathbb{E}[\upsilon_t(j,\ell)\upsilon_{t-s}(j,\ell)] \leq c_\circ s^{-m}\mathbb{E}[\upsilon_t^2(j,\ell)]$ for $s \geq 1$, and furthermore, $\mathbb{E}[\|\langle x_t, f_j\rangle z_t\|^2] \leq c_\circ \lambda_j^2$ and $\mathbb{E}[\|\langle z_t, \xi_j\rangle x_t\|^2] \leq c_\circ \lambda_j^2$.

Assumptions E2(a) and E2(b) restrict the eigenstructure of $\mathcal{C}_{xz}$ (or equivalently $\mathcal{C}_{xz}^*\mathcal{C}_{xz}$), which are adapted from similar conditions in Hall and Horowitz (2007) and Imaizumi and Kato (2018).[1] Assumption E2(c) is a very natural condition given that $\langle \mathcal{A} f_j, \xi_\ell\rangle$ must be square-summable with respect to both j and ℓ; in this assumption, it is worth mentioning that ς is the parameter determining the smoothness of $\mathcal{A}$ on $\operatorname{ran}\mathcal{C}_{xz}^*\mathcal{C}_{xz}$. As may be deduced from the definition of $\upsilon_t(j,\ell)$ and Assumption M(b), $\{\upsilon_t(j,\ell)\}_{t\geq 1}$ is a stationary sequence in $\mathbb{R}$ for each j and ℓ, and the former condition of Assumption E2(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 both $\mathbb{E}[\|\langle x_t, f_j\rangle z_t\|^2]$ and $\mathbb{E}[\|\langle z_t, \xi_j\rangle x_t\|^2]$ naturally decrease as j gets larger and its decay rate is restricted by Assumption E2(d). Specifically, we require that the second moments of $\|\langle x_t, f_j\rangle z_t\|$ and $\|\langle z_t, \xi_j\rangle x_t\|$ as functions of j have a constant multiple of λ_j^2 as their upper envelope; a similar condition for the i.i.d. case can be found in, for example, Hall and Horowitz (2007).

The following theorem refines the result given in Theorem 1 under Assumption E2.

THEOREM 3. *Suppose that Assumptions M and E2 are satisfied and* $\alpha = o(T^{\rho/(2\rho+2)})$. *Then,* $\|\widehat{\mathcal{A}} - \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\|_{\mathrm{op}}^2 = O_p(T^{-1}\alpha)$ *as in Theorem 1, and*

$$\|\mathcal{A}(\mathcal{I} - \widehat{\Pi}_{\mathrm{K}})\|_{\mathrm{op}}^2 = O_p(T^{-1}\alpha \max\{1, \alpha^{(3-2\varsigma)/\rho}\} + \alpha^{(1-2\varsigma)/\rho}). \tag{3.6}$$

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

Some comments on the requirement $\alpha = o(T^{\rho/(2\rho+2)})$ are first in order. This condition is needed in our proof of Theorem 3 to deal with estimation errors associated with $\widehat{\lambda}_j$ (see Remark 8). This may be replaced by a sufficient and more convenient condition given by $\alpha = o(T^{1/3})$, which does not depend on the value of ρ under Assumption E2 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 E2 can affect the convergence rate of the FIVE. Specifically, in the above theorem, the convergence rate is characterized by the regularization parameter α, the smoothness ς of $\mathcal{A}$ on $\operatorname{ran}\mathcal{C}_{xz}^*\mathcal{C}_{xz}$, and the decay rate ρ of λ_j^2 (as a

[1] In fact, Assumptions E2(a) and E2(b) can be replaced by the following conditions: $|\lambda_j| \leq c_\circ j^{-\rho/2}$ and $|\lambda_j| - |\lambda_{j+1}| \geq c_\circ j^{-\rho/2-1}$ 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).

function of j). From (3.6), it is evident that if α grows to infinity at a sufficiently slow rate, then the convergence rate of the RB term will be dominantly determined by the second term (appearing in (3.6)), whose convergence rate is positively related to α. Therefore, in this case, we expect a slower convergence rate of the RB component; 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. negatively) related to ς (resp. ρ); the former is immediately seen from (3.6), and the latter is discussed in detail in Remark 9.

Remark 7. In fact, the results given in Theorems 1 and 3 hold even if $\|\cdot\|_{\mathrm{op}}$ is replaced by $\|\cdot\|_{\mathrm{HS}}$, which can be seen in our proofs of those theorems (see Section S2 of the Supplementary Material).

Remark 8. The requirement $\alpha = o(T^{\rho/(2\rho+2)})$ is used to ensure the existence of a small constant, say $\tilde{c}_\circ$, such that $\mathbb{P}(|\widehat{\lambda}_j - \lambda_\ell| \geq \tilde{c}_\circ|\lambda_j - \lambda_\ell|$ for all $1 \leq j \leq \mathrm{K}$ and $\ell \neq j) \to 1$; the detailed reason why this result can be obtained from the requirement on α is given in our proof of Theorem 3, which is quite similar to the discussion given by Imaizumi and Kato (2018) following their Theorem 1.

Remark 9. If the eigenvalues of $\mathcal{C}_{xz}^*\mathcal{C}_{xz}$ decay to zero at a fast rate (and thus the eigenvalues of $\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz}$ tend to do so), then the rank-regularized inverse $(\widehat{\mathcal{C}}_{xz}^*\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}$ 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 as ρ increases. This can be seen from the asymptotic results given in Theorem 3. Let ρ change with the other parameters being fixed. In addition, it is assumed that α satisfies the condition $\alpha = o(T^{\rho/(2\rho+2)})$ both before and after any change in ρ and thus no adjustment in α is required; note that this can always be done by letting $\alpha = o(T^{1/3})$ if necessary. From (3.6), it can be shown that the RB component is (i) $O_p(T^{-1}\alpha) + O_p(\alpha^{(1-2\varsigma)/\rho})$ if $\varsigma \geq 3/2$ and (ii) $O_p(T^{-1}\alpha^{(\rho-2\varsigma+3)/\rho}) + O_p(\alpha^{(1-2\varsigma)/\rho})$ if $\varsigma \in (1/2, 3/2)$. In case (i), the decay rate of the second term is negatively related to ρ, and thus an increase in ρ does not yield a faster convergence rate. In case (ii), the decay rates of both terms are negatively related to ρ since $(\rho - 2\varsigma + 3)/\rho$ is positive and strictly increasing in ρ.

We next refine our pointwise asymptotic normality result under Assumption E2. To this end, it is convenient to decompose the RB component again as follows:

$$\mathcal{A}(\widehat{\Pi}_{\mathrm{K}} - \mathcal{I}) = \mathcal{A}(\widehat{\Pi}_{\mathrm{K}} - \Pi_{\mathrm{K}}) + \mathcal{A}(\Pi_{\mathrm{K}} - \mathcal{I}), \tag{3.7}$$

where $\Pi_{\mathrm{K}} = \sum_{j=1}^{\mathrm{K}} f_j \otimes f_j$, and this may be understood as the population counterpart of $\widehat{\Pi}_{\mathrm{K}}$. The next theorem refines the results in Theorem 2, but in order to simplify the subsequent discussion, we for now only consider the case where $\rho/2 + 2 < \varsigma + \delta_\varsigma$; the result without this condition is given in Section S2.3.2 of the Supplementary Material.

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

THEOREM 4. *Suppose that Assumptions M and E2 are satisfied, $\zeta \in \mathcal{H}$ satisfies $\langle f_j, \zeta \rangle \leq c_\zeta j^{-\delta_\zeta}$ for some $c_\zeta > 0$ and $\delta_\zeta > 1/2$, $\rho/2+2 < \varsigma + \delta_\zeta$ and*

$$T^{-1} \max\left\{\alpha^{(3\rho-2\delta_\zeta+1)/\rho}, \alpha^{(\rho+1)/\rho}\right\} = o(1). \tag{3.8}$$

Then, Theorem 2 holds and

$$\|\mathcal{A}(\widehat{\Pi}_K - \Pi_K)\zeta\| = O_p(T^{-1/2}) \quad \text{and} \quad \|\mathcal{A}(\Pi_K - \mathcal{I})\zeta\| = O_p(\alpha^{(1/2-\varsigma-\delta_\zeta)/\rho}).$$

Theorem 4 refines Theorem 2 by providing a detailed asymptotic order of the RB component. Some remarks on the theorem are given in Remarks 10 and 11; particularly, in the latter remark, an improvement of the asymptotic normality result in Theorem 2 is discussed.

Remark 10. The growing rate of α required for Theorem 4 depends on both ρ and δ_ζ. If δ_ζ is sufficiently large so that $2\delta_\zeta \geq \rho - 1$, then (3.8) can be simplified to $\alpha = o(T^{1/3})$. Moreover, the condition $\delta_\zeta > 1/2$ is natural since $\langle f_j, \zeta \rangle$ must be square-summable with respect to j.

Remark 11 (Pointwise asymptotic normality of the FIVE). A consequence of Theorem 4 is that, if $\mathcal{A}$ is smooth enough and $\langle f_j, \zeta \rangle$ decays to zero at a sufficiently fast rate as j increases, $\sqrt{T/\theta_K(\zeta)}\|\mathcal{A}(\widehat{\Pi}_K - \mathcal{I})\zeta\| \xrightarrow{p} 0$ and thus the result given in Theorem 2(i) can be strengthened to the following:

$$\sqrt{T/\theta_K(\zeta)}(\widehat{\mathcal{A}} - \mathcal{A})\zeta \xrightarrow{d} \mathcal{N}(0, \mathcal{C}_{uu}). \tag{3.9}$$

In this case, of course, (3.5) is understood as a confidence interval for $\langle \mathcal{A}\zeta, \psi \rangle$. In particular, if (i) $\varsigma + \delta_\zeta > \rho/2+2$ and (ii) $T\alpha^{(1-2\varsigma-\delta_\zeta)/\rho} = O(1)$ (note that both conditions are easier to hold if ς and δ_ζ are large), we have

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

The above quantity converges to zero if $\theta_K(\zeta) \xrightarrow{p} \infty$, which is likely to happen in practice for many possible choices of ζ; for example, if we assume that ζ is arbitrarily and randomly chosen from $\mathcal{H}$, $\mathbb{P}\{\theta_K(\zeta) < c < \infty\} \to 0$ as $K \to \infty$ since θ_K is convergent only on a strict subspace of $\mathcal{H}$. Given that $\delta_\zeta > 1/2$, the aforementioned conditions for (3.9) are satisfied if (i)$'$ $\varsigma > \rho/2+3/2$ and (ii)$'$ $T\alpha^{-2\varsigma/\rho} = O(1)$ regardless of the value of δ_ζ. The former condition (i)$'$ requires $\mathcal{A}$ to have a sufficient smoothness depending on the decay rate of the eigenvalues of $\mathcal{C}_{xz}^*\mathcal{C}_{xz}$, and this seems not to be restrictive; in the literature on functional linear models, it is common to impose such a smoothness condition on $\mathcal{A}$ 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).

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

4. FUNCTIONAL TWO-STAGE LEAST SQUARE ESTIMATOR

4.1. The Proposed Estimator

In the context of Euclidean space, where the covariance matrix of z_t is assumed to be invertible, our FIVE reduces to a certain IV estimator, as noted in Remark 1. As expected from existing results (see, e.g., Hausman, 1983), this IV estimator generally exhibits a larger asymptotic variance compared to the two-stage least squares estimator (2SLSE) when the model is over-identified. In a Hilbert space setting, covariance operators are not invertible and have eigenvalues that decay to zero, making it challenging to extend the conventional 2SLSE approach straightforwardly. In this section, we show that a functional version of the 2SLSE can be constructed using appropriate regularization and discuss its asymptotic properties. To the best of the authors' knowledge, a similar extension of the 2SLSE was considered by Florens and Van Bellegem (2015) for a scalar-valued dependent variable but has not been explored in the context of function-on-function regression. As detailed in Remark 12, despite their resemblance, the theoretical properties of the proposed estimator differ from those of the conventional 2SLSE, which is mainly due to the noninvertible covariance of the IV and its eigenvalues decaying to zero.

In this section, we assume that x_t and z_t satisfy the so-called first-stage relationship:

Assumption M*. (a) Assumption M holds, (b) $x_t = \mathcal{B}z_t + v_t$, where $\mathcal{B} \in \mathcal{L}_{\mathcal{H}}$ and $\mathbb{E}[v_t|\mathfrak{F}_{t-1}] = 0$ ($\mathfrak{F}_{t-1}$ is defined in Assumption M), (c) $\widehat{\mathcal{C}}_{vz} = T^{-1}\sum_{t=1}^{T} v_t \otimes z_t$ satisfies that $\|\widehat{\mathcal{C}}_{vz}\|_{\mathrm{HS}} = O_p(T^{-1/2})$.

If we consider the case $\mathcal{H} = \mathbb{R}^n$, then the 2SLSE is defined as follows:

$$\widetilde{\mathcal{A}}^{\circ} = \widehat{\mathcal{C}}_{yz}^{*}\widehat{\mathcal{C}}_{zz}^{-1}\widehat{\mathcal{C}}_{xz}\left(\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{zz}^{-1}\widehat{\mathcal{C}}_{xz}\right)^{-1},$$

and it is widely known that $\widetilde{\mathcal{A}}^{\circ}$ has many desirable properties as an estimator of $\mathcal{A}$. Coming back to our functional setting, it is not difficult to see that the use of the standard 2SLSE is problematic since it involves $\widehat{\mathcal{C}}_{zz}^{-1}$ and $(\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{zz}^{-1}\widehat{\mathcal{C}}_{xz})^{-1}$ 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 $\mathcal{T}_{\mathrm{K}}^{-1}$ denote the regularized inverse of a compact operator $\mathcal{T}$ based on its first K eigenelements (this is defined in the same way as $(\widehat{\mathcal{C}}_{xz}^{*}\widehat{\mathcal{C}}_{xz})_{\mathrm{K}}^{-1}$ given in (3.3)). Our proposed estimator is defined as follows:

$$\widetilde{\mathcal{A}} = \widehat{\mathcal{P}}\widehat{\mathcal{Q}}_{\mathrm{K}_2}^{-1}, \quad \text{where} \quad \widehat{\mathcal{P}} = \widehat{\mathcal{C}}_{yz}^{*}(\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1}\widehat{\mathcal{C}}_{xz} \quad \text{and} \quad \widehat{\mathcal{Q}} = \widehat{\mathcal{C}}_{xz}^{*}(\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1}\widehat{\mathcal{C}}_{xz},$$

and, if we let $\{\widehat{\mu}_j\}_{j\geq 1}$ (resp. $\{\widehat{\nu}_j\}_{j\geq 1}$) be the ordered (from the largest to the smallest) eigenvalues of $\widehat{\mathcal{C}}_{zz}$ (resp. $\widehat{\mathcal{Q}}$),[2] then K_1 and K_2 are defined as

[2] The eigenvalues of $\widehat{\mathcal{C}}_{zz}$ and $\widehat{\mathcal{Q}}$ are almost surely positive since they are nonnegative and self-adjoint by construction.

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

$$\mathrm{K}_1 = \#\{j : \widehat{\mu}_j^2 > 1/\alpha_1\} \quad \text{and} \quad \mathrm{K}_2 = \#\{j : \widehat{\nu}_j^2 > 1/\alpha_2\}.$$

Note that by definition, $\mathrm{K}_2 \leq \mathrm{K}_1$ holds almost surely because $(\widehat{\mathcal{C}}_{zz})_{\mathrm{K}_1}^{-1}$ is of finite rank K_1. We conveniently call $\widetilde{\mathcal{A}}$ 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 $\mathcal{Q}$ defined by $\mathcal{Q} = \mathcal{C}_{xz}^* \mathcal{C}_{zz}^{-1} \mathcal{C}_{xz}$ can be understood as a well-defined compact operator. To be more specific, Lemma S3 (and the following discussion given in Section S3 of the Supplementary Material) shows that $\mathcal{C}_{zz}^{-1/2}\mathcal{C}_{xz} = \mathcal{R}_{xz}\mathcal{C}_{xx}^{1/2}$ for some unique bounded linear operator $\mathcal{R}_{xz}$, which may be understood as the correlation operator of x_t and z_t, and thus $\mathcal{Q} = \mathcal{C}_{xx}^{1/2}\mathcal{R}_{xz}^*\mathcal{R}_{xz}\mathcal{C}_{xx}^{1/2}$. From similar arguments, it can be easily shown that the operator $\mathcal{P} = \mathcal{C}_{yz}^*\mathcal{C}_{zz}^{-1}\mathcal{C}_{xz}$ is also well defined. We then let $\{\mu_j, g_j\}_{j\geq 1}$ (resp. $\{\nu_j, h_j\}_{j\geq 1}$) be the eigenelements of $\mathcal{C}_{zz}$ (resp. $\mathcal{Q}$), that is,

$$\mathcal{C}_{zz} = \sum_{j=1}^{\infty} \mu_j g_j \otimes g_j \quad \text{and} \quad \mathcal{Q} = \sum_{j=1}^{\infty} \nu_j h_j \otimes h_j.$$

Since $\mathcal{C}_{zz}$ and $\mathcal{Q}$ are self-adjoint and nonnegative, μ_j and ν_j are all nonnegative. We know from (2.2) that the population relationship $\mathcal{P} = \mathcal{A}\mathcal{Q}$ holds.

4.2. General Asymptotic Properties

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

Assumption E1*. $\mu_1 > \mu_2 > \cdots > 0$ and $\nu_1 > \nu_2 > \cdots > 0$.

It should be noted that the operator $\mathcal{A}$ satisfying $\mathcal{P} = \mathcal{A}\mathcal{Q}$ is uniquely identified if Assumptions M* and E1* are satisfied. To see this in detail, note that $\mathcal{A}$ satisfying $\mathcal{C}_{yz}^*\mathcal{C}_{xz} = \mathcal{A}\mathcal{C}_{xz}^*\mathcal{C}_{xz}$ (resp. $\mathcal{P} = \mathcal{A}\mathcal{Q}$) is identified if and only if $\mathcal{C}_{xz}$ (resp. $\mathcal{C}_{zz}^{-1/2}\mathcal{C}_{xz}$) is injective. If Assumption E1* is satisfied and hence $\mathcal{C}_{zz}$ is injective, then $\ker \mathcal{C}_{xz} (= \ker \mathcal{C}_{xz}^*\mathcal{C}_{xz})$ becomes identical to $\ker \mathcal{C}_{zz}^{-1/2}\mathcal{C}_{xz} (= \ker \mathcal{Q})$.

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

$$\widetilde{\mathcal{A}} - \mathcal{A} = (\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}) - \mathcal{A}(\mathcal{I} - \widetilde{\Pi}_{\mathrm{K}_2}), \tag{4.1}$$

where $\widetilde{\Pi}_{\mathrm{K}_2}$ denotes the orthogonal projection defined by $\widetilde{\Pi}_{\mathrm{K}_2} = \sum_{j=1}^{\mathrm{K}_2} \widehat{h}_j \otimes \widehat{h}_j$ and $\{\widehat{h}_j\}_{j=1}^{\mathrm{K}_2}$ is the collection of the eigenvectors of $\widehat{\mathcal{Q}}$ corresponding to the first K_2 leading eigenvalues. The two terms in (4.1) are similarly interpreted as in the case of the FIVE (see (3.7)), and we thus call the first (resp. second) term the DR (resp. RB) component.

We first show that both the DR and RB components are asymptotically negligible (and thus $\widehat{\mathcal{A}}$ is weakly consistent) if the regularization parameters α_1 and α_2 diverge to infinity 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 $\tau_{2,j} = 2\sqrt{2}\max\{(\nu_{j-1} - \nu_j)^{-1}, (\nu_j - \nu_{j+1})^{-1}\}$.

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

THEOREM 5. *Suppose that Assumptions M* and E1* are satisfied, and* $T^{-1/2}(\sum_{j=1}^{K_1}\mu_j\tau_{1,j})(\sum_{j=1}^{K_2}\tau_{2,j}) \xrightarrow{p} 0$, $(\sum_{j=K_1+1}^{\infty}\mu_j)(\sum_{j=1}^{K_2}\tau_{2,j}) \xrightarrow{p} 0$, $\alpha_1^{-1}\alpha_2 \to 0$, *and* $T^{-1}\alpha_1 \to 0$ *as* $\alpha_1 \to \infty$, $\alpha_2 \to \infty$ *and* $T \to \infty$. *Then*

$$\|\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{K_2}\|_{op}^2 = O_p(T^{-1}\alpha_1^{1/2}\alpha_2^{1/2}) \quad \text{and} \quad \|\mathcal{A}(\mathcal{I} - \widetilde{\Pi}_{K_2})\|_{op}^2 = o_p(1).$$

An immediate consequence of Theorem 5 is given as follows:

COROLLARY 2. *Suppose that the assumptions in Theorem 2 are satisfied and let* $\tilde{u}_t = y_t - \widetilde{\mathcal{A}}x_t$. *Then* $\|T^{-1}\sum_{t=1}^{T}\tilde{u}_t \otimes \tilde{u}_t - \mathcal{C}_{uu}\|_{op} \xrightarrow{p} 0$.

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 growing rates of α_1 and α_2 in our asymptotic theory. Specifically, the condition on the former quantity merely requires α_1 and α_2 to increase slowly so that K_1 and K_2 tend to grow with sufficiently slower rates than T. Moreover, given that, for fixed α_2, $(\sum_{j=K_1+1}^{\infty}\mu_j)(\sum_{j=1}^{K_2}\tau_{2,j})$ can be arbitrarily small by choosing α_1 large enough, the condition on the latter quantity merely tells us that the growing rate of α_1 needs to be sufficiently higher than that of α_2. In addition to the weak consistency given by Theorem 5, the strong consistency of the F2SLSE can be derived under additional conditions; this is discussed in Section 3.3.1 of 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,* $T^{-1/2}\alpha_1^{1/2}\sum_{j=1}^{K_1}\tau_{1,j} \xrightarrow{p} 0$, $T^{-1/2}\alpha_2^{1/2}(\sum_{j=1}^{K_1}\mu_j\tau_{1,j})(\sum_{j=1}^{K_2}\tau_{2,j}) \xrightarrow{p} 0$, $\alpha_2^{1/2}(\sum_{j=K_1+1}^{\infty}\mu_j)$ $(\sum_{j=1}^{K_2}\tau_{2,j}) \xrightarrow{p} 0$, $\alpha_1^{-1}\alpha_2 \to 0$, *and* $T^{-1}\alpha_1 \to 0$ *as* $\alpha_1 \to \infty$, $\alpha_2 \to \infty$ *and* $T \to \infty$. *Then the following hold for any* $\zeta \in \mathcal{H}$.

(i) $\sqrt{T/\phi_{K_2}(\zeta)}(\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{K_2})\zeta \xrightarrow{d} \mathcal{N}(0, \mathcal{C}_{uu})$, *where* $\phi_{K_2}(\zeta) = \langle\zeta, \mathcal{Q}_{K_2}^{-1}\zeta\rangle$.

(ii) *If* $\widehat{\phi}_{K_2}(\zeta) := \langle\zeta, \widehat{\mathcal{Q}}_{K_2}^{-1}\zeta\rangle$, *then* $|\widehat{\phi}_{K_2}(\zeta) - \phi_{K_2}(\zeta)| \xrightarrow{p} 0$.

As shown, we need more stringent requirements on the growing rates of α_1 and α_2. This is mainly due to that the F2SLSE involves the doubly regularized inverse $\widehat{\mathcal{Q}}_{K_2}^{-1}$, which may be ill-behaved if α_1 and α_2 do not grow 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 the Euclidean space setting.[3]

[3]Moreover, the two estimators have different RB terms whose magnitudes depend on various parameters (e.g., the eigenvalues of $\mathcal{C}_{xz}^*\mathcal{C}_{xz}$ and $\mathcal{Q}$), and thus the RB component of the FIVE can have a smaller asymptotic order.

Remark 12. While the functional form of the F2SLSE closely resembles the conventional 2SLSE in the Euclidean space setting, a crucial distinction exists. In the conventional case, the 2SLSE gains its theoretical superiority by efficiently combining regressors using a larger number of instruments. However, in our setting, one endogenous functional regressor is instrumented by another functional variable. This explains, at least to some degree, why the F2SLSE does not supersede the FIVE and why the standard properties of the 2SLSE do not naturally extend to our functional framework. For example, for any element $\zeta \in \mathcal{H}$, Theorems 2 and 6 tell us that the DR components of the FIVE and F2SLSE, respectively, converge to the same Gaussian random element with $\sqrt{T/\theta_{\mathrm{K}}(\zeta)}$-rate and $\sqrt{T/\phi_{\mathrm{K}_2}(\zeta)}$-rate, which are possibly random quantities depending on ζ. Given this, it is not generally possible to conclude that the F2SLSE is asymptotically better than the FIVE, although we disregard the RB components of the FIVE and F2SLSE. As our extension to the F2SLSE, it may be possible to extend the generalized least squares (GLS) methods by employing a regularized inverse of the (estimated) covariance of the error term. However, due to the additional regularization, we expect that the well-known properties of the GLS estimator in a finite-dimensional setting may not be easily translated into the considered functional setting. The insight of this remark originates from comments by the editor and an anonymous referee, to whom we are indebted.

4.3. Refinements of the General Asymptotic Results

We provide refinements of Theorems 5 and 6 under the following set of assumptions which is stronger than Assumption E1*: below, we let $\tilde{\upsilon}_t(j,\ell) = \langle z_t, g_j\rangle\langle z_t, g_\ell\rangle - \mathbb{E}[\langle z_t, g_j\rangle\langle z_t, g_\ell\rangle]$ for $j,\ell \geq 1$.

Assumption E2*. There exist constants $c_\circ > 0$, $\rho_\mu > 2$, $\rho_\nu > 2$, $\varsigma_\mu > 1/2$, $\varsigma_\nu > 1/2$, $\gamma_\mu > 1/2$, $\gamma_\nu > 1/2$ and $m > 1$ satisfying the following: (a) $\mu_j^2 \leq c_\circ j^{-\rho_\mu}$, (b) $\mu_j^2 - \mu_{j+1}^2 \geq c_\circ^{-1} j^{-\rho_\mu - 1}$, (c) $\nu_j^2 \leq c_\circ j^{-\rho_\nu}$, (d) $\nu_j^2 - \nu_{j+1}^2 \geq c_\circ^{-1} j^{-\rho_\nu - 1}$, (e) $\langle h_j, \mathcal{A}h_\ell\rangle \leq c_\circ j^{-\gamma_\nu}\ell^{-\varsigma_\nu}$, (f) $\langle h_j, \mathcal{B}g_\ell\rangle \leq c_\circ j^{-\gamma_\mu}\ell^{-\varsigma_\mu}$ and $\gamma_\mu \leq \rho_\nu/4 + 1/2$, (g) $\mathbb{E}[\tilde{\upsilon}_t(j,\ell)\tilde{\upsilon}_{t-s}(j,\ell)] \leq c_\circ s^{-m}\mathbb{E}[\tilde{\upsilon}_t^2(j,\ell)]$ for $s \geq 1$, $\mathbb{E}[\|\langle z_t, g_j\rangle z_t\|^2] \leq c_\circ \mu_j^2$, and $\mathbb{E}[\|\langle x_t, h_j\rangle z_t\|^2] \leq c_\circ \|\mathcal{C}_{xz} h_j\|^2$.

The conditions are somewhat similar to those in Assumption E2, and thus we omit detailed comments except the following two points: (i) from a technical point of view, Assumption E2*(g) is similar to Assumption E2(d) employed for our study of the FIVE and helps us obtain convergence rates of the eigenelements of $\widehat{\mathcal{Q}}$ (which are crucial inputs to our main results of the F2SLSE), and (ii) we require a smoothness condition on $\mathcal{B}$ which characterizes the linear relationship between x_t and z_t whereas such a condition is not necessary in the case of the FIVE. This reveals that the DGP is more restricted for our asymptotic analysis of the F2SLSE.

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

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

THEOREM 7. *Suppose that Assumptions M* and E2* are satisfied,* $\alpha_1 = o(T^{\rho_\mu/(2\rho_\mu+2)})$ *and* $\alpha_2 = o(\alpha_1^{\rho_\nu/(2\rho_\nu+2)})$. *Then,* $\|\widetilde{\mathcal{A}} - \mathcal{A}\widetilde{\Pi}_{K_2}\|_{op}^2 = O_p(T^{-1}\alpha_1^{1/2}\alpha_2^{1/2})$ *as in Theorem 5, and*

$$\|\mathcal{A}(\mathcal{I} - \widetilde{\Pi}_{K_2})\|_{op}^2 = O_p(\alpha_1^{-1}\alpha_2 \max\{1, \alpha_2^{(3-2\varsigma_\nu)/\rho_\nu}\} + \alpha_2^{(1-2\varsigma_\nu)/\rho_\nu}). \tag{4.2}$$

Thus, $\|\widetilde{\mathcal{A}} - \mathcal{A}\|_{op} = o_p(1)$ *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 also on smoothness of $\mathcal{A}$ as in the case of the FIVE. However, the convergence rate described in (4.2) 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 competing estimators (including the FIVE), and thus this estimator can also be used in practice.

Using Assumption E2*, the next theorem refines Theorem 6, but as in Section 3.3, we for now only focus on the case where $\rho_\nu/2 + 2 < \varsigma_\nu + \delta_\zeta$. The result without this condition is provided in the Supplementary Material (see Section S3.3.2). In the theorem below, we, as in (3.7), consider the decomposition of the RB component given by

$$\mathcal{A}(\widetilde{\Pi}_{K_2} - \mathcal{I}) = \mathcal{A}(\widetilde{\Pi}_{K_2} - \Pi_{K_2}) + \mathcal{A}(\Pi_{K_2} - \mathcal{I}),$$

where $\Pi_{K_2} = \sum_{j=1}^{K_2} h_j \otimes h_j$ is understood as the population counterpart of $\widetilde{\Pi}_{K_2}$.

THEOREM 8. *Suppose that Assumptions M* and E2* are satisfied,* $\zeta \in \mathcal{H}$ *satisfies* $\langle h_j, \zeta \rangle \leq c_\zeta j^{-\delta_\zeta}$ *for some* $c_\zeta > 0$ *and* $\delta_\zeta > 1/2$, $\rho_\nu/2 + 2 < \varsigma_\nu + \delta_\zeta$, *and the following hold:*

$$\alpha_1 = o(T^{\rho_\mu/(2\rho_\mu+2)}), \quad \alpha_1^{-1} \max\left\{\alpha_2^{(3\rho_\nu - 2\delta_\zeta + 1)/\rho_\nu}, \alpha_2^{(\rho_\nu+1)/\rho_\nu}\right\} = o(1). \tag{4.3}$$

Then Theorem 6 holds, and furthermore,

$$\|\mathcal{A}(\widetilde{\Pi}_{K_2} - \Pi_{K_2})\zeta\| = O_p(\alpha_1^{-1/2}) \quad \text{and} \quad \|\mathcal{A}(\Pi_{K_2} - \mathcal{I})\zeta\| = O_p(\alpha_2^{(1/2 - \varsigma_\nu - \delta_\zeta)/\rho_\nu}).$$

Obtaining a pointwise asymptotic normality result that is not dependent on the RB component as in Remark 11 requires more stringent conditions, which will be detailed in Remark 13.

Remark 13 (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, as in Remark 11, suppose that $\mathcal{A}$ is smooth enough so that $\varsigma_\nu > \rho_\nu/2 + 3/2$ and $T\alpha_2^{(1-2\varsigma_\nu - 2\delta_\zeta)/\rho_\nu} = O(1)$. We know from Theorem 8 that $\sqrt{T/\phi_{K_2}(\zeta)}\|\mathcal{A}(\widetilde{\Pi}_{K_2} - \mathcal{I})\zeta\| = O_p(\sqrt{T\alpha_1^{-1}/\phi_{K_2}(\zeta)})$,

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

and find that $\sqrt{T/\phi_{\mathrm{K}_2}(\zeta)}(\mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}-\mathcal{A})\zeta \xrightarrow{d} \mathcal{N}(0,\mathcal{C}_{uu})$ if $\phi_{\mathrm{K}_2}(\zeta)$ diverges at a faster rate than that of $T\alpha_1^{-1}$. In the case of the FIVE and under an analogous smoothness condition, recall that only $\theta_{\mathrm{K}}(\zeta) \xrightarrow{p} \infty$ is needed to obtain a similar result; see Remark 11.

5. NUMERICAL STUDIES

We first investigate the finite sample performance of our estimators via Monte Carlo studies. In Sections 5.1–5.3, the number of replications is set to 1,000 and all the considered random variables are demeaned before computing the estimators of $\mathcal{A}$. Section 5.4 provides an empirical application.

5.1. Experiment 1: Functional Linear Simultaneous Equation Model

We consider the following functional linear simultaneous equation model: for $t \geq 1$,

$$y_t = \mathcal{A}x_t + u_t, \qquad x_t = \vartheta\mathcal{B}z_t + v_t, \tag{5.1}$$

where $u_t = 0.8v_t + 0.6\varepsilon_t$, $\{v_t\}_{t\geq 1}$ and $\{\varepsilon_t\}_{t\geq 1}$ are mutually independent i.i.d. sequences of standard Brownian bridges satisfying $\mathbb{E}[v_t \otimes \varepsilon_\ell] = 0$ for all $t,\ell \geq 1$. The constant ϑ is chosen in such a way that the first-stage functional coefficient of determination (see Yao et al., 2005), defined by $\mathbb{E}[\|\vartheta\mathcal{B}z_t\|^2]/\mathbb{E}[\|x_t\|^2]$, has a specific value of r^2. In this section, we will focus on empirical MSEs of a few estimators at various levels of r^2, and in particular we consider $\mathrm{r}^2 \in \{0.1, 0.2, \ldots, 0.5\}$.

The DGP here is specially designed to examine the performance of our estimators when the employed assumptions (Assumptions M, E1, E2, M*, E1*, and E2*) are satisfied (see Section S4.1 of the Supplementary Material). Specifically, we let $\{z_t\}_{t\geq 1}$ be an i.i.d. sequence of standard Brownian bridges satisfying $\mathbb{E}[z_t \otimes v_t] = \mathbb{E}[z_t \otimes u_t] = 0$. Then, we have $\mu_j = (j\pi)^{-2}$ and $g_j(s) = \sqrt{2}\sin(j\pi s)$ for $s \in [0,1]$, see, for example, Jaimez and Bonnet (1987). The operators $\mathcal{A}$ and $\mathcal{B}$ are defined as follows:

$$\mathcal{A} = \sum_{j=1}^{\infty} a_j g_j \otimes g_j, \quad \mathcal{B} = \sum_{j=1}^{\infty} b_j g_j \otimes g_j, \quad a_j = j^{-n_a}, \quad b_j = j^{-n_b}, \quad n_a \in \{3,5\}, \quad n_b \in \{0.75, 1.5\}.$$

In this setup, $f_j = g_j$. In view of the fact that function-valued random variables are only partially observed in practice, we assume that the discrete realizations of y_t, x_t, and z_t at 50 equally spaced points on $[0,1]$ are available. Then, following the literature, for example, Ramsay and Silverman (2005, Chap. 5), we represent functional variables y_t, x_t and z_t by using 31 Fourier basis functions.

We will compare the performance of our estimators with the ridge regularized IV estimator (RIVE) of Benatia et al. (2017, eqn. (34)) with denoting their regularization parameter to α^{-1} to keep notational consistency. To compute the

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

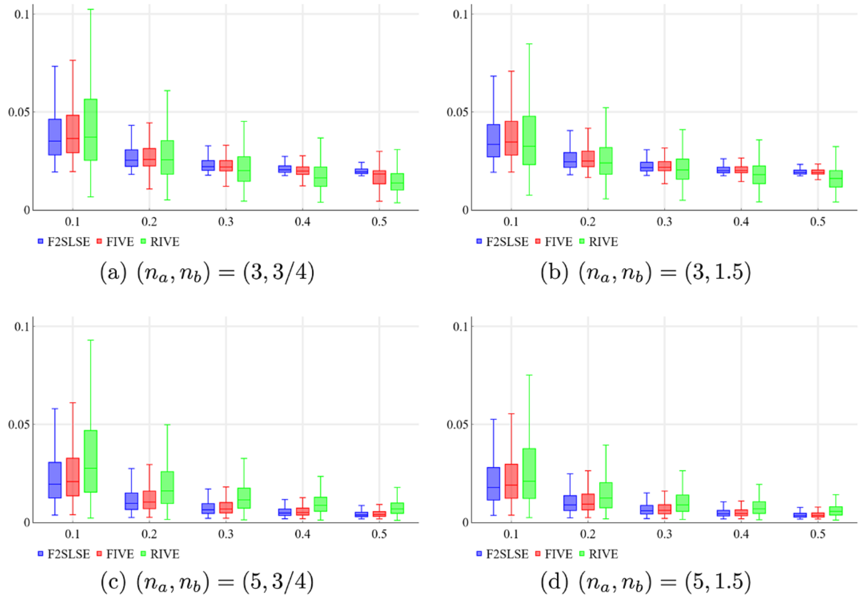


FIGURE 1. Boxplots of the empirical MSEs ($T = 500$).
Notes: Boxplots of the empirical mean squared errors (MSEs) of the FIVE (red), the F2SLSE (blue), and the RIVE (green) are reported for each value of the first-stage functional coefficient of determination $\mathrm{r}^2 \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

FIVE and RIVE, we consider $\delta_\alpha T^{-0.4}\|\widehat{\mathcal{C}}_{xz}\|_{\mathrm{HS}}^2$ as candidates for the inverse of α. This candidate value is calculated at 20 equidistant points of δ_α ranging from 0.1 to $T^{0.2}$. Among such candidates, we choose the value that minimizes the empirical MSE of each estimator. The F2SLSE needs two regularization parameters: α_1 and α_2. The parameter α_1 is chosen as the FIVE and RIVE with $\|\widehat{\mathcal{C}}_{xz}\|_{\mathrm{HS}}^2$ being replaced by $\|\widehat{\mathcal{C}}_{zz}\|_{\mathrm{HS}}^2$. Once α_1 is chosen, we similarly choose the inverse of α_2 from $\delta_{\alpha_2}(\alpha_1^{-1}\|\mathcal{C}_{zz}\|_{\mathrm{HS}}^2)^{1/2}\|\widehat{\mathcal{Q}}_{\mathrm{K}_1}\|_{\mathrm{HS}}^2$ with δ_{α_2} being 20 equidistant points between $T^{0.05}$ and $T^{0.2}$. This setup enforces α_1 to grow at a faster rate than that of α_2.

To save space, we report estimation results only for the case with $T = 500$; the results with a smaller sample size are qualitatively similar and are reported in Section S4.4 of the Supplementary Material. Figure 1 reports boxplots (without outliers) of the empirical MSE estimated with the FIVE (red), the F2SLSE (blue), and the RIVE (green). The first interesting observation in the figure is that our estimators tend to produce smaller MSEs when the signal from x_t to y_t is more concentrated on the first few components, that is, when $n_a = 5$. This observation is consistent regardless of the values of n_b and r^2. This may not be surprising because when n_a is large, the first few f_j's ($=g_j$'s) summarize the most significant information of $\mathcal{A}$.

In the figure, as the value of r^2 decreases, the considered estimators tend to exhibit larger MSEs. Similar observations can be found in the standard IV literature, in which the so-called concentration parameter is used to measure the strength of IVs. Given that the coefficient of determination r^2 is closely related to the concentration parameter in the IV literature, such a larger MSE may be understood as the distortion related to weak instruments.

In the subsequent sections, we will consider a more general setting to investigate the robustness of our estimators when the assumptions are unlikely to hold. Unlike the DGP considered in this section, it is nearly impossible to verify if all the required conditions are satisfied for the DGPs under consideration. Given the practical challenges of confirming these conditions, practitioners may find it valuable to observe the performance of our estimators in the presence of potential violations of the required conditions.

5.2. Experiment 2: A Modification of Benatia et al.'s (2017) Simulation DGP

In this section, we consider a simulation DGP similar to that in Benatia et al.'s (2017). Specifically, we let $\mathcal{B}$ in (5.1) be the identity operator $\mathcal{I}$ and let $\mathcal{A}$ be the integral operator with kernel $\kappa_{\mathcal{A}}(s_1,s_2) = 1-|s_1-s_2|^2$ for $s_1,s_2 \in [0,1]$. In this setup, the first-stage signal is solely determined by the constant ϑ. The IV z_t is given as follows:

$$z_t(s) = \tilde{z}_t(s;a_t,b_t)+\eta_t(s) \quad \text{and} \quad \tilde{z}_t(s;a_t,b_t) = \frac{\Gamma(a_t)\Gamma(b_t)}{\Gamma(a_t+b_t)} s^{a_t-1}(1-s)^{b_t-1} \tag{5.2}$$

for $s \in [0,1]$, 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 η_t to the beta density function with parameters a_t and b_t. The IV in (5.2) is analogous to that used for the simulation experiments in Benatia et al. (2017), in which the additive noise $\eta_t(s)$ is given by q_t for all $s \in [0,1]$ with q_t being randomly drawn from $\mathcal{N}(0,1)$. In this section, we allow a more general form of η_t by letting $\eta_t = \sum_{j=1}^{n_J} \sigma_j q_{t,j} \xi_j$ where $n_J = 31$, $\{\xi_j\}_{j\geq 1}$ is the Fourier basis functions with the constant basis function ξ_1, and $q_{t,j} \sim_{\text{i.i.d.}} \mathcal{N}(0,1)$ across t and j. Benatia et al.'s (2017) setup can be understood to the case $n_J = 1$ under our notation. Then, we consider three different designs of $\{\sigma_j\}_{j\geq 1}$. First, we consider the case $\sigma_j = c_1\sigma_\eta$ for $j \leq 2$ and $\sigma_j = c_1\sigma_\eta(0.1)^{j-2}$ for $j > 2$; this is called the sparse design. Second, we set $\sigma_j = \sigma_\eta(0.9)^{j-1}$ and call this setting the exponential design. In the last design, which we call the geometric design, we let $\sigma_j = c_2\sigma_\eta j^{-1}$. The parameter σ_η is set to 0.5 and 0.9, and the constants c_1 and c_2 are chosen in such a way as to have the same Hilbert–Schmidt norm of $\mathbb{E}[\eta_t \otimes \eta_t]$ in all three designs. Lastly, the parameter ϑ is chosen as in Section 5.1 with r^2 being set to 0.5; this can be done by using that $\mathbb{E}[\|v_t\|^2] = 1/6$, $\mathbb{E}[\|\eta_t\|^2] = \sum_{j=1}^{31}\sigma_j^2$, and the value of $\mathbb{E}[\|\tilde{z}_t(s_i;a_t,b_t)\|^2]$ can be approximated from a large number of simulations.

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

Table 1. Simulation results for Experiment 2: empirical MSEs and coverage probabilities

	Sparse design				Exponential design				Geometric design			
σ_η	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
Coverage probability for $\langle \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\zeta, \psi\rangle$ or $\langle \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}\zeta, \psi\rangle$												
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 probability for $\langle \mathcal{A}\zeta, \psi\rangle$												
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

Note: Based on 1,000 replications. In the top, each cell reports the empirical mean squared error (MSE) of three estimators: FIVE, F2SLSE, and Benatia et al.'s (2017) RIVE. The last four rows report the coverage probabilities of the designated quantities; the nominal level is 95%.

Table 1 summarizes simulation results. Overall, the MSEs of our estimators and those of the RIVE are similar to each other. However, in the exponential design, our estimators tend to have smaller MSEs compared to the RIVE.

We note that our estimators and related asymptotic results can be used to discuss the coverage probability of the interval (3.5) that is computed from the FIVE. The interval is expected to contain the random quantity $\langle \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\zeta, \psi\rangle$ with $(100-\varpi)\%$ of probability; moreover, if certain conditions are satisfied (see Remark 11) the interval (3.5) can be understood as the $(100-\varpi)\%$ confidence interval for $\langle \mathcal{A}\zeta, \psi\rangle$ which is nonrandom. Based on Theorem 6 (and also Remark 13), we may construct a similar interval with the F2SLSE and the interval is expected to include $\langle \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}\zeta, \psi\rangle$ (and also $\langle \mathcal{A}\zeta, \psi\rangle$ under certain conditions) with $(100-\varpi)\%$ 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 ζ be randomly generated by $\zeta = \sum_{j=1}^{11} \ddot{q}_{1,j}\ell_j$ for each realization of the DGP, where $\{\ell_j\}_{j\geq 1}$ is the polynomial basis with the constant basis function ℓ_1 and $\ddot{q}_j \sim_{\text{i.i.d.}} \mathcal{N}(0, j^{-4})$ 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 $\langle \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\zeta, \psi\rangle$ or $\langle \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}\zeta, \psi\rangle$ 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 $\langle \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\zeta, \psi\rangle$ or $\langle \mathcal{A}\widetilde{\Pi}_{\mathrm{K}_2}\zeta, \psi\rangle$, they are still reasonably close to the nominal level of 95%. This is what can be expected

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

from Remarks 11 and 13. In unreported simulations, we further experimented with different choices of ζ and ψ, but found no significant difference.

5.3. Experiment 3: AR(1) Model of Probability Density Functions

In this section, we 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 Sections 5.1 and 5.2 is that endogeneity is not explicitly imposed, but implicitly introduced by estimation errors.

We let $\{p_t^\circ\}_{t\geq 1}$ be a sequence of probability densities supported on $[0,1]$, and consider the linear prediction model of p_t° given p_{t-1}°. 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. As a result, a direct application of the statistical methods developed in a Hilbert space setting may not be recommended; see, for example, Delicado (2011), Petersen and Müller (2016), Hron, Menafoglia, Templ, Hrůzová and Filzmoser (2016), Kokoszka et al. (2019), and Zhang, Kokoszka, and Petersen (2021). As a way to circumvent such issues, we consider the centered-log-ratio (clr) transformation $y_t^\circ(s) = \log p_t^\circ(s) - \int \log p_t^\circ(s)ds$, $s \in [0,1]$ (see, e.g., Egozcue, Díaz-Barrero, and Pawlowsky-Glahn, 2006). Then, $\{y_t^\circ\}_{t\geq 1}$ turns out to be 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. Any element in $\mathcal{H}_c$ may be understood as a probability density via the inverse transformation $y_t^\circ(s) \mapsto \exp(y_t^\circ(s))/\int_0^1 \exp(y_t^\circ(s))ds$. Thus, the linear prediction model of p_t° given p_{t-1}° may be recast into that of y_t° given y_{t-1}° in $\mathcal{H}_c$. We thus consider the following prediction model:

$$y_t^\circ = c_y + \mathcal{A}(y_{t-1}^\circ - c_y) + \varepsilon_t,$$

where y_{t-1}° and ε_t are uncorrelated. To mimic situations commonly encountered in practice, we assume that p_t° (and thus y_t°) is not observed, but only random samples $\{s_{i,t}\}_{i=1}^{n_t}$ drawn from p_t° are available. If so, by replacing the density p_t° or the log-density $\log p_t^\circ$ with its proper nonparametric estimate, we may obtain an estimate y_t of y_t°. As shown in Example 2, $\{y_t\}_{t\geq 1}$ satisfies $y_t = c_y + \mathcal{A}(y_{t-1} - c_y) + u_t$, but now y_{t-1} and u_t are generally correlated due to errors arising from the nonparametric estimation. 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 very often have no choice but to replace p_t° or $\log p_t^\circ$ with a standard nonparametric estimate, to see if a naive use of our estimators can make any actual improvements in estimating $\mathcal{A}$. This is the purpose of simulation experiments in this section.

Specifically, we first estimate y_t° from n random samples that are generated from p_t° by (i) the local likelihood density estimation method proposed by Loader (1996) (see Section S4.2 of the Supplementary Material for more details) and (ii) the standard kernel density estimation method with the Gaussian kernel and Silverman's rule-of-thumb bandwidth (Silverman, 1998). Even if the former is more suitable for

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

the estimation of y_t° (Seo and Beare, 2019, Sect. 4.2), the latter is considered as well because of its popularity in empirical studies. Once y_t is computed, it is represented by the first 30 nonconstant Fourier basis functions for implementation of the FPCA in $\mathcal{H}_c$. We let c_y be the clr transformation of the normal density function with mean 0.5 and variance 0.25^2 that is truncated on $[0,1]$. In addition, $\varepsilon_t = \sum_{j=1}^{\infty} \sigma_j q_{t,j} \xi_j^c$, where $\{\xi_j^c\}_{j\geq 1}$ is the set of Fourier basis functions except for the constant basis function, and $q_{t,j} \sim_{\text{i.i.d.}} \mathcal{N}(0,1)$ across t and j.[4] Below we consider two different specifications of σ_j, which are, respectively, called the exponential design and the sparse 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_σ is chosen to ensure the same Hilbert–Schmidt norm of $\mathbb{E}[\varepsilon_t \otimes \varepsilon_t]$ in both designs. These two designs are respectively obtained by setting σ_η to 0.1 in the sparse and exponential designs considered for η_t in Section 5.2; the reason that we choose a relatively smaller scale of σ_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 $\sum_{j=1}^{\infty} a_j \xi_j^c \otimes \xi_j^c$,[5] and, for each realization of the DGP, the coefficients $\{a_j\}_{j\geq 1}$ are independently determined across j as follows:

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

Note that we let the first two coefficients a_1 and a_2 be bounded below by 0.4 to ensure that the operator norm of the cross-covariance of y_{t-1}° and y_{t-2}° 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 article.

Table 2 reports the empirical MSEs of the proposed estimators and Park and Qian's (2012) functional least squares estimator (FLSE) when $n = 100$ and 150 (recall that n is the number of random samples drawn from the distribution p_t° to estimate $\log p_t^\circ$ or p_t°). The IV estimators tend to exhibit smaller empirical MSEs than the FLSE. The superior 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°, and hence the estimators that address the possible endogeneity caused by estimation errors will work better. The FIVE or the F2SLSE exhibits the smallest MSE in most cases (see also Table S2 in the Supplementary Material reporting the simulation results for the case where the lower bound of a_1 and a_2 increases to 0.6). However, it is hard to conclude the relative performance between the 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 these IV estimators complementarily in practice.

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

[5] $\mathcal{A}$ is approximated by $\sum_{j=1}^{50} a_j \xi_j^c \otimes \xi_j^c$ in actual computation as in the case of ε_t.

Table 2. Simulation results for Experiment 3: empirical MSEs ($a_1, a_2 \geq 0.4$)

		Sparse design				Exponential design			
	n	100		150		100		150	
	T	200	500	200	500	200	500	200	500
Loader's	FIVE	0.206	0.158	0.187	0.152	0.400	0.227	0.315	0.194
	F2SLSE	0.204	0.157	0.186	0.151	0.392	0.219	0.310	0.189
	FLSE	0.255	0.228	0.208	0.187	0.427	0.354	0.333	0.267
Silverman's	FIVE	0.272	0.218	0.223	0.188	0.395	0.257	0.314	0.215
	F2SLSE	0.272	0.215	0.223	0.186	0.396	0.247	0.315	0.209
	FLSE	0.326	0.291	0.251	0.226	0.418	0.351	0.322	0.258

Note: Based on 1,000 replications. Each cell reports the empirical mean squared error (MSE) of the three considered estimators: FIVE, F2SLSE, and Park and Qian's (2012) FLSE. The RIVE considered in Sections 5.1 and 5.2 is excluded in this experiment since the estimator is developed for i.i.d. functional data.

5.4. Empirical Application: Effect of Immigration on Native Wages in the US

In this section, we use our estimation methods to investigate the effect of immigrant inflows on the US native labor market outcomes of workers with heterogeneous skills, which has received due attention from both researchers and policymakers, see, for example, Card (2009), Borjas, Grogger, and Hanson (2011), Ottaviano and Peri (2012), and Glitz (2012). To begin with, we use national-level data and generalize a widely used empirical model by viewing the variables of interest as functions depending on a measure of relative communication skill provision. Our measure of relative communication skill provision is similar to Peri and Sparber's (2009) measure of occupation-specific relative provision of communication versus manual skills. The number of distinct skill levels, denoted s_j, is 223, and by construction, each occupation is uniquely identified by the skill score $s_j \in [0,1]$. Its formal definition is provided in the Supplementary Material.

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)[6], 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 ($\ell^\circ_{it}(s_j)$) and that of natives ($\ell^\circ_{nt}(s_j)$) are computed by the total hours of work per week (weighted by the variable WTFINL) provided by the foreign- and

[6]Flood, King, Rodgers, Ruggles, and Warren (2020).

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

native-born workers for each s_j. The skill-dependent native wage is computed by weighted averaging the weekly wages of native workers[7] 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, β_j° is the parameter of interest, the explanatory variable $\Delta h_t^\circ(s_j)$ is the first difference of $h_t^\circ(s_j)$, and $h_t^\circ(s_j) = \ell_{it}^\circ(s_j)/(\ell_{nt}^\circ(s_j) + \ell_{it}^\circ(s_j))$. In this model, an inflow of immigrants in the occupation with s_j is assumed to affect only 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. As an alternative, we view observations $w_t^\circ(s_j)$ and $h_t^\circ(s_j)$ 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 that minimizes the least squares 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:

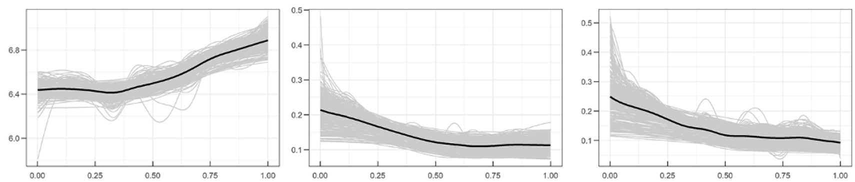
$$\Delta w_t = \mathcal{A} \Delta h_t + u_t, \tag{5.3}$$

where $\Delta w_t = w_t - w_{t-1}$, $\Delta h_t = h_t - h_{t-1}$, and Δh_t is likely to be correlated with u_t due to, for example, the self-selection bias pointed out by Borjas (1987) and Llull (2018). Thus, 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 $z_t^\circ(s_j)$, is defined as follows:

$$z_t^\circ(s_j) = \frac{\tilde{\ell}_{it}^\circ(s_j)}{\ell_{nt}(s_j) + \tilde{\ell}_{it}^\circ(s_j)} \quad \text{and} \quad \tilde{\ell}_{it}^\circ(s_j) = \frac{1}{12} \sum_{b=1}^{B} \sum_{t=1}^{12} \frac{\ell_{it1994,b}^\circ(s_j)}{\ell_{it1994,b}^\circ} \ell_{it,b}^\circ,$$

where b denotes the country of birth of immigrants, $\ell_{it1994,b}^\circ(s_j)$ 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 $\ell_{it1994,b}^\circ$ is its aggregation over s_j. The curve of imputed shares of immigrants, denoted z_t, is obtained by smoothing $z_t^\circ(s_j)$, and the instrument, denoted Δz_t, is the first difference of z_t.

[7]The weekly wage of a native worker is computed as (hourly wage) × (usual hours of work), and the variables required to compute this quantity are also available in the CPS. We use the variable EARNWT as a weight.



(a) average log native wages (b) relative share of immigrants (c) imputed share of immigrants

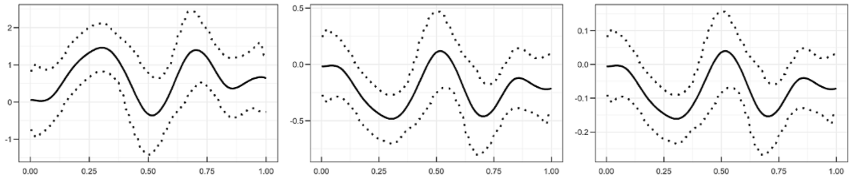
FIGURE 2. Functional data (gray) and their mean functions (black).

The smoothed curves are reported in Figure 2. 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 decrease in such occupations, and so does the imputed share of immigrants; this may be because natives have a comparative advantage in communication intensive tasks.

Then, we apply our estimation method to study if an inflow of immigrants has a heterogeneous impact on native workers depending on the value of s. For the ease of interpretation, we focus on the case where immigrants are fully concentrated in a group of occupations with low, medium or high communication skill intensity, but they are evenly distributed within the group; that is, we set ζ in Theorem 2 to $1\{0 \le s < 1/3\}$, $1\{1/3 \le s < 2/3\}$, and $1\{2/3 \le s < 1\}$. The regularization parameter is chosen as in Section 5.1. The results computed from the FIVE are summarized in Figure 3. The estimation results from the F2SLSE are similar and thus omitted. Overall, our findings in Figure 3 reconfirm the existing evidence that an inflow of immigrants heterogeneously affects the labor market outcomes of native workers according to workers' skills. For example, in Figure 3a, if the share of immigrants increases in occupations with a low value of s_j, then native wages are overall positively affected, although the size of this effect depends on the value of native workers' s_j as well. In particular, Figure 3a suggests that native workers in occupations with $s \in [0.1, 0.4]$ will experience the most significant positive wage effects. This is somewhat consistent with Peri and Sparber's (2009) finding that native workers in occupations intensive in manual skills take advantage of having better-paid jobs when similarly skilled immigrants enter into the market. On the other hand, in Figure 3b, it seems that the natives in occupations with $s \in [0.1, 0.4]$ are negatively affected if the share of immigrants increases in occupations of which $s \in [1/3, 2/3]$.

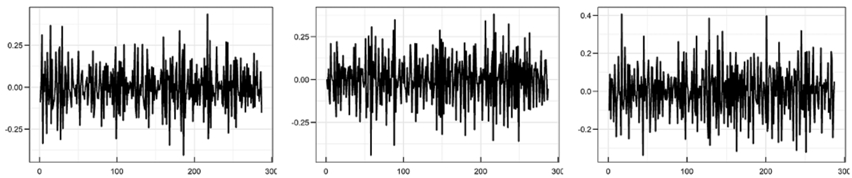
Before concluding this section, recall that the earlier literature mostly relies on the strategy of reducing the dimensionality of the model by classifying workers into a few groups according to a measure of their skill. We will make a comparison of our estimation results with those based on this strategy. To this end, we let $\zeta_{j(J)} = J \times 1\{j-1 < Js \le j\}$. Then, the inner product $\langle \Delta w_t, \zeta_{j(J)} \rangle$ computes the average of changes in the log wages of natives in occupations of which s is between $(j-1)/J$

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press



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

FIGURE 3. Estimated effects of immigration computed from the FIVE.
Note: The solid line reports $\widehat{\mathcal{A}}\zeta$. The dashed lines represent the collection of confidence intervals for the local averages of $\mathcal{A}\widehat{\Pi}_{\mathrm{K}}\zeta$ over finely defined interval $[(m-1)/M, m/M]$, where $M = 50$. For each m, the interval is constructed as in (3.5) with 95% significance level by noting that the local average is given by $\langle \mathcal{A}\widehat{\Pi}_{\mathrm{K}}\zeta, \zeta_{m(M)} \rangle$, and this interval, of course, may be viewed as the confidence interval for the local average of $\mathcal{A}\zeta$ under certain conditions (see Remark 11).



(a) $\zeta = \zeta_{1(3)}$ (b) $\zeta = \zeta_{2(3)}$ (c) $\zeta = \zeta_{3(3)}$

FIGURE 4. Group characteristics (first difference of log native wages, $\langle \Delta w_t, \zeta_{j(3)} \rangle$).

and j/J, at time t. We then estimate the following using the standard 2SLSE:

$$\Delta w_{t(J)} = \beta \Delta h_{t(J)} + u_{t(J)}, \tag{5.4}$$

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 example, if $J = 3$, we have $\langle \Delta w_t, \zeta_{1(3)} \rangle$, $\langle \Delta w_t, \zeta_{2(3)} \rangle$, and $\langle \Delta w_t, \zeta_{3(3)} \rangle$, which are plotted in Figure 4. In the following, we consider three different values for J: 3, 7, and 11.

We compare the performance of estimators by using the root mean squared prediction error (RMSPE), which is computed with a rolling window for three test sets, with setting their starting points respectively as 2013/01, 2015/01, and 2017/01. Specifically, if we let $\bar{u}_h$ be the forecasting error computed from the FIVE, F2SLSE or RIVE, then the RMSPE is given by $(H^{-1}\sum_{h=1}^{H}\int \bar{u}_h(s)^2 ds)^{1/2}$ and their regularization parameters are chosen in such a way as to minimize the RMSPE. Let $\widehat{u}_{h,j(J)}$ denote the jth element of the forecasting error $\widehat{u}_{h(J)}$ computed from the 2SLSE for each J. The standard RMSPE of the 2SLSE is given by $(H^{-1}\sum_{h=1}^{H}\sum_{j=1}^{J}\widehat{u}_{h,j(J)}^2)^{1/2}$. Because this measure is nondecreasing in J and thus does not provide a fair comparison between RMSPEs, we instead consider the normalized RMSPE, given by $((JH)^{-1}\sum_{h=1}^{H}\sum_{j=1}^{J}\widehat{u}_{h,j(J)}^2)^{1/2}$, which

https://doi.org/10.1017/S0266466624000252 Published online by Cambridge University Press

Table 3. Root mean squared prediction errors

Test period	FIVE	F2SLSE	RIVE	2SLSE		
				$J=3$	$J=7$	$J=11$
2013/01∼	0.1886	0.1889	0.1883	0.1514	0.2025	2.5877
2015/01∼	0.1828	0.1829	0.1823	0.1432	0.1884	3.4587
2017/01∼	0.1679	0.1678	0.1676	0.1236	0.1734	1.0399

Note: Each cell reports the estimated (normalized) RMSPE which is computed using three test sets.

can be reasonably compared to the RMSPEs of our estimators. This is because for each $h \geq 1$, (i) both $\widehat{u}_{h,j(J)}$ and $\langle \bar{u}_h, \zeta_{j(J)} \rangle$ are estimates of the local average of u_h over the interval $[(j-1)/J, j/J]$ and (ii) $\int \bar{u}_h(s)^2 ds$ may be approximated by $J^{-1} \sum_{j=1}^{J} \langle \bar{u}_h, \zeta_{j(J)} \rangle^2$.

Estimation results are reported in Table 3. We first note that the results from our estimators and those from the RIVE are very similar to each other. These estimators report smaller RMSPEs than those of the 2SLSE except for the case $J = 3$. Even if the 2SLSE reports the smallest RMSPEs when $J = 3$, we should note that in this case, 223 different skill levels are aggregated into only three groups, resulting in a lot of information loss. Moreover, the normalized RMSPE of the 2SLSE rapidly increases as we consider more finely defined skill groups. This may be because, as J gets larger, the number of parameters to be estimated rapidly increases. In addition, for a large J, the sample (cross-)covariance matrices for computing the 2SLSE tend to be singular, and thus the 2SLSE is expected to perform poorly. This result also suggests that the pre-classification strategy can have a significant effect on the estimation results and their interpretation. Therefore, the results given by Table 3 imply that our functional IV methodology can be an appealing alternative to practitioners.

6. CONCLUSION

This article extends the existing results on the endogenous functional linear model to a more general setup, allowing for weakly dependent errors and without assuming a specific type of endogeneity. Additionally, it suitably extends the asymptotic approach of Hall and Horowitz (2007) to encompass cases with endogeneity, a common occurrence in practical applications. Consequently, this article proposes two novel estimators and provides their detailed asymptotic properties under this broader setting. Notably, even in the case where there is no endogeneity and hence the FIVE reduces to the FLSE of Park and Qian (2012) (see Sect. 3.1), most of the asymptotic results that we obtain in the present article have not been explored in the literature, to the best of the authors' knowledge. Given the potential prevalence of endogeneity in the functional linear model (see Sect. 2.1), we believe that the theoretical results presented in this article hold value beyond the existing findings. From a practical perspective, our methodology can be applied to study relationships between economic functional variables of which availability has been being increasing; potential examples include the density-on-density regression model (see, e.g., Park and Qian, 2012).

SUPPLEMENTARY MATERIAL

Seong, D. and W.-K. Seo (2024): Supplement to "Functional instrumental variable regression with an application to estimating the impact of immigration on native wages", *Econometric Theory Supplementary Material*. To view, please visit: https://doi.org/10.1017/10.1017/S0266466624000252.

REFERENCES

Autor, D. H., & 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. (2022). High-dimensional mixed-frequency IV regression. *Journal of Business & Economic Statistics*, 40, 1470–1483.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62(3), 657–681.

Benatia, D., Carrasco, M., & 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., & 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.

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

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

Conway, J. B. (2007). *A course in functional analysis*. Springer.

Crambes, C., & 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., & Preston, I. P. (2013). The effect of immigration along the distribution of wages. *The Review of Economic Studies*, 80(1), 145–173.

Egozcue, J. J., Díaz-Barrero, J. L., & 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., & Warren, J. R. (2020). Integrated public use microdata series, current population survey: Version 8.0 [dataset]. Minneapolis, MN : IPUMS. https://doi.org/10.18128/D030.V8.0.

Florens, J.-P., & 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.

Hall, P., & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35(1), 70–91.

Hausman, J. A. (1983). Specification and estimation of simultaneous equation models. *Handbook of Econometrics*, 1, 391–448.

Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer.

Hron, K., Menafoglio, A., Templ, M., Hrůzová, K., & Filzmoser, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. *Computational Statistics & Data Analysis*, 94, 330–350.

Hu, B., & Park, J. Y. (2016). *Econometric analysis of functional dynamics in the presence of persistence*. Indiana University.

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

Jaimez, R. G., & Bonnet, M. J. V. (1987). On the Karhunen-Loeve expansion for transformed processes. *Trabajos de Estadistica*, 2, 81–90.

Kokoszka, P., Miao, H., Petersen, A., & 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.

Mas, A. (2007). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis*, 98(6), 1231–1261.

Newey, W. K., & Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, 77(3), 687–719.

Nielsen, M. Ø., Seo, W.-K., & Seong, D. (2023). Inference on the dimension of the nonstationary subspace in functional time series. *Econometric Theory*, 39(3), 443–480.

Ottaviano, G. I. P., & Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1), 152–197.

Park, J. Y., & Qian, J. (2012). Functional regression of continuous state distributions. *Journal of Econometrics*, 167(2), 397–412.

Peri, G., & Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1(3), 135–169.

Petersen, A., & 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., & Silverman, B. W. (2005). *Functional data analysis*. Springer.

Seo, W.-K. (2024). Functional principal component analysis for cointegrated functional time series. *Journal of Time Series Analysis*, 45(2), 320–330.

Seo, W.-K., & 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., & 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.

Staudenmayer, J., & 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., & 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.

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

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

Zhang, C., Kokoszka, P., & Petersen, A. (2021). Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis*, 43(1), 30–52.