Statistica Sinica Preprint No: SS-2022-0154	
Title	Testing Exogeneity in the Functional Linear Regression Model
Manuscript ID	SS-2022-0154
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0154
Complete List of Authors	Manuela Dorn, Melanie Birke and Carsten Jentsch
Corresponding Authors	Melanie Birke
E-mails	Melanie.Birke@uni-bayreuth.de

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

Testing exogeneity in the functional linear regression model

Manuela Dorn[1], Melanie Birke[1], Carsten Jentsch[2]

[1] *University of Bayreuth,* [2] *TU Dortmund University*

Abstract: We propose a novel test statistic for testing exogeneity in the functional linear regression model. In contrast to Hausman-type tests in finite dimensional linear regression setups, we show that a direct extension to the functional linear regression model is not possible. Instead, we propose a test statistic based on the sum of squared differences of projections of the two estimators used for testing the null hypothesis of exogeneity in the functional linear regression model. We derive asymptotic normality under the null and show consistency under general alternatives. Moreover, we establish bootstrap consistency results for residual-based bootstrap approaches. In simulations, we investigate the finite sample performance of the proposed exogeneity tests and illustrate the superiority of bootstrap-based approaches. In particular, the bootstrap-based results turn out to be much more robust with respect to the choice of the regularization parameter.

Key words and phrases: Asymptotic theory, bootstrap inference, endogeneity, Hausman test, instrumental variables, inverse problem.

1. Introduction

The construction of goodness-of-fit tests in functional regression models is much more complicated than e.g. in the multiple linear setting. This is particularly due to the fact that, in functional linear regression models, the L_2-distance of the slope function estimator to the true function has no proper limiting distribution. Under exogeneity, this was shown in Cardot et al. (2007) and Ruymgaart et al. (2011) for two slope function estimators in the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

classical functional linear regression model. It turns out that this lack of a proper limiting distribution also remains for other such estimators based on different model assumptions such as endogeneity. This phenomenon inherent to (infinite-dimensional) functional data setups is probably the main reasons why testing in general and goodness-of-fit testing in particular is less developed for functional regression models. In particular, desirable and seemingly natural counterparts of standard and well-established tests from the (finite-dimensional) multiple linear regression model are still missing in functional linear regression setups.

In functional data settings, existing goodness-of-fit tests are described in Müller and Stadtmüller (2005), who use a suitable scalar product to transform the functions to a different space using the autocovariance operator to obtain a test statistic having a proper limiting distribution. Further approaches are given in Cuesta-Albertos et al. (2019), García-Portugués et al. (2014) and García-Portugués et al. (2020), who use random projections together with empirical process techniques.

In practice, a crucial assumptions that is usually imposed on the regression model to guarantee unbiased estimation, is the exogeneity of the regressor. In particular, in empirical applications from economics, this assumption is often violated, because (some) regressors are correlated with the error terms leading to endogeneity issues. While the estimation in endogeneous functional regression models is an inverse problem, ignoring potential endogeneity generally results in biased and inconsistent estimators. However, using an estimator that is robust to endogeneity at all times is not desirable, because such estimators are usually less efficient and, although consistent, in finite samples the quality of the estimator strongly depends on the strength and the functional form of the instrument (see e.g. Reiss (2016)) even under exogeneity. We present a small simulation example demonstrating this in Section 5.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

Hence, confronted with potential endogeneity issues in the (functional) data, it is important to test for exogeneity first. If the null of exogeneity is rejected, other estimators that account for endogeneity have to be used, that e.g. rely on instrumental variables (IVs), to achieve consistent estimation. In functional regression setups, IV estimators have been considered by Johannes (2016) or Florens and Van Bellegem (2015), who derive asymptotic theory as minimax rates and consistency respectively asymptotic normality of the prediction error.

In the multiple linear regression model, the Hausman test (see Hausman (1978) and Wu (1974)) is based on the norm of the difference of two parameter estimators, where one estimator is consistent under both exogeneity and endogeneity, while the other is only consistent under exogeneity, but biased and inconsistent under endogeneity. While this original Hausman test is a standard and natural approach for testing exogeneity in the multiple linear regression model, it is *not possible* to transfer this testing approach directly to the functional linear model. This is because a natural extension to the functional regression setup would rely on the L_2-distance of two different slope function estimators, which suffers from a lack of a proper limiting distribution according to Proposition 3.1 below, which transfers the findings obtained by Cardot et al. (2007) and Ruymgaart et al. (2011) to the present setting. Making use of the fact, that, in contrast to the L_2-distance, the projection error typically has an asymptotic distribution (see e.g. Müller and Stadtmüller, 2005; Florens and Van Bellegem, 2015), we propose to construct such a Hausman-type test based on the sum of the squared differences of projections of the estimators in Johannes (2016) and Johannes (2008) as test statistic.

The rest of the paper is organized as follows. In Section 2, we state the model assumptions and discuss the estimators we use to construct the test statistic. In Section 3,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

we show that a direct extension of the Hausman test to the functional linear regression model is not possible. Then, we define a modified Hausman-type test statistic and derive its asymptotic distribution. At the end of this section, we also discuss possible extensions to other related estimators and observation schemes. As the limiting distribution turns out to depend on unknown functional nuisance parameters, which are difficult to estimate, we propose residual-based bootstrap methods in Section 4 and prove their consistency. The finite sample performance of all tests discussed in Sections 3 and 4 is investigated in Section 5, while the practical relevance is demonstrated with a real data example in Section 6. All proofs are deferred to the Appendix. Additional auxiliary results and additional simulation results can be found in supplementary material.

2. Model setup and slope function estimation

We consider the functional linear regression model

$$Y = \int_{[0,1]} X(t)\beta(t)dt + U = \langle \beta, X \rangle + U, \tag{2.1}$$

where Y is a real-valued random variable, U is a real-valued error term with $E(U) = 0$ and $E(U^2) = \sigma^2 \in (0, \infty)$, X is a functional random variable with values in $L_2([0,1])$ such that $\int_0^1 E|X(t)|^2\,dt < \infty$. In this setup, the error variance σ^2 is unknown, and β is an unknown slope function from the Sobolev space of periodically extendable square integrable functions denoted by $\mathcal{W}_\nu = \Big\{ f \in L^2[0,1] : \;\; \|f\|_\nu^2 = \sum_{k\in\mathbb{Z}} \gamma_k^\nu |\langle f, \phi_k \rangle|^2 < \infty \Big\}$, where $(\phi_k)_{k\in\mathbb{Z}}$ is the Fourier basis of $L^2([0,1])$, $\nu \in \mathbb{R}$ and $\gamma_k = 1 + |2\pi k|$, $k \in \mathbb{Z}$, see e.g. Neubauer (1988), Mair and Ruymgaart (1996) or Tsybakov (2004). In the setup of (2.1), we will speak of *exogeneity* (and call X an exogenous regressor), if

$$H_0 : \;\; E\{X(t)U\} = 0 \text{ for all } t \in [0,1]. \tag{2.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

Otherwise, we will speak of *endogeneity* (and call X an endogenous regressor), if

$$H_1: \ E\{X(t)U\} \neq 0 \text{ for at least one } t \in [0,1]. \tag{2.3}$$

2.1 Estimation of β under endogeneity

We begin with the estimation setup for β under endogeneity. For this purpose, we assume to additionally have a functional instrumental variable W with values in $L_2([0,1])$ such that $\int_0^1 E|W(t)|^2\,dt < \infty$ and $E\{UW(t)\} = 0$ for all $t \in [0,1]$. While, for the sake of simplicity, it is often assumed in the literature to have that $E\{X(t)\} = E\{W(t)\} = 0$ holds for all $t \in [0,1]$, this assumption turns out to be not restrictive and the general case that allows for $E\{X(t)\} \neq 0$ and $E\{W(t)\} \neq 0$ can be handled along the same lines by first centering with the sample mean. Nevertheless, in the following, we state all results for the general case. For the estimation of the cross-covariance operator, we also assume that (X, W) is second-order stationary, see Johannes (2016).

Assumption 1. *There exist functions $c_X, c_W, c_{WX} : [-1,1] \to \mathbb{R}$, such that $Cov\{X(s), X(t)\} = c_X(t-s)$, $Cov\{W(s), W(t)\} = c_W(t-s)$ and $Cov\{W(s), X(t)\} = c_{WX}(t-s)$ for all $s, t \in [0,1]$, respectively, where c_X is assumed to be continuous.*

Under the alternative (2.3), the imposed continuity of c_X immediately implies $E\{X(t)U\} \neq 0$ on some set with positive Lebesgue measure. This condition ensures, that the test statistic proposed in the following can be used to consistently test for the null hypothesis H_0 in (2.2) against general alternatives H_1 in (2.3). Note that c_X and c_W define the kernels of the covariance operators Γ_X of X and Γ_W of W, respectively, and c_{WX} is the kernel of the cross covariance operator Γ_{WX} of X and W. The (joint) weak stationarity of (X, W) ensures, that both covariance operators as well as the cross covariance operator have the same exponen-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

tial system of eigenfunctions, which we denote by $(\phi_k)_{k\in\mathbb{N}}$. While this assumption of (joint) second order stationarity seems to be quite restrictive, it is a common assumption when deriving consistency and especially the convergence rates of the slope function estimators, see Johannes (2008), Johannes (2009), Johannes (2016) or Seong and Seo (2021). On the other hand, it is well known, that the scalar product in Hilbert spaces is independent of the chosen basis. Therefore, we hypothesize that the results derived below under the assumption of second order stationarity remain also valid without this assumption, with the price to pay that completely different arguments are required in the proofs, see e.g. Müller and Stadtmüller (2005). Our hypothesis is backed by the additional simulations provided in the supplementary material. Hence, in the following, let $(x_k, \phi_k)_{k\in\mathbb{N}}$ be the eigensystem of Γ_X, $(w_k, \phi_k)_{k\in\mathbb{N}}$ the eigensystem of Γ_W, and $(c_k, \phi_k)_{k\in\mathbb{N}}$ the eigensystem of Γ_{WX}. Furthermore, let $\lambda_k = \frac{|c_k|^2}{w_k}$ which can be bounded by $\lambda_k = \frac{|c_k|^2}{w_k} \le x_k$ using the Cauchy-Schwarz inequality, and additionally are assumed to fulfill the following regularity conditions.

Assumption 2. *Throughout the paper, we assume that all eigenvalues* $(x_k)_{k\in\mathbb{Z}}$ *are strictly positive and that*

$$\sum_{k\in\mathbb{Z}} \frac{|E(Y\langle X, \phi_k\rangle)|^2}{x_k^2} < \infty.$$

Furthermore, let $\mu_X = \sum_{k\in\mathbb{Z}} \langle \mu_X, \phi_k\rangle \phi_k$ *and* $\mu_W = \sum_{k\in\mathbb{Z}} \langle \mu_W, \phi_k\rangle \phi_k$ *denote the expectations of* X *and* W*, respectively, and assume that there exists some* $0 < \tau < \infty$ *such that*

$$\sup_{k\in\mathbb{Z}} \left|\frac{\lambda_k}{w_k}\right| = \sup_{k\in\mathbb{Z}} \frac{|c_k|^2}{w_k^2} \le \tau. \tag{2.4}$$

The last assumption ensures, that the linear prediction of X with respect to the instrumental variable W is well defined. In principle, if it were available, IV estimation would be based on the *optimal* instrument $\widetilde{W} = \Gamma_{WX}\Gamma_W^{-1} W = \sum_{k\in\mathbb{Z}} \frac{\overline{c_k}}{w_k} \langle W, \phi_k\rangle \phi_k$ and the eigenvalues $(\widetilde{\lambda}_k)_{k\in\mathbb{N}}$ of the corresponding cross covariance operator $\Gamma_{\widetilde{W}X}$. However, this is usually not

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

the case and the optimal instrument $\widetilde{W}$ respectively the corresponding eigenvalues $(\widetilde{\lambda}_k)_{k\in\mathbb{N}}$ of the cross covariance operator have to be estimated. Note that $\widetilde{W}$ could be exactly computed from X and W, if the (cross) covariance operators were known.

In the following, let $\mathcal{S}_n = \{(X_i, W_i, Y_i)\}_{i=1,\dots,n}$ be independent and identically distributed (i.i.d.) copies of (X, W, Y) and suppose that the alternative H_1 in (2.3) holds. Then, due to Johannes (2016), the unknown slope function β can still be consistently estimated. For this purpose, let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence of regularization parameters such that $\alpha_n > 0$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \alpha_n = 0$. To simplify notation, we will write α for the regularization keeping in mind that it still depends on n. Since the covariance operators and therefore the corresponding eigenvalues are unknown, they have to be estimated in a first step. Further, let $\Gamma_{WX,n}, \Gamma_{X,n}, \Gamma_{W,n} : L_2([0,1]) \to L_2([0,1])$ denote the empirical versions of Γ_{WX}, Γ_X and Γ_W, respectively, defined by

$$\Gamma_{WX,n} f = \frac{1}{n}\sum_{i=1}^{n} \langle W_i, f\rangle X_i, \quad \Gamma_{X,n} f = \frac{1}{n}\sum_{i=1}^{n} \langle X_i, f\rangle X_i, \quad \text{and} \quad \Gamma_{W,n} f = \frac{1}{n}\sum_{i=1}^{n} \langle W_i, f\rangle W_i$$

for $f \in L_2([0,1])$. These estimators as well as the deduced estimators

$$\widehat{w}_k = \frac{1}{n}\sum_{i=1}^{n} |\langle W_i, \phi_k\rangle|^2, \qquad \widehat{x}_k = \frac{1}{n}\sum_{i=1}^{n} |\langle X_i, \phi_k\rangle|^2,$$
$$\widehat{c}_k = \frac{1}{n}\sum_{i=1}^{n} \langle \phi_k, X_i\rangle\langle W_i, \phi_k\rangle, \qquad \widehat{\lambda}_k = \frac{|\widehat{c}_k|^2}{\widehat{w}_k} I\{\widehat{w}_k \geq \alpha\}$$

for the eigenvalues w_k, x_k, c_k and λ_k, respectively, are consistent for all $k \in \mathbb{Z}$. Hence, observations of the optimal linear instrument $\widetilde{W}$ can be estimated by

$$\widetilde{W}_{n,i} = \sum_{k\in\mathbb{Z}} \frac{\overline{\widehat{c}_k}}{\widehat{w}_k} I\{\widehat{w}_k \geq \alpha\}\langle W_i, \phi_k\rangle\phi_k, \quad i = 1, \dots, n,$$

and the corresponding cross covariance operator by

$$\widetilde{\Gamma}_n = \frac{1}{n}\sum_{i=1}^{n} \langle \widetilde{W}_{n,i}, \cdot\rangle X_i = \frac{1}{n}\sum_{k\in\mathbb{Z}} \frac{\overline{\widehat{c}_k}}{\widehat{w}_k} I\{\widehat{w}_k \geq \alpha\} \sum_{i=1}^{n} \langle \cdot, X_i\rangle\langle W_i, \phi_k\rangle\phi_k. \tag{2.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

This allows to construct the IV-based estimator $\widehat{\beta}_{IV}$ of the slope function β defined by

$$\widehat{\beta}_{IV} = \sum_{k\in\mathbb{Z}} \frac{\widehat{g}_k}{\widehat{\lambda}_k} I\{\widehat{\lambda}_k \geq \gamma_k^\nu \alpha\}\phi_k = \sum_{k\in\mathbb{Z}} \frac{\frac{1}{n}\sum_{i=1}^n \langle W_i, \phi_k\rangle Y_i}{\widehat{c}_k} I\{\widehat{\lambda}_k \geq \gamma_k^\nu \alpha\} I\{\widehat{w}_k \geq \alpha\}\phi_k, \qquad (2.6)$$

where $\widehat{g}_k = \frac{1}{n}\sum_{i=1}^n Y_i \langle \widetilde{W}_{n,i}, \phi_k\rangle$.

2.2 Estimation of β under exogeneity

As shown in Johannes (2016), under Assumptions 1 and 2, the estimator $\widehat{\beta}_{IV}$ is consistent under exogeneity in (2.2) *and* under endogeneity in (2.3). In contrast, again under Assumptions 1 and 2, the estimator

$$\widehat{\beta} = \sum_{k\in\mathbb{Z}} \frac{\frac{1}{n}\sum_{i=1}^n \langle X_i, \phi_k\rangle Y_i}{\widehat{x}_k} I\{\widehat{\lambda}_k \geq \alpha\gamma_k^\nu\}\phi_k \qquad (2.7)$$

is only consistent under the exogeneity assumption (2.2) (see Johannes (2008)) and inconsistent under endogeneity in (2.3). Note that in comparison to the original definition of $\widehat{\beta}$ in Johannes (2008), we use the same indicator function $I\{\widehat{\lambda}_k \geq \alpha\gamma_k^\nu\}$ as in $\widehat{\beta}_{IV}$. It turned out, that the tests proposed below tend to perform better if the same regularization is used in both estimators $\widehat{\beta}_{IV}$ and $\widehat{\beta}$, although it might not be the best choice for estimating β by $\widehat{\beta}$ under assumption (2.2).

3. Construction of the test statistic and asymptotic theory

A direct generaliziation of the Hausman test from the multiple linear to the function linear regression model is not possible. This is an immediate consequence of the following negative result, which can be obtained along the same lines as the ones in Cardot et al. (2007) or Ruymgaart et al. (2011), see also Dorn (2021) for further details.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

Proposition 3.1. *In the functional linear regression model* (2.1)*, under both hypotheses* (2.2) *and* (2.3)*, there exists no random variable Z with non-degenerate distribution, such that for the estimators $\widehat{\beta}_{IV}$ and $\widehat{\beta}$ defined in* (2.6) *and* (2.7)*, $s_n\|\widehat{\beta}_{IV}-\widehat{\beta}\| \xrightarrow{\mathcal{D}} Z$ for some real sequence $(s_n)_{n\in\mathbb{N}}$ with $lim_{n\to\infty}s_n=\infty$, where $||\cdot||$ denotes the norm of the Hilbert space.*

We use this result as motivation for a different approach to construct a Hausman-type test for exogeneity in the functional linear regression setup in the following. Based on the two estimators (2.6) and (2.7), since it is known that the prediction error has an asymptotic distribution, we construct the test statistic as

$$T_n = \frac{1}{n}\sum_{i=1}^{n}\left|\left\langle \widehat{\beta}_{IV}-\widehat{\beta}, X_i\right\rangle\right|^2 = \left\langle \widehat{\beta}_{IV}-\widehat{\beta}, \Gamma_{X,n}\left(\widehat{\beta}_{IV}-\widehat{\beta}\right)\right\rangle. \tag{3.1}$$

The last representation above corresponds to the idea used in Müller and Stadtmüller (2005) to construct a goodness-of-fit test. The equivalence of both approaches can be seen by using the singular value decomposition for the estimators and for the covariance operator.

Assumption 3. *For the sequence of regularization parameters, we assume*

$$\alpha_n = \alpha > 0\,\forall\, n\in\mathbb{N},\ \ \alpha = o(1)\ \ \textit{and}\ \ \frac{1}{n\alpha^2} = o(1).$$

For the next results, different moment conditions for X, W and U are required. To simplify the notation, we introduce the following sets. In doing so, we assume, that all conditions on X and W mentioned above are fulfilled and define

$$\mathcal{F}_\eta^m = \left\{(X,W)\middle|\sup_{k\in\mathbb{Z}}\mathrm{E}\left|\frac{\langle X,\phi_k\rangle}{\sqrt{x_k}}\right|^m \le \eta \text{ and } \sup_{k\in\mathbb{Z}}\mathrm{E}\left|\frac{\langle W,\phi_k\rangle}{\sqrt{w_k}}\right|^m \le \eta\right\}, \tag{3.2}$$

$$\mathcal{G}_\eta^m = \left\{X\middle|\Gamma_X > 0 \text{ and } \sup_{k\in\mathbb{Z}}\mathrm{E}\left|\frac{\langle X,\phi_k\rangle}{\sqrt{x_k}}\right|^m \le \eta\right\}. \tag{3.3}$$

In the following, for an operator Δ and its sequence of eigenvalues $(\delta_k)_{k\in\mathbb{Z}}$, let $\Delta^\dagger$ denote its

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

regularized inverse, that is $\Delta^\dagger = \sum_{k\in\mathbb{Z}} \frac{1}{\delta_k} I\{|\delta_k| > \alpha\gamma_k^\nu\}\langle\cdot,\phi_k\rangle\phi_k$ and we define

$$t_n^2 = \|(\widetilde{\Gamma}^\dagger_{X,n} - \Gamma^\dagger_X)\Gamma_X\|^2_{HS} = \sum_{k\in\mathcal{K}_n} \left(\frac{x_k w_k}{|c_k|^2} - 1\right)^2, \tag{3.4}$$

where $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt norm and $\mathcal{K}_n = \{k \in \mathbb{Z} \mid \lambda_k \geq \alpha\gamma_k^\nu\}$. Note, that $\frac{n}{t_n} > Cn\alpha^2$ for some $C > 0$. Now, we are in a position to state an asymptotic result for the test statistic. The proofs of the next results are deferred to the Appendix 7.

Theorem 3.2. *In model (2.1), under Assumptions 1-3, let* $\mathcal{S}_n = \{(X_i, W_i, Y_i)\}_{i=1,\dots,n}$ *be i.i.d. copies of* (X, W, Y) *with* $(X, W) \in \mathcal{F}_\eta^{128}$ *and* $E|U|^{128} \leq \eta < \infty$*. Furthermore, let* $t_n \to \infty$ *as* $n \to \infty$*, and*

$$\frac{1}{t_n^4}\sum_{k\in\mathcal{K}_n}\left(\frac{x_k w_k}{|c_k|^2} - 1\right)^4 = o(1), \quad \sum_{k\in\mathbb{Z}} |\langle\beta,\phi_k\rangle| \frac{x_k^{3/2} w_k}{|c_k|^2} < \infty, \quad \sum_{k\in\mathbb{Z}} \frac{x_k^2 w_k}{|c_k|^2} < \infty.$$

Then, under H_0 *in (2.2), we have* $\frac{n}{t_n}(T_n - \mathfrak{B}_n - \mathfrak{R}_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathfrak{V})$*, where*

$$\mathfrak{B}_n = \frac{n}{2t_n}\langle\beta,\mu_X\rangle^2 \sum_{k\in\mathbb{Z}}\left(\frac{\langle\mu_W,\phi_k\rangle}{c_k} - \frac{\langle\mu_X,\phi_k\rangle}{x_k}\right)^2 x_k I\{\lambda_k \geq \alpha\gamma_k^\nu\},$$

$$\mathfrak{R}_n = \frac{1}{n}\left(\sigma^2 + \sum_{m\in\mathbb{Z}}|\langle\beta,\phi_m\rangle|^2 x_m\right)\sum_{k\in\mathbb{Z}}\left(\frac{x_k w_k}{|c_k|^2} - 1\right) I\{\lambda_k \geq \alpha\gamma_k^\nu\},$$

$$\mathfrak{V} = \left(\sigma^2 + \sum_{m\in\mathbb{Z}}|\langle\beta,\phi_m\rangle|^2 x_m\right)^2.$$

Further, if X *is centered, that is,* $E\{X(t)\} = 0$ *for all* $t \in [0,1]$*, we have* $\mu_X = 0$ *and* $\mathfrak{B}_n = 0$*.*

To make use of the asymptotic normality result in Theorem 3.2 for testing purposes, the bias terms $\mathfrak{B}_n$ and $\mathfrak{R}_n$ and the variance term $\mathfrak{V}_n$ have to be estimated to implement the test. To this end, note that σ^2 can be consistently estimated by $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \langle\widehat{\beta}_{IV}, X_i\rangle)^2$ due to the law of large numbers and since $\frac{1}{n}\sum_{i=1}^n\langle\beta - \widehat{\beta}_{IV}, X_i\rangle^2 = o_P(1)$ by similar calculations as in the derivation of the asymptotic distribution of T_n.

Corollary 3.3. *Suppose all assumptions of Theorem 3.2 hold. Then, under H_0, we have*

$$\frac{n}{\widehat{t}_n}\frac{T_n - \widehat{\mathfrak{B}}_n - \widehat{\mathfrak{R}}_n}{\sqrt{\widehat{\mathfrak{V}}_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where, using $\widehat{\sigma}_n^2$ defined in (**??**),

$$\widehat{t}_n^2 = \sum_{k\in\mathbb{Z}}\left(\frac{\widehat{x}_k\widehat{w}_k}{|\widehat{c}_k|^2} - 1\right)^2 I\{\widehat{\lambda}_k \geq \alpha\gamma_k^\nu\},\ \widehat{\mathfrak{V}}_n = \left(\widehat{\sigma}_n^2 + \|\Gamma_{X,n}^{1/2}\widehat{\beta}_{IV}\|^2\right)^2.$$

$$\widehat{\mathfrak{B}}_n = \frac{n}{2\widehat{t}_n}\langle\widehat{\beta}_{IV}, \widehat{\mu}_X\rangle^2 \sum_{k\in\mathbb{Z}}\left(\frac{\langle\widehat{\mu}_W, \phi_k\rangle}{c_k} - \frac{\langle\widehat{\mu}_X, \phi_k\rangle}{x_k}\right)^2 \widehat{x}_k I\{\widehat{\lambda}_k \geq \alpha\gamma_k^\nu\},$$

$$\widehat{\mathfrak{R}}_n = \frac{1}{n}\left(\widehat{\sigma}_n^2 + \|\Gamma_{X,n}^{1/2}\widehat{\beta}_{IV}\|^2\right)\sum_{k\in\mathbb{Z}}\left(\frac{\widehat{x}_k\widehat{w}_k}{|\widehat{c}_k|^2} - 1\right) I\{\widehat{\lambda}_k \geq \alpha\gamma_k^\nu\}.$$

Using Corollary 3.3, we can construct a (one-sided) test for the null hypothesis H_0 in (2.2) against the alternative H_1 in (2.3). That is, for given size $\gamma \in (0,1)$, we reject H_0 if

$$\frac{n}{\widehat{t}_n}\frac{T_n - \widehat{\mathfrak{B}}_n - \widehat{\mathfrak{R}}_n}{\sqrt{\widehat{\mathfrak{V}}_n}} > u_{1-\gamma}, \tag{3.5}$$

where $u_{1-\gamma}$ denotes the $(1-\gamma)$-quantile of the standard normal distribution. In the special case of $\mu_X = 0$, we can neglect the additional bias term $\mathfrak{B}_n$, which avoids the use of its plug-in estimator $\widehat{\mathfrak{B}}_n$ such that the test simplifies and we reject H_0 if $\frac{n}{\widehat{t}_n}(T_n - \widehat{\mathfrak{R}}_n)/\sqrt{\widehat{\mathfrak{V}}_n} > u_{1-\gamma}$.

Theorem 3.4. *Suppose all assumptions of Theorem 3.2 and Corollary 3.3 hold. Then, under the alternative H_1, the test constructed in* (3.5) *is consistent.*

In practice, we do not know, whether $\mu_X = 0$ holds such that a naive application of the asymptotic test without estimating $\mathfrak{B}_n$ could result in wrong decisions. In addition, due to the additional estimation step, asymptotic tests based on plug-in methods as above usually exhibit a smaller power compared to other methods. Hence, as discussed in Section 4, the bootstrap version of the test is expected to have better finite sample behavior, since it is not required to estimate the unknown bias and variance. Moreover, we do not need to distinguish between the cases $\mu_X = 0$ and $\mu_X \neq 0$ which is a clear advantage of the bootstrap test.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

Remark 3.5. *The above results can be generalized to other types of estimators as long as a classical as well as an IV-type estimator of the same kind is available. Besides the spectral-cut-off estimators as proposed in Johannes (2008, 2016), there is little literature where this is the case. A quite general approach for an estimator under exogeneity is given in Cardot et al. (2007) using a sequence of regularization functions* $f_n : [c_n, \infty) \to \mathbb{R}_0^+$ *such that* f_n *is decreasing on* $[c_n, 2z_1 - z_2]$, *where* $(z_j)_{j\in\mathbb{Z}}$ *are the eigenvalues of the relevant covariance operator and* $(c_n)_{n\in\mathbb{N}}$ *is a decreasing sequence of positive values with* $c_n < z_1$. *Furthermore, it is required that* $\lim_{n\to\infty} \sup_{z\geq c_n} |zf_n(z) - 1| = o(1/\sqrt{n})$ *and* f_n *is differentiable on* $[c_n, \infty)$. *This also covers cut-off estimators as discussed in Müller and Stadtmüller (2005) or estimators based on Tikhonov or ridge-type regularization. It is straightforward to modify this approach for IV estimators in the endogeneous functional linear model and to choose a similar regularization scheme such that* $\widetilde{\beta} = \sum_{k\in\mathbb{Z}} \widehat{g}_k f_{1,n}(\widehat{x}_k, \widehat{\lambda}_k)$ *and* $\widetilde{\beta}_{IV} = \sum_{k\in\mathbb{Z}} \widehat{g}_k f_{2,n}(\widehat{x}_k, \widehat{\lambda}_k)$ *with* $f_{n,1}(x_k, \lambda_k) = g_1(x_k, \lambda_k)\widetilde{f}_n(\lambda_k)$ *and* $f_{n,2}(x_k, \lambda_k) = g_2(x_k, \lambda_k)\widetilde{f}_n(\lambda_k)$ *for suitable functions* g_1, g_2 *and* $\widetilde{f}_n$. *For the sake of shorter notation, we assume here* $\mu_X \equiv 0$. *Hence, in concordance with Theorem 3.2, we need as regularity assumptions*

$$\frac{\sum_{k\in\mathbb{Z}} m_k^2}{\left(\sum_{k\in\mathbb{Z}} m_k^2\right)^2} = o(1), \quad \sum_{k\in\mathbb{Z}} |\langle \beta, \phi_k\rangle| E\{|f_{1,n}(\widehat{z}_k) - f_{2,n}(\widehat{z}_k)|\} x_k < \infty,$$
$$\text{and} \quad \sum_{k\in\mathbb{Z}} x_k E\{f_{1,n}(\widehat{z}_k) - f_{2,n}(\widehat{z}_k)\}^2 < \infty$$

with $m_k = \{\lambda_k g_2^2(x_k, \lambda_k) - 2\lambda_k g_1(x_k, \lambda_k) g_2(x_k, \lambda_k) + x_k g_1^2(x_k, \lambda_k)\}^2 \widetilde{f}_n^4(\lambda_k)$, $z_k = (x_k, \lambda_k)$, $\widehat{z}_k = (\widehat{x}_k, \widehat{\lambda}_k)$. *If, additionally, Assumption 1, the first part of Assumption 2 and the moment conditions in Theorem 3.2 hold, we also have*

$$\frac{n}{\widetilde{t}_n}\left(\frac{1}{n}\sum_{i=1}^n \left|\left\langle \widetilde{\beta}_{IV} - \widetilde{\beta}, X_i\right\rangle\right|^2 - \mathfrak{R}_n\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathfrak{V}),$$

with $\mathfrak{V}$ as in Theorem 3.2 and

$$\widetilde{t}_n = \sum_{k\in\mathbb{Z}} \{\lambda_k g_2^2(x_k, \lambda_k) - 2\lambda_k g_1(x_k, \lambda_k) g_2(x_k, \lambda_k) + x_k g_1^2(x_k, \lambda_k)\}^2 \widetilde{f}_n^4(\lambda_k),$$

$$\mathfrak{R}_n = \frac{1}{n}\mathfrak{V}^{1/2} \sum_{k\in\mathbb{Z}} \{\lambda_k g_2^2(x_k, \lambda_k) - 2\lambda_k g_1(x_k, \lambda_k) g_2(x_k, \lambda_k) + x_k g_1^2(x_k, \lambda_k)\} \widetilde{f}_n^2(\lambda_k).$$

Again, it is straightforward to derive empirical versions and even bootstrap results as proposed in Section 4. Furthermore, with similar approaches, one could derive goodness-of-fit tests based e.g. on the estimator in Cardot et al. (2007). The centered test statistic converges to a normally distributed random variable with rate $\frac{n}{s_n}$, where $s_n = \sum_{k\in\mathbb{Z}} x_k^4 f_n^4(x_k)$.

Remark 3.6. *Another topic of interest is to consider different observation schemes. In practice, the functional variable would be observed in a perhaps highly frequent, but time-discrete way with additional observation errors in each time point. Such observation schemes have widely been discussed in the context of estimating the mean and covariance kernel and only recently for estimating in the functional linear model, see e.g. Zhou et al. (2022) or Hörmann and Jammoul (2023). Zhou et al. (2022) propose an estimator which is consistent in the exogeneous functional linear model by first estimating the covariance kernel of the functional variable using kernel smoothing. The slope function is then estimated by least squares based on its series expansion with respect to the principle components of the estimated covariance operator. As discussed earlier, the idea proposed in this article for testing for exogeneity could be used as long as two estimators are available, where the first is consistent under exogeneity as well as endogeneity and the second is only consistent under exogeneity. We presume that if such an estimator would be constructed and has a similar structure as e.g. the one in Zhou et al. (2022), it should be possible to derive an asymptotic distribution of the test statistic and again to construct a test for exogeneity with a similar behavior as*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

the one discussed in this article.

Hörmann and Jammoul (2023) show, that if the functional variable is discretely observed and one is only interested in prediction, but not the slope function itself, it is still efficient to work in the corresponding factor model. Since the test statistic is based on predictions, it could also be an alternative to use the classical methods for testing for exogeneity. But this would run into problems, if the number of observation points of the functional variable is large compared to the sample size.

Remark 3.7. *Finally, as in the multiple linear regression model, one could think of a multiple functional linear model as in Chiou et al. (2016) and the situation that not all, but only some functional regressors are endogeneous. By reorganizing the model, one can think of one functional variable* $\widetilde{X}$ *observed on a larger interval instead of several functions observed on* $[0,1]$*. This means, that there are some intervals where* $E\{\widetilde{X}(t)U\} \neq 0$ *for some* t *in those intervals. Since our approach can easily be generalized to other compact intervals than* $[0,1]$*, this situation is already covered by our method.*

4. Bootstrap inference

In this section, we propose (fixed-design) residual-based bootstrap procedures to estimate the distribution of $\frac{n}{t_n}(T_n - \mathfrak{B}_n - \mathfrak{R}_n)$ under the null hypothesis H_0 of exogeneity in (2.2). To this end, we first compute the residuals $\widehat{U}_i$, $i = 1, \ldots, n$, from the data by defining $\widehat{U}_i = Y_i - \langle \widehat{\beta}_{IV}, X_i \rangle$. For computing the residuals, we use the IV-based estimator $\widehat{\beta}_{IV}$ (and not $\widehat{\beta}$), because it is consistent under both the null H_0 and under the alternative H_1. Nevertheless, as the bootstrap errors U_i^*, $i = 1, \ldots, n$ will be drawn independently from the residuals, using the classical estimator $\widehat{\beta}$ would also result in a proper bootstrap scheme to approximate the distribution of the test statistic under the null, since the uncorrelatedness of error and regressor in the bootstrap sample is guaranteed by the (fixed-design) bootstrap procedure itself. However, under the null, the construction of residuals $\widehat{U}_i$ that approximate the true errors U_i, usually results in better finite sample performance of the bootstrap approximation of distribution of the test statistic under the null. In the following, different versions of residual-based bootstraps are considered that all will follow these steps:

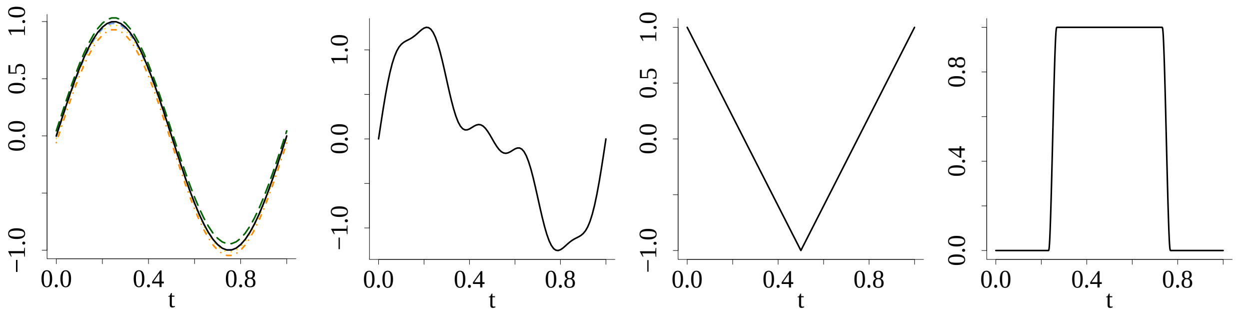
Step 1.) Given i.i.d. copies $\mathcal{S}_n = \{(X_i, W_i, Y_i)\}_{i=1,\ldots,n}$ of (X, W, Y), we generate a bootstrap sample (X_i, W_i, Y_i^*), $i = 1, \ldots, n$, by computing $Y_i^* = \langle \widehat{\beta}_{IV}, X_i \rangle + U_i^*$, where the bootstrap errors U_i^* are generated from the residuals $\widehat{U}_1, \ldots, \widehat{U}_n$ in such a way that, conditional on the data $\mathcal{S}_n$, the uncorrelatedness of U_i^* and (X_i, W_i) is assured.

Step 2.) According to (3.1), a bootstrap test statistic T_n^* is calculated from (X_i, W_i, Y_i^*), $i = 1, \ldots, n$.

Step 3.) Repeat Steps 1.) and 2.) B times, where B is large, to get bootstrap realizations $T_n^{*,1}, \ldots, T_n^{*,B}$ of the test statistic and denote by $q_{1-\gamma}^* = T_n^{*,(\lfloor B(1-\gamma) \rfloor)}$ the corresponding empirical $(1-\gamma)$-quantile.

As the bootstrap errors are generated such that uncorrelatedness of U_i^* and (X_i, W_i) conditionally on the original sample is ensured, the bootstrap automatically adopts the exogeneity assumption. For the naive (Efron-type) residual-based bootstrap, this is trivially the case, because the bootstrap errors are drawn independently with replacement from the residuals, and for the wild bootstrap, since suitable bootstrap multiplier variables V_i will also be drawn independently from X_i and W_i.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154



(a) $\widehat{\beta}$ and $\widehat{\beta}_{IV}$ (b) Slope functions used in the tests

Figure 1: Estimators and slope functions used in the test.

Theorem 4.1. *Under the assumptions of Theorem 3.2, let $\mathcal{S}_n$ be a set of i.i.d. copies of (X, W, Y), where $(X, W) \in \mathcal{F}_\eta^{128}$ and $E|U|^{128} \leq \eta < \infty$ and let $(t_n)_{n\in\mathbb{N}}$ from (3.4) be such that $t_n \to \infty$ as $n \to \infty$. Additionally, suppose that*

$$\frac{1}{t_n^4}\sum_{k\in\mathcal{K}_n}\left(\frac{x_k w_k}{|c_k|^2}-1\right)^4 = o(1), \quad \sum_{k\in\mathcal{K}_n}\left(x_k^2 \mathrm{E}|\langle \beta - \widehat{\beta}_{IV}, \phi_k\rangle|^4\right)^{1/4}\frac{x_k^4 w_k^4}{|c_k|^8} = O(1),$$
$$\sum_{k\in\mathbb{Z}}\frac{x_k w_k^{1/2}}{|c_k|} < \infty, \quad \text{and} \quad \frac{1}{t_n}\sum_{k\in\mathcal{K}_n}\frac{x_k^{3/2} w_k}{|c_k|^2} = O(1)$$

hold. Then, under both the null H_0 and the alternative H_1, we have

$$\sup_{y\in\mathbb{R}}\left|P\left\{\frac{n}{t_n}\left(T_n^* - \mathfrak{B}_n^* - \mathfrak{R}_n^*\right) \leq y \mid \mathcal{S}_n\right\} - P_{H_0}\left\{\frac{n}{t_n}\left(T_n - \mathfrak{B}_n - \mathfrak{R}_n\right) \leq y\right\}\right| \xrightarrow{\mathbb{P}} 0,$$

where $\mathfrak{B}_n^*$ and $\mathfrak{R}_n^*$ denote the bootstrap versions of $\mathfrak{B}_n$ and $\mathfrak{R}_n$ defined in Theorem 3.2 and P_{H_0} is the distribution of $\frac{n}{t_n}(T_n - \mathfrak{B}_n - \mathfrak{R}_n)$ under H_0.

Based on this result, as $T_n, T_n^* \geq 0$ and both have asymptotically the same bias and variance, we can construct a one-sided bootstrap test of size $\gamma \in (0, 1)$ for the null H_0 in (2.2), by rejecting H_0 if $T_n > q_{1-\gamma}^*$, where $q_{1-\gamma}^*$ is the $(1-\gamma)$ bootstrap quantile from Step 3.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

5. Monte Carlo Simulations

In this section, we investigate the finite sample behavior of the asymptotic test proposed in Section 3 and its bootstrap versions from Section 4 under several degrees of endogeneity and for different slope functions. We generate data from the model

$$X(t) = \cos(t)A + \sin(t)B, \quad W(t) = \cos(t)C + \sin(t)D + H$$

and $Y = \frac{1}{p+1}\sum_{l=0}^{p} X(l/p) \cdot \beta(l/p) + U$, for $p = 50$. To control all correlations in the model, we generate i.i.d. copies of

$$\begin{pmatrix} A \\ B \\ C \\ D \\ \varepsilon \end{pmatrix} \sim \mathcal{N}_5 \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 & 0 & \nu\sqrt{6} & 0 & \rho\sqrt{3} \\ 0 & 3 & 0 & \nu\sqrt{6} & 0 \\ \nu\sqrt{6} & 0 & 2 & 0 & 0 \\ 0 & \nu\sqrt{6} & 0 & 2 & 0 \\ \rho\sqrt{3} & 0 & 0 & 0 & 1 \end{pmatrix} \right\}$$

such that $corr(A, C) = \nu = corr(B, D)$, $corr(A, \varepsilon) = \rho$ and $U = \frac{7}{5}\varepsilon$. The random variable H is uniformly distributed on $(-1/2, 1/2)$ and independent of $(A, B, C, D, \varepsilon)'$. The parameter ρ controls the severity of endogeneity with $\rho = 0$ representing the exogenous case under H_0, while ν controls the strength of the instrument W. It is straightforward to show, that (X, W) is second-order stationary, which meets Assumption 1. Additional results for non-stationary (X, W) indicating a good performance can be also found in the supplementary material. Before investigating the performance of the proposed tests for different slope functions, we begin with a short motivation why it is reasonable to test for endogeneity instead of just always using the instrumental variable estimator. Consider the slope function $\beta(t) = \sin(2\pi t)$ and a sample size of $n = 100$ in the exogeneous functional linear model. As shown in Figure 1a, even in this simple situation the instrumental variable estimator has problems in estimating the true slope function. There, the black line represents the true slope function. The estimator $\widehat{\beta}$ is given by the shortly dashed line while the dotted, dot dashed

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

and long dashed lines represent $\widehat{\beta}_{IV}$ for an instrument with correlation $\nu = 0.6, 0.4, 0.2$ with X, respectively. All regularization parameters are chosen data driven by cross validation. It can be observed, that the estimation by $\widehat{\beta}_{IV}$ gets worse with decreasing correlation. In the following simulation study, we will use three different slope functions β_1, β_2 and β_3 defined by

$$\begin{aligned} \beta_1(t) &= \sin(4\pi t) + \frac{1}{2}\sin(8\pi t) + \frac{1}{7}\sin(20\pi t), \\ \beta_2(t) &= \frac{2}{\pi}\arcsin(\cos(2\pi t)), \\ \beta_3(t) &= \sum_{j\in\mathbb{Z}} \int_{\mathbb{R}} r_j(s)\frac{1}{h}k_j((t-s)/h)\, ds, \end{aligned} \tag{5.1}$$

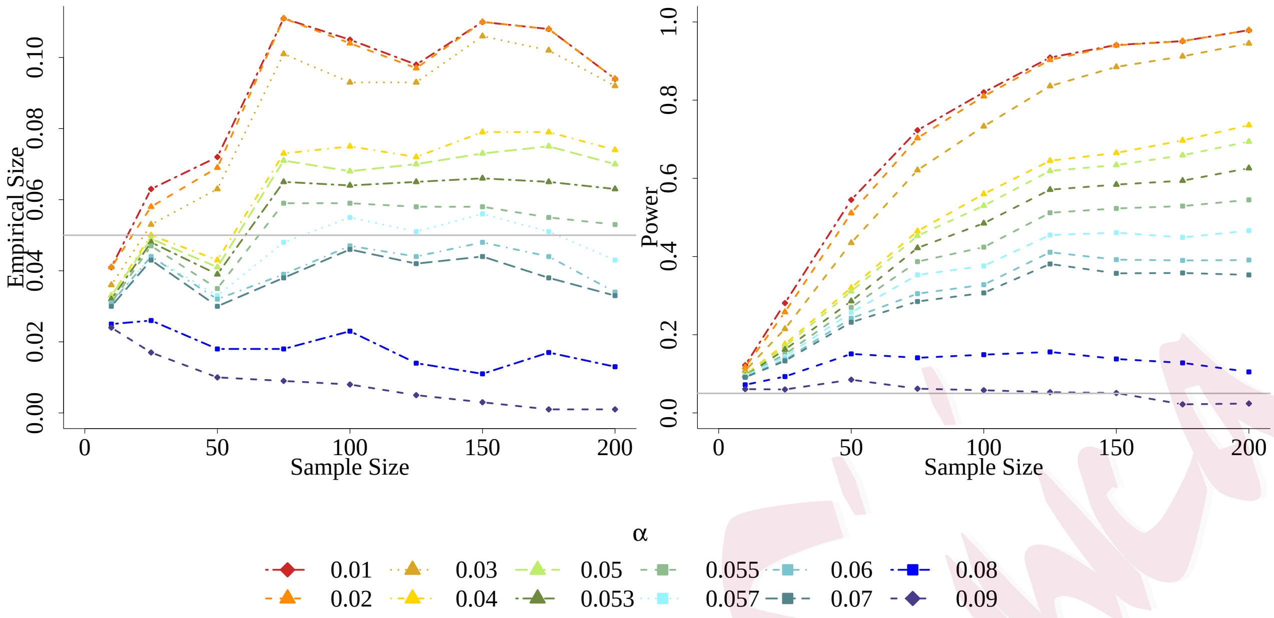
where $r_k(t) = I_{[j+\frac{1}{4}, j+\frac{3}{4})}(t)$, $k_j(t) = \frac{1}{C}\exp\left(-\frac{1}{1-(t-2j)^2}\right) I_{(-1+2j,2j+1)}(t)$ and $C = \int_{\mathbb{R}} k_0(s)\, ds$ which are illustrated in Figure 1b. For all simulations, we generate 1000 Monte Carlo realizations and use $B = 500$ bootstrap replications. Besides an Efron-type residual-based bootstrap, which draws the bootstrap errors U_i^*, $i = 1, \ldots, n$ independently with replacement from the residuals $\widehat{U}_1, \ldots, \widehat{U}_n$, we consider also several versions of a residual-based wild bootstrap, where $U_i^* = V_i\widehat{U}_i, \quad i = 1, \ldots, n$, and the V_i's are i.i.d. with $\mathrm{E}[V_1] = 0$ and $\mathrm{E}[V_1^2] = 1$ independent of $(X_i, W_i, Y_i)_{i=1,\ldots,n}$. We consider three different choices (a) - (c) for the distribution of the V_i's, which are commonly used in the literature, see e.g. Mammen (1993),

$$(a)\ P\left(V_1 = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}}, \quad P\left(V_1 = \frac{\sqrt{5}+1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}, \tag{5.2}$$

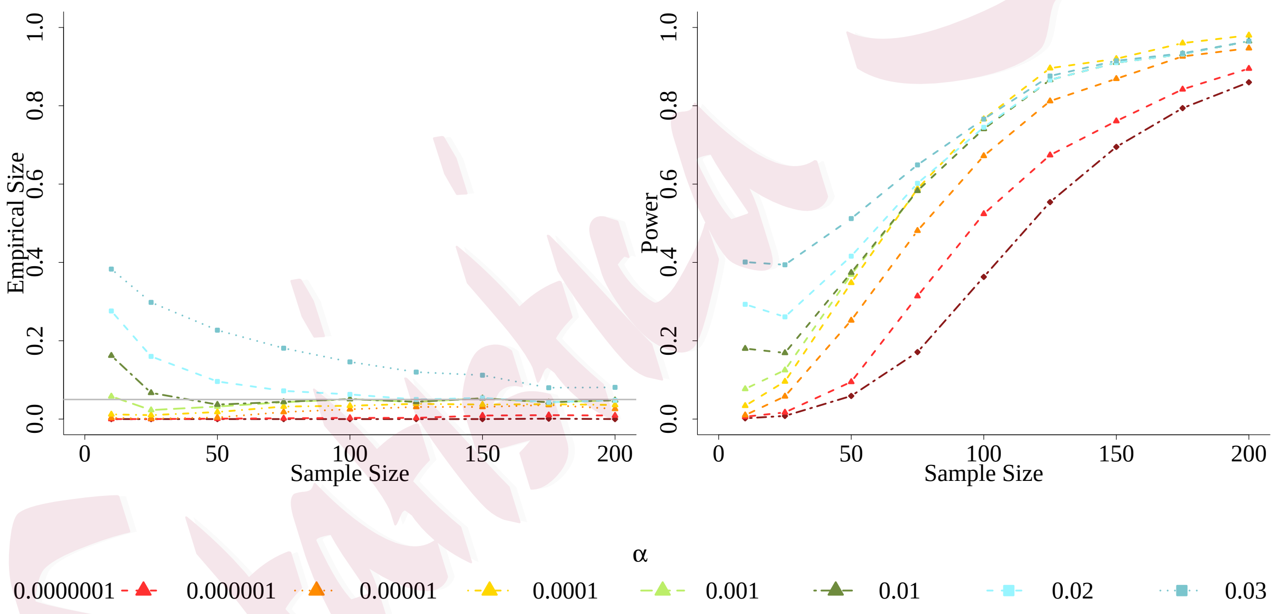
$$(b)\ P(V_1 = 1) = 0.5 = P(V_1 = -1), \quad (c)\ V_1 \sim \mathcal{N}(0,1). \tag{5.3}$$

The results for the asymptotic test for different choices of α are shown for the slope parameter function β_1 in Figure 2a, while the behavior for β_2 and β_3 with respect to the choice of α is very similar. Simulation results for all slope functions and fixed α are presented in Table

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154



(a) asymptotic test



(b) naive residual bootstrap

Figure 2: For the true slope parameter function β_1, empirical size and power of the tests for several choices of α is shown. The solid line shows the target level $\gamma = 0.05$.

1. From Figure 2a, we see that the best results are obtained for α between 0.05 and 0.055. For smaller α, the test does not hold the prescribed level, while for larger α the power is comparably small up to biased tests for α larger than 0.07. Even for a good choice of α, the asymptotic test has only moderate power, also for larger sample sizes. This is a well known effect with asymptotic tests using plug-in estimators.

The way out is typically a bootstrap-based test. The results for the residual-based bootstrap proposed in Section 4 and again β_1 are shown in Figure 2b. It turns out, that the regularization parameter can be chosen considerably smaller than for the asymptotic test and the procedure is much more robust with respect to the choice of α. Nearly all tests hold the size of $\gamma = 0.05$ for larger sample sizes and the power increases with sample size for most choices of α up to a value close to 1 already for $n = 200$. Again we can get an idea of choosing a good α depending on the sample size which varies from $\alpha = 0.01$ for $n = 25, 50$ to $\alpha = 0.0001$ for $n = 75, 100, 200$ and 300. Apparently, all bootstrap procedures discussed in Section 4 perform comparably well, which can be seen in Figure 3 for a choice of $\alpha = 0.0001$.

Comparing the performance of the bootstrap test for different slope functions, we find that the bootstrap test holds the size $\gamma = 0.05$ for all models, while we see in Table 1 that the power is similarly good for all settings with only slight disadvantages for the smoothed indicator function β_3. Finally, we inspect the effect of the degree of endogeneity and the strength of the instrument on the performance of the test. In the left panel of Figure 4, we see that the power of the bootstrap test increases with increasing degree ρ of endogeneity being already acceptable for $\rho = 0.3$. The middle and right panels of Figure 4 show, that the performance of the test highly depends on the strength of the instrument. If the instrument is weak, i.e. the correlation with the errors is low, the power is also low and the test does not

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

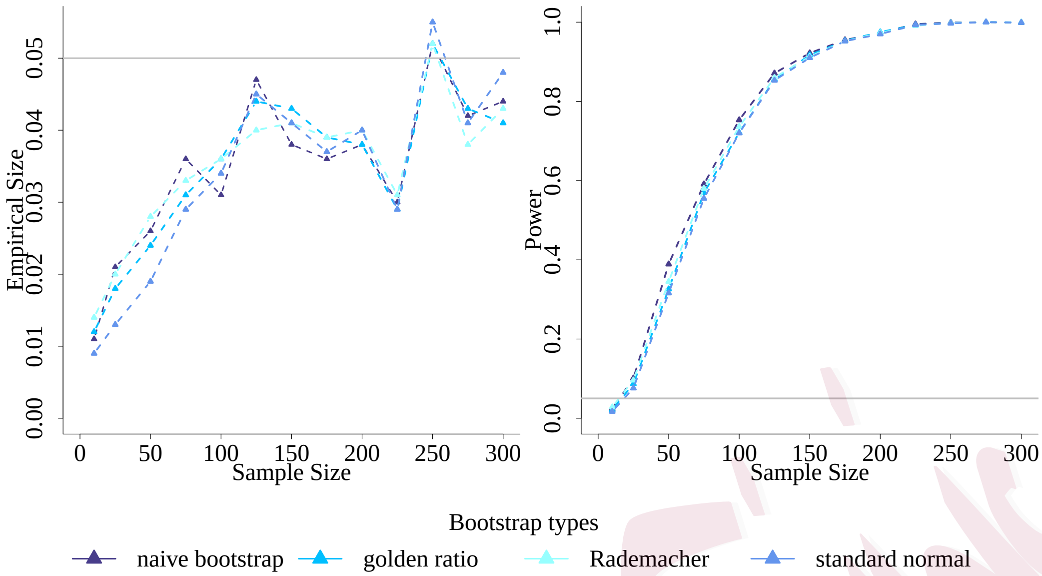


Figure 3: Empirical size and power of the different bootstrap tests proposed in (5.2)-(5.3) for the true slope function β_1 and for several choices of α. The solid line shows the target level $\gamma = 0.05$.

hold the size. It turns out, that for the setting with slope function β_1, $\rho = 0.4$ and $\alpha = 0.0001$, the bootstrap test performs best for a strength of the instrument around $\nu = 0.7$.

We have seen, that several parameters substantially influence the performance of the test. While some of them are determined by the model itself such as the slope function or the correlation of regressors and errors, others have to be chosen at least partly by the user. Since there are typically only few reasonable instruments available, the user has only limited possibilities to improve the performance via choosing a stronger instrument. The case is different for the regularization parameter. To get a better overview, α is kept fixed with growing n in the simulations although this does not meet our assumptions in theory. As an effect, we observe in some cases an empirical power, which does not increase for growing sample size (see e.g. Fig. 2a for $n = 175$). Since the optimal regularization parameter is only a theoretical value, a data driven choice is generally desirable. As for other estimation

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

	n	25	50	75	100	125	150	175	200	225	250	275	300
β_1	$\rho = 0.4$	0.125	0.369	0.583	0.741	0.866	0.91	0.932	0.965	0.976	0.986	0.993	0.996
	$\rho = 0$	0.023	0.032	0.043	0.051	0.044	0.053	0.043	0.048	0.052	0.042	0.05	0.037
β_2	$\rho = 0.4$	0.14	0.406	0.63	0.76	0.88	0.932	0.944	0.975	0.981	0.991	0.995	0.996
	$\rho = 0$	0.023	0.033	0.04	0.051	0.055	0.053	0.039	0.047	0.054	0.044	0.051	0.0311
β_3	$\rho = 0.4$	0.14	0.424	0.634	0.786	0.865	0.909	0.953	0.976	0.984	0.992	0.996	0.997
	$\rho = 0$	0.022	0.035	0.036	0.047	0.044	0.069	0.057	0.062	0.054	0.042	0.068	0.053

Table 1: Empirical power and size of the naive residual based bootstrap test for the slope functions defined in (5.1) using $\rho = 0.4$, $\nu = 0.6$ and $\alpha = 0.0001$.

techniques, least squares cross validation would be a natural approach, that is, to choose α by minimizing

$$CV(\alpha) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle \widehat{\beta}_{-i}, X_i\rangle)^2 + \frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle \widehat{\beta}_{IV,-i}, X_i\rangle)^2,$$

where $\widehat{\beta}_{-i}$ and $\widehat{\beta}_{IV,-i}$, $i = 1, \ldots, n$, denote the classical and the instrumental variable estimator from the sample $\mathcal{S}_n$ with (X_i, W_i, Y_i) removed. The simulation results for the bootstrap test with slope function β_1 and α chosen by cross validation for different significance levels γ are presented in Table 2. The test using cross validation for the choice of α is slightly conservative. This might be due to the fact, that the optimal parameter for estimation is not always the ideal parameter for testing. This means, that choosing the regularization parameter for testing is even more challenging than choosing it for estimation. On the other hand, the power of the test is quite high even for moderate sample sizes like $n = 150$ and the test holds the size. Therefore, cross validation seems to be a reasonable method for a data-driven choice of the regularization parameter in the bootstrap testing approach. It is possible to

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

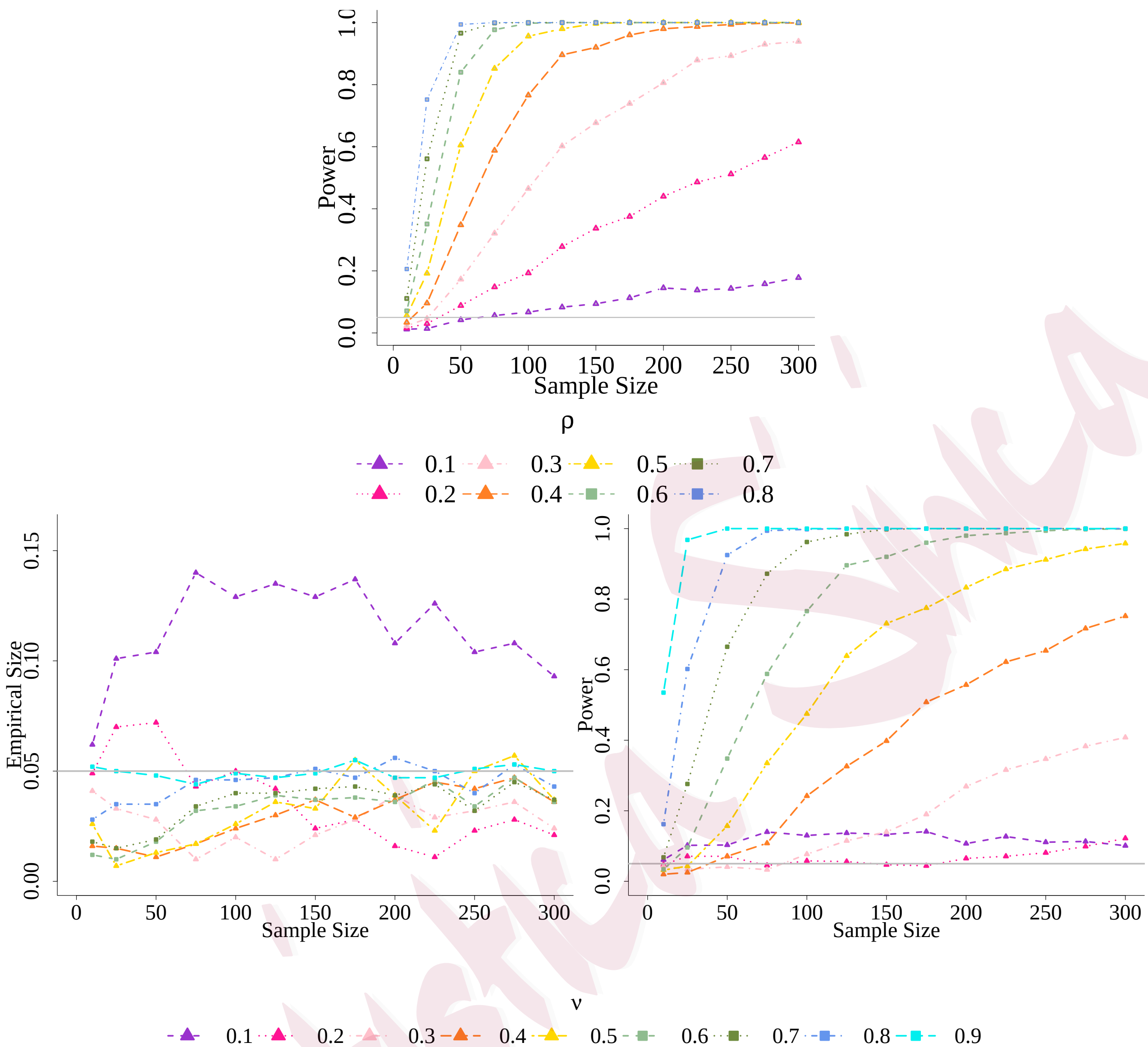


Figure 4: Simulation results for the naive bootstrap test. Above: Power for different degrees ρ of endogeneity, Below: Size and power for different strengths ν of the instrument.

still increase the power by using a slightly higher regularization parameter in the calculation of the bootstrap test statistics than for the original one. But, as we discovered in simulation studies the optimal order of the bootstrap regularization parameter highly depends on the underlying model without the possibility of choosing it data driven. For the asymptotic test

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

	n	25	50	75	100	125	150	175	200	225	250	275	300
$\gamma = 0.025$	$\rho = 0.4$	0.043	0.224	0.442	0.661	0.805	0.871	0.917	0.96	0.977	0.984	0.993	0.997
	$\rho = 0$	0.012	0.008	0.009	0.018	0.014	0.013	0.015	0.012	0.018	0.011	0.009	0.015
$\gamma = 0.05$	$\rho = 0.4$	0.108	0.363	0.586	0.768	0.891	0.911	0.957	0.974	0.983	0.993	0.997	0.999
	$\rho = 0$	0.017	0.021	0.03	0.037	0.028	0.027	0.027	0.03	0.039	0.026	0.034	0.028
$\gamma = 0.1$	$\rho = 0.4$	0.247	0.535	0.75	0.859	0.938	0.959	0.975	0.987	0.992	0.999	1	1
	$\rho = 0$	0.06	0.053	0.08	0.075	0.071	0.069	0.062	0.064	0.073	0.067	0.071	0.072

Table 2: Empirical power and sizes of the naive residual based bootstrap test for the slope function β_1 using $\rho = 0.4$, $\nu = 0.6$ and α chosen by cross validation considering significance levels $\gamma = 0.025$, 0.05 and 0.1.

the a cross validation approach is also possible and cures the observed effect of nonincreasing power for increasing sample size. But having in mind the latter discussion about optimal bandwidth choice for testig versus estimation and the sensibility to bandwidth choice of the asymptotic test discussed above combined with the addidional challenge of estimating bias and variance terms, we refrain from discussiong this test further.

6. Real Data Example

Florens and Van Bellegem (2015) analyze the impact of fertility on economic growth in the United Kingdom. There, fertility rates as functions of the mothers' age are considered for the years 1966 to 2012. Each curve is observed for the ages 15 to 44 and 45 and older. For each year, the corresponding GDP growth rate is given as real variable. Due to several publications, the fertility rate is considered to be an endogenous regressor for the growth

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

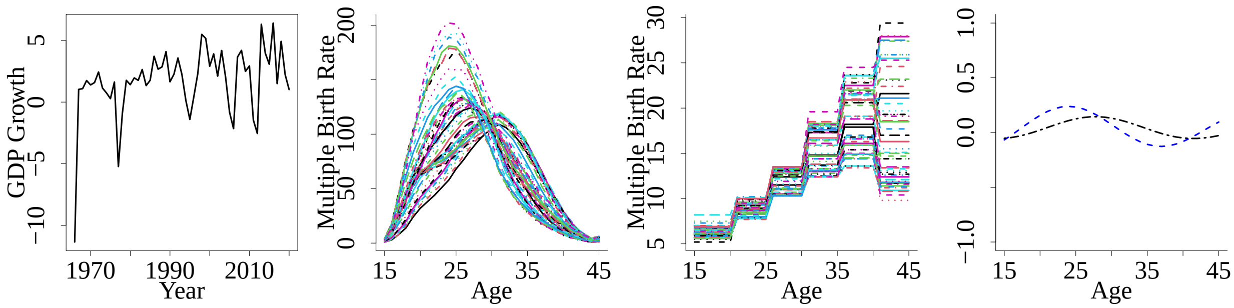


Figure 5: From left to right: GDP growth (annual %), Fertility Rates (live births per 1000 women), Multiple births rates (maternities with multiple births per 1000 maternities), $\widehat{\beta}$ (dashed) and $\widehat{\beta}_{IV}$ (dot dashed). Source: World Bank and Uk Office of National Statistics.

rate, see e.g. Braakmann and Wildman (2016), but this has never been tested in a functional context. Florens and Van Bellegem (2015) use multiple birth rates as instrumental variable interpreted as piecewise constant functions on the range of ages $[15, 20)$, $[20, 25)$, $[25, 30)$ $[30, 35)$, $[35, 40)$ $[40, 45)$ and older than 45. They analyze the data with the instrumental variable estimator proposed there. We use the same data, but for the augmented range of years 1966 to 2020. For the year 1981, the availability of birth rates is limited due to a registrars strike. Therefore, following Florens and Van Bellegem (2015), the year 1981 is excluded from the analysis. This results in $n = 54$ observations. For the test we have chosen $\alpha = 0.033$ by cross validation as described in Section 5. By using the Efron-type bootstrap approach with $B = 10000$ replications, we get a p-value of 0.095. This indicates, that we cannot reject the null hypothesis of exogeneity for a significance level of $\gamma = 0.05$, but would reject for $\gamma = 0.1$. As a consequence, based on the test decision with $\gamma = 0.05$, one could use the classical estimator for the slope function instead of the less efficient instrumental variable estimator, whereas the instrumental variable estimator should be used insteas for $\gamma = 0.1$. The right panel in Figure 5 shows $\widehat{\beta}$ as dashed line and $\widehat{\beta}_{IV}$ as dot dashed line with

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

α chosen by cross validation for each estimator seperately. Both estimators are quite close together and exhibit very similar characteristics. By inspecting them closer one observes that the IV estimator has a steeper increase at the beginning but a moderater effect at the higher ages compared to the classical estimator. This might be an effect of endogeneity, but in view of the test decision for $\gamma = 0.05$ it may also be caused by the small sample effect for $\widehat{\beta}_{IV}$ under exogeneity described at the beginning of Section 5.

The whole evaluation took 76.94 *sec.* on a desktop PC with an Intel(R) Core(TM) i5-2500 CPU. A detailed runtime analysis for different sample sizes and bootstrap repetitions can be found in the supplementary material S1.2.

7. Concluding remarks

The underlying work is the first approach of testing for endogeneity in a functional regression setup. For this purpose, the classical Hausman test designed to test for endogeneity in multiple linear regression models has to be suitably modified. This modification is required, because the L_2-distance of two slope function estimators in functional linear regression models are shown to have no proper limiting distribution. We prove asymptotic normality for the proposed modified Hausman-type test statistic, which allows for the construction of asymptotic tests for exogeneity. As the asymptotic test has several drawbacks such as many nuisance parameters, which are cumbersome to estimate, an additional bias term, which diverges when multiplied with the rate of convergence, and a high sensitivity to the choice of the regularization parameter, we propose suitable bootstrap versions of the test to approximate the null distribution. This avoids the additional estimation of nuisance parameters and turns out to be much more robust to the choice of the regularization parameter in sim-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

ulations. This behavior is demonstrated in a detailed simulation study. Topics of ongoing work are the choice of the instrument, a data driven choice of the regularization parameter and the transfer to other regression models.

Supplementary Material

The "Supplement to Testing exogeneity in the functional linear regression model" contains additional simulation results and details of the proofs of Propositions A.5, C.1 and C.2.

Acknowledgments

We thank Jan Johannes for helpful discussions on the different estimation techniques in the functional linear regression model with and without endogeneity. Furthermore, we highly appreciate the comments of two anonymous referees which helped us improve the paper.

A. Auxiliary Results for the Proof of Theorem 3.2

We assume for the sake of simplicity $E[X(t)] = E[W(t)] = 0$ for all $t \in [0,1]$ and remember from Section 3 the decomposition of the test statistic with

$$R_{n,1} = \frac{1}{n^2}\sum_{k\in\mathbb{Z}}(\widehat{x}_k - x_k)\Big|\sum_{i=1}^{n} D_{i,k} I\{\lambda_k \geq \alpha\gamma_k^\nu\}\Big(\sigma U_i + \sum_{\substack{m\in\mathbb{Z},\\ |m|\neq|k|}} S_{i,m}\Big)\Big|^2,$$

$$R_{n,2} = \frac{1}{n^2}\sum_{k\in\mathbb{Z}} x_k I\{\lambda_k \geq \alpha\gamma_k^\nu\}\sum_{i=1}^{n}|D_{i,k}|^2\Big|\sigma U_i + \sum_{\substack{m\in\mathbb{Z},\\ |m|\neq|k|}} S_{i,m}\Big|^2,$$

$$R_{n,3} = \frac{1}{n^2}\sum_{k\in\mathbb{Z}} x_k I\{\lambda_k \geq \alpha\gamma_k^\nu\}\sum_{\substack{i,j=1,\\ i\neq j}}^{n} D_{i,k}\Big(\sigma U_i + \sum_{\substack{m\in\mathbb{Z},\\ |m|\neq|k|}} S_{i,m}\Big)\overline{D_{j,k}}\Big(\sigma U_j + \sum_{\substack{m\in\mathbb{Z},\\ |m|\neq|k|}} \overline{S_{j,m}}\Big),$$

$$R_{n,4} = \frac{1}{n^3}\sum_{\substack{k,l\in\mathbb{Z},\\ |k|\neq|l|}}\sum_{j=1}^{n}\langle\phi_k, X_j\rangle\langle X_j,\phi_l\rangle I\{\lambda_k \geq \alpha\gamma_k^\nu\} I\{\lambda_l \geq \alpha\gamma_l^\nu\}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

$$\times \sum_{i=1}^{n} D_{i,k}\Big(\sigma U_i + \sum_{\substack{m\in\mathbb{Z},\\ |m|\neq|k|}} S_{i,m}\Big)\overline{D_{i,l}}\Big(\sigma U_i + \sum_{\substack{m\in\mathbb{Z},\\ |m|\neq|l|}} \overline{S_{i,m}}\Big),$$

$$R_{n,5} = \frac{1}{n^3}\sum_{\substack{k,l\in\mathbb{Z},\\ k\neq l}}\sum_{j=1}^{n}\langle\phi_k, X_j\rangle \sum_{\substack{i_1,i_2=1,\\ i_1\neq i_2}}^{n} D_{i_1,k}\Big(\sigma U_{i_1} + \sum_{\substack{m\in\mathbb{Z},\\ |m|\neq|k|}} S_{i_1,m}\Big)\overline{D_{i_2,l}}\Big(\sigma U_{i_2} + \sum_{\substack{m\in\mathbb{Z},\\ |m|\neq|l|}} \overline{S_{i_2,m}}\Big) \quad \text{(A.1)}$$

and define

$$D_{i,k,n} = \frac{\langle W_i, \phi_k\rangle}{\widehat{c}_k} I\{\widehat{w}_k \geq \alpha\} - \frac{\langle X_i, \phi_k\rangle}{\widehat{x}_k},\ D_{i,k} = \frac{\langle W_i, \phi_k\rangle}{c_k} - \frac{1}{x_k}\langle X_i, \phi_k\rangle,\ S_{i,m} = \langle\beta, \phi_m\rangle\langle\phi_m, X_i\rangle.$$

The first result establishes the asymptotic distribution of the test statistic.

Theorem A.1. *Under the assumptions of Theorem 3.2, under the null hypothesis, and for* $(X, W) \in \mathcal{F}_\eta^4$ *and* $E\{X(t)\} = E\{W(t)\} = 0$ *for all* $t \in [0, 1]$, *we have* $\frac{n}{t_n} R_{n,3} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathfrak{V})$.

The remaining results are required to show that the remainder terms are negligible.

Proposition A.2. *Under the assumptions of Theorem 3.2, and if* $(X, W) \in \mathcal{F}_\eta^{128}$ *and* $\mathrm{E}|U|^{128} \leq \eta < \infty$, *we have* $\frac{1}{n}\sum_{j=1}^n |\langle T_{n,1}, X_j\rangle|^2 = o_P\left(\frac{1}{n}\right)$.

Proposition A.3. *Under the assumptions of Theorem 3.2 and if* $(X, W) \in \mathcal{F}_\eta^{64}$ *and* $\mathrm{E}|U|^{64} \leq \eta < \infty$, *we have* $\frac{1}{n}\sum_{j=1}^n |\langle T_{n,2}, X_j\rangle|^2 = o_P\left(\frac{t_n}{n}\right)$.

Proposition A.4. *Under the assumptions of Theorem 3.2, and if* $(X, W) \in \mathcal{F}_\eta^{8}$ *and* $\mathrm{E}|U|^{8} \leq \eta < \infty$, *we have* $\frac{1}{n}\sum_{j=1}^n |\langle T_{n,3}, X_j\rangle|^2 = o_P\left(\frac{t_n}{n}\right)$.

Proposition A.5. *Under the assumptions of Theorem 3.2, and if* $E|U|^4 \leq \eta < \infty$ *and* $(X, W) \in \mathcal{F}_\eta^4$, *we have*

$$R_{n,1} = o_P\left(\frac{1}{n}\right), R_{n,4} = o_P\left(\frac{1}{n^{3/2}}\right), R_{n,5} = o_P\left(\frac{1}{n}\right), R_{n,2} = \mathfrak{R}_n + o_P\left(\frac{t_n}{n}\right), \mathfrak{R}_n = o\left(\frac{1}{\sqrt{n}}\right).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

B. Auxiliary results

The results in this section are used at several places in the proofs. They follow from Lemma A.1 in Johannes (2016).

Lemma B.1. *Let X and W have finite second moments and let $m \in \mathbb{N}$. Then, we have $\sum_{k\in\mathbb{Z}} x_k^{2m} < \infty$ and $\sum_{k\in\mathbb{Z}} x_k^{2m} w_k < \infty$. If additionally $X \in \mathcal{G}_\eta^{2m}$ und $\beta \in L_2([0,1])$, we have*

$$\mathrm{E}\left|\sum_{k\in\mathbb{Z}} \langle \beta, \phi_k\rangle\langle\phi_k, X\rangle\right|^{2m} < \infty.$$

Lemma B.2. *Let $p \in \mathbb{N}$ be fixed and suppose $(X, W) \in \mathcal{F}_\eta^{8p}$ and $\mathrm{E}|U|^{8p} \le \eta < \infty$. Then, there is a positive constant $C = C_p < \infty$ such that, for all $k \in \mathbb{Z}$, we have*

$$\mathrm{E}|I\{\widehat{\lambda}_k \ge \alpha\gamma_k^\nu\}\left(D_{i,k,n} - D_{i,k}\right)|^p \le \frac{C}{n^{p/2}}\left(\frac{w_k^p x_k^{p/2}}{|c_k|^{2p}} + \frac{1}{x_k^{p/2}}\right)(1 + o(1)) \quad \textit{and} \tag{B.1}$$

$$\mathrm{E}\left|I\{\widehat{\lambda}_k \ge \alpha\gamma_k^\nu\}D_{i,k,n}\right|^p \le C_p\left\{\frac{w_k^{p/2}}{|c_k|^p} + \frac{1}{x_k^{p/2}} + \frac{C}{n^{p/2}}\left(\frac{w_k^p x_k^{p/2}}{|c_k|^{2p}} + \frac{1}{x_k^{p/2}}\right)(1 + o(1))\right\}. \tag{B.2}$$

C. Proof of Theorem A.1

The proof follows by using a central limit theorem for martingale difference sequences with respect to $(\mathcal{F}_{n,j})_{n\in\mathbb{N},0\le j\le n}$, where $\mathcal{F}_{n,j} = \sigma\left(X_1, W_1, Y_1, \ldots, X_j, W_j, Y_j\right)$ and $\mathcal{F}_{n,0} = \sigma\left(\emptyset, \Omega\right)$, see Hall and Heyde (1980), Theorem 3.2 and Corollary 3.1, for

$$\frac{n}{t_n} R_{n,3} = \sum_{j=2}^{n} \frac{1}{t_n n} \sum_{k\in\mathbb{Z}} \mathscr{U}_{j,k} D_{j,k} \sum_{i=1}^{j-1} \overline{\mathscr{U}_{i,k} D_{i,k}} x_k I\{\lambda_k \ge \alpha\gamma_k^\nu\} = \sum_{j=2}^{n} Y_{n,j},$$

where

$$Y_{n,j} = \frac{1}{t_n n} \sum_{k\in\mathbb{Z}} \mathscr{U}_{j,k} D_{j,k} Z_{n,j,k}, \text{ and } Z_{n,j,k} = \sum_{i=1}^{j-1} \overline{\mathscr{U}_{i,k} D_{i,k}} x_k I\{\lambda_k \ge \alpha\gamma_k^\nu\}.$$

In a first step, we consider the conditional variance of the martingale difference scheme.

Proposition C.1. *Under the assumptions of Theorem 3.2, under the null hypothesis and for $(X, W) \in \mathcal{F}_\eta^4$, we have $\mathfrak{V}_n = \sum_{j=2}^n \mathrm{E}\left(Y_{n,j}^2 \mid \mathcal{F}_{n,j-1}\right) \xrightarrow{P} \mathfrak{V}$ as $n \to \infty$.*

Proof. Using that $\mathscr{U}_{j,k}D_{j,k}\overline{\mathscr{U}_{j,l}D_{j,l}}$ is independent of $(\mathcal{F}_{n,j-1})_{j=1,\dots,n}$, we can decompose

$$\begin{aligned}
\mathfrak{V}_n &= \frac{1}{t_n^2 n^2}\sum_{j=2}^{n} \mathrm{E}\Big(\Big|\sum_{k\in\mathbb{Z}} \mathscr{U}_{j,k}D_{j,k}Z_{n,j,k}\Big|^2 \mid \mathcal{F}_{n,j-1}\Big)\\
&= \frac{1}{t_n^2 n}\sum_{k\in\mathbb{Z}} x_k\left(\frac{x_k w_k}{|c_k|^2}-1\right) I\{\lambda_k \geq \alpha\gamma_k^{\nu}\}\mathrm{E}|\mathscr{U}_{1,k}|^2\left(\sum_{i=1}^{n-1}|\mathscr{U}_{i,k}D_{i,k}|^2 + \sum_{\substack{i,p=1,\\ i\neq p}}^{n-1} \mathscr{U}_{i,k}D_{i,k}\overline{\mathscr{U}_{p,k}D_{p,k}}\right)\\
&= \mathfrak{V}_{n,1}+\mathfrak{V}_{n,2}.
\end{aligned}$$

We define $\mathfrak{H}_n = \frac{\mathfrak{V}}{t_n^2 n}\sum_{k\in\mathbb{Z}} x_k\left(\frac{x_k w_k}{|c_k|^2}-1\right) I\{\lambda_k \geq \alpha\gamma_k^{\nu}\}\sum_{i=1}^{n-1}\mathrm{E}|D_{i,k}|^2$ and show $\mathfrak{V}_{n,1} = \mathfrak{H}_n + o_P(1)$ by proving the corresponding L_2-convergence and afterwards that $\mathfrak{H}_n$ converges in probability to $\mathfrak{V}$. For $i\in\{1,\dots,n\}$ and $k\in\mathbb{Z}$, we have

$$|\mathscr{U}_{i,k}D_{i,k}|^2\mathrm{E}|\mathscr{U}_{1,k}|^2 - \mathfrak{V}\mathrm{E}|D_{i,k}|^2 = \mathfrak{V}^{1/2}\Big[|\mathscr{U}_{i,k}D_{i,k}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,k}|^2\Big] - |\mathscr{U}_{i,k}D_{i,k}|^2|\langle\beta,\phi_k\rangle|^2 x_k.$$

and, observing that $\sigma^2+\sum_{m\in\mathbb{Z}}|\langle\beta,\phi_m\rangle|^2 x_m \leq C_1$ for some constant $C_1>0$, we get $\mathrm{E}\left(\mathfrak{V}_{n,1}-\mathfrak{H}_n\right)^2 \leq \mathbb{V}_{n,1}+\mathbb{V}_{n,2}+\mathbb{V}_{n,3}$, where

$$\begin{aligned}
\mathbb{V}_{n,1} &= \frac{C}{t_n^4 n^2}\sum_{k\in\mathbb{Z}} x_k^2\left(\frac{x_k w_k}{|c_k|^2}-1\right)^2 I\{\lambda_k \geq \alpha\gamma_k^{\nu}\}\Bigg\{\sum_{i=1}^{n-1}\mathrm{E}\Big(|\mathscr{U}_{i,k}D_{i,k}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,k}|^2\Big)^2\\
&\qquad + \sum_{\substack{i,p=1,\\ i\neq p}}^{n-1}\mathrm{E}\Big[|\mathscr{U}_{i,k}D_{i,k}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{1,k}|^2\Big]\mathrm{E}\Big[(|\mathscr{U}_{p,k}D_{p,k}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{1,k}|^2\Big]\Bigg\},\\
\mathbb{V}_{n,2} &= \frac{C}{t_n^4 n^2}\sum_{\substack{k,l\in\mathbb{Z},\\ |k|\neq|l|}} x_k\left(\frac{x_k w_k}{|c_k|^2}-1\right) I\{\lambda_k \geq \alpha\gamma_k^{\nu}\} x_l\left(\frac{x_l w_l}{|c_l|^2}-1\right) I\{\lambda_l \geq \alpha\gamma_l^{\nu}\}\\
&\qquad \Bigg[\sum_{i=1}^{n-1}\mathrm{E}\Big\{\Big(|\mathscr{U}_{i,k}D_{i,k}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,k}|^2\Big)\Big(|\mathscr{U}_{i,l}D_{i,l}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,l}|^2\Big)\Big\}\\
&\qquad + \sum_{\substack{i,p=1,\\ i\neq p}}^{n-1}\mathrm{E}\Big(|\mathscr{U}_{i,k}D_{i,k}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,k}|^2\Big)\mathrm{E}\Big(|\mathscr{U}_{i,l}D_{i,l}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,l}|^2\Big)\Bigg],\\
\mathbb{V}_{n,3} &= \frac{2}{t_n^4 n^2}\mathrm{E}\Bigg\{\sum_{k\in\mathbb{Z}} x_k^2\left(\frac{x_k w_k}{|c_k|^2}-1\right)|\langle\beta,\phi_k\rangle|^2 I\{\lambda_k \geq \alpha\gamma_k^{\nu}\}\sum_{i=1}^{n-1}|\mathscr{U}_{i,k}D_{i,k}|^2\Bigg\}^2.
\end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

We have

$$\mathrm{E}|\mathscr{U}_{j,k}D_{j,k}|^2 = \left(\sigma^2 + \sum_{\substack{m\in\mathbb{Z},\\ |m|\neq|k|}} |\langle\beta,\phi_m\rangle|^2 x_m\right)\left(\frac{w_k}{|c_k|^2} - \frac{1}{x_k}\right), \tag{C.1}$$

because $|\mathscr{U}_{j,k}|^2$ and $|D_{j,k}|^2$ are uncorrelated for all $k \in \mathbb{Z}$ and all $j \in \{1,\ldots,n\}$. Hence, with Lemma B.1 and (C.6), for all $i \in \{1,\ldots,n\}$ and $k \in \mathcal{K}_n$, we have

$$\mathrm{E}\left(|\mathscr{U}_{i,k}D_{i,k}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,k}|^2\right)^2 \leq C\left\{\mathrm{E}|D_{1,k}|^4 - \left(\mathrm{E}|D_{1,k}|^2\right)^2\right\} \leq C\mathrm{E}|D_{1,k}|^4 \leq \frac{C}{\alpha^2}, \tag{C.2}$$

$$\mathrm{E}\left(|\mathscr{U}_{i,k}D_{i,k}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,k}|^2\right) = -\mathrm{E}|D_{1,k}|^2|\langle\beta,\phi_k\rangle|^2 x_k = -\left(\frac{x_k w_k}{|c_k|^2} - 1\right)|\langle\beta,\phi_k\rangle|^2. \tag{C.3}$$

For the mixed terms with $k,l \in \mathbb{Z}, |k| \neq |l|$ and $i \in \{1,\ldots,n\}$ and $\frac{w_k}{|c_k|^2} - \frac{1}{x_k} \geq 0$, for all $k \in \mathbb{Z}$, we get

$$\begin{aligned}
&\mathrm{E}\Big\{\Big(|\mathscr{U}_{i,k}D_{i,k}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,k}|^2\Big)\Big(|\mathscr{U}_{i,l}D_{i,l}|^2 - \mathfrak{V}^{1/2}\mathrm{E}|D_{i,l}|^2\Big)\Big\} \\
&\leq \mathrm{E}\Big(|\mathscr{U}_{1,k}D_{1,k}\mathscr{U}_{1,l}D_{1,l}|^2\Big) + \Big(\frac{w_k}{|c_k|^2} - \frac{1}{x_k}\Big)\Big(\frac{w_l}{|c_l|^2} - \frac{1}{x_l}\Big) \\
&\leq C\Bigg\{\frac{1}{\alpha^2}|\langle\beta,\phi_k\rangle|^2 x_k|\langle\beta,\phi_l\rangle|^2 x_l + \frac{x_l}{\alpha}|\langle\beta,\phi_l\rangle|^2\Big(\frac{w_k}{|c_k|^2} - \frac{1}{x_k}\Big) \\
&\qquad + \frac{x_k}{\alpha}|\langle\beta,\phi_k\rangle|^2\Big(\frac{w_l}{|c_l|^2} - \frac{1}{x_l}\Big) + \Big(\frac{w_k}{|c_k|^2} - \frac{1}{x_k}\Big)\Big(\frac{w_l}{|c_l|^2} - \frac{1}{x_l}\Big)\Bigg\}.
\end{aligned} \tag{C.4}$$

Using this, we have

$$\begin{aligned}
\mathbb{V}_{n,1} &\leq \frac{C}{t_n^4 n^2}\sum_{k\in\mathbb{Z}} x_k^2\left(\frac{x_k w_k}{|c_k|^2} - 1\right)^2 I\{\lambda_k \geq \alpha\gamma_k^\nu\}\left\{\frac{n}{\alpha^2} + n^2\left(\frac{x_k w_k}{|c_k|^2} - 1\right)^2|\langle\beta,\phi_k\rangle|^4\right\} \\
&\leq \frac{C}{t_n^4 n\alpha^2}\sum_{k\in\mathbb{Z}} x_k^2\left(\frac{x_k w_k}{|c_k|^2} - 1\right)^2 I\{\lambda_k \geq \alpha\gamma_k^\nu\} \\
&\quad + \frac{C}{t_n^4}\sum_{k\in\mathbb{Z}} x_k^2\left(\frac{x_k w_k}{|c_k|^2} - 1\right)^4|\langle\beta,\phi_k\rangle|^4 I\{\lambda_k \geq \alpha\gamma_k^\nu\} = o\left(1 + \frac{1}{t_n^2}\right),
\end{aligned}$$

for some constant $C > 0$. Using similar arguments, we obtain $\mathbb{V}_{n,2} = o\left(1 + \frac{1}{t_n^2} + \frac{1}{\sqrt{n}t_n}\right) + \mathcal{O}\left(\frac{1}{n}\right)$ and $\mathbb{V}_{n,3} = o\left(1 + \frac{1}{t_n^2} + \frac{1}{n} + \frac{1}{\sqrt{n}t_n}\right)$, which altogether results in $\mathfrak{V}_{n,1} = \mathfrak{H}_n + o_P(1)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

Finally, for $n \to \infty$, the stochastic convergence of $\mathfrak{H}_n$ follows by

$$\mathfrak{H}_n = \mathfrak{V}\frac{n-1}{t_n^2 n} \sum_{k\in\mathbb{Z}} \left(\frac{x_k w_k}{|c_k|^2} - 1\right)^2 I\{\lambda_k \geq \alpha\gamma_k^\nu\} \xrightarrow{P} \mathfrak{V}.$$

For proving, that $\mathfrak{V}_{n,2}$ converges to 0 in probability, we show again the corresponding L_2-convergence. To this end, for all $i \in \{1,\ldots,n\}$ and all $k \in \mathbb{Z}$, we bound the term $\mathrm{E}|\mathscr{U}_{1,k}|^2$ by a constant $C < \infty$ using the centeredness of U and Lemma B.1, to obtain $\mathfrak{V}_{n,2} = o_P(1)$. The detailed arguments can be found in the supplementary material. □

The second step is to show the conditional Lindeberg condition by verifying an unconditional Lyapunov condition.

Proposition C.2. *Under the assumptions of Theorem 3.2, under the null hypothesis, and with* $(X,W) \in \mathcal{F}_\eta^4$, *we have*

$$\forall\, \varepsilon > 0: \sum_{j=2}^n \mathrm{E}\left(Y_{n,j}^2 I\{|Y_{n,j}| > \varepsilon\} \mid \mathcal{F}_{n,j-1}\right) \xrightarrow{P} 0 \quad as \quad n \to \infty. \tag{C.5}$$

Proof. It is shown in Alj et al. (2014) and Gaenssler et al. (1978) that the conditional Lindeberg condition follows from the unconditional Lyapunov condition. We will show in the following, that $\sum_{j=2}^n \mathrm{E}|Y_{n,j}|^4 = o(1)$ and decompose $\sum_{j=2}^n \mathrm{E}|Y_{n,j}|^4 = L_{n,1} + L_{n,2} + L_{n,3} + L_{n,4}$, where

$$L_{n,1} = \frac{1}{t_n^4 n^4} \sum_{j=2}^n \sum_{k\in\mathbb{Z}} \mathrm{E}\left|\mathscr{U}_{j,k} D_{j,k} Z_{n,j,k}\right|^4,\ L_{n,2} = \frac{1}{t_n^4 n^4} \sum_{j=2}^n \sum_{\substack{k,l\in\mathbb{Z},\\ |k|\neq|l|}} \mathrm{E}\left|\mathscr{U}_{j,k} D_{j,k} Z_{n,j,k} \overline{\mathscr{U}_{j,l} D_{j,l} Z_{n,j,l}}\right|^2,$$

$$L_{n,3} = \frac{1}{t_n^4 n^4} \sum_{j=2}^n \sum_{\substack{k,l,q\in\mathbb{Z},\\ |k|,|l|\neq|q|,|k|\neq|l|}} \mathrm{E}\Big(|\mathscr{U}_{j,k} D_{j,k} Z_{n,j,k}|^2 \mathscr{U}_{j,l} D_{j,l} Z_{n,j,l} \overline{\mathscr{U}_{j,q} D_{j,q} Z_{n,j,q}}\Big),$$

$$L_{n,4} = \frac{1}{t_n^4 n^4} \sum_{j=2}^n \sum_{\substack{k,l,p,q\in\mathbb{Z},\\ |k|,|l|,|p|\neq|q|,\\ |k|,|l|\neq|p|,|k|\neq|l|}} \mathrm{E}\Big(\mathscr{U}_{j,k} D_{j,k} Z_{n,j,k} \overline{\mathscr{U}_{j,l} D_{j,l} Z_{n,j,l}} \mathscr{U}_{j,p} D_{j,p} Z_{n,j,p} \overline{\mathscr{U}_{j,q} D_{j,q} Z_{n,j,q}}\Big).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

For $L_{n,1}$, we use that for all $k \in \mathbb{Z}, n \in \mathbb{N}, j \in \{1, \dots, n\}$, $Z_{n,j,k}$ are stochastically independent of $\mathscr{U}_{j,k} D_{j,k}$ and that $\mathscr{U}_{j,k}$ and $D_{j,k}$ are uncorrelated. Further, the fourth absolute moment of $\mathscr{U}_{j,k}$ is uniformly bounded due to the centeredness of U and Lemma B.1. The fourth absolute moment of $D_{j,k}$ can be estimated using Assumption 3 and $(X, W) \in \mathcal{F}_\eta^4$ as

$$\mathrm{E}|D_{j,k}|^4 \leq C\left(\frac{\mathrm{E}|\langle W, \phi_k\rangle|^4}{|c_k|^4} + \frac{\mathrm{E}|\langle X, \phi_k\rangle|^4}{x_k^4}\right) \leq C\eta\left(\frac{w_k^2}{|c_k|^4} + \frac{1}{x_k^2}\right) \leq \frac{C\eta}{\alpha^2}. \tag{C.6}$$

By similar arguments, we get $\mathrm{E}\left|\mathscr{U}_{i_1,k} D_{i_1,k}\right|^2 = \mathrm{E}|\mathscr{U}_{i_1,k}|^2 \mathrm{E}|D_{i_1,k}|^2 \leq C\left(\frac{w_k}{|c_k|^2} - \frac{1}{x_k}\right)$ leading to

$$\begin{aligned}
&\mathrm{E}\Big|\sum_{i=1}^{j-1} \mathscr{U}_{i,k} D_{i,k} x_k I\{\lambda_k \geq \alpha\gamma_k^\nu\}\Big|^4 \\
&= x_k^4 I\{\lambda_k \geq \alpha\gamma_k^\nu\}\left(\sum_{i=1}^{j-1} \mathrm{E}|\mathscr{U}_{i,k}|^4 \mathrm{E}|D_{i,k}|^4 + 2\sum_{1\leq i_1<i_2\leq j-1} \mathrm{E}|\mathscr{U}_{i_1,k} D_{i_1,k}|^2 \mathrm{E}|\mathscr{U}_{i_2,k} D_{i_2,k}|^2\right) \\
&\leq \frac{Cn}{\alpha^2} x_k^4 I\{\lambda_k \geq \alpha\gamma_k^\nu\} + Cn^2 x_k^2\left(\frac{x_k w_k}{|c_k|^2} - 1\right)^2 I\{\lambda_k \geq \alpha\gamma_k^\nu\}.
\end{aligned} \tag{C.7}$$

Putting these results together, we get

$$\begin{aligned}
L_{n,1} &= \frac{1}{t_n^4 n^4} \sum_{j=2}^{n} \sum_{k\in\mathbb{Z}} \mathrm{E}|\mathscr{U}_{j,k}|^4 \mathrm{E}|D_{j,k}|^4 \mathrm{E}|Z_{n,j,k}|^4 \\
&\leq \frac{C}{t_n^4 n^4 \alpha^2} \sum_{j=2}^{n} \sum_{k\in\mathbb{Z}} \mathrm{E}\Big|\sum_{i=1}^{j-1} \mathscr{U}_{i,k} D_{i,k} x_k I\{\lambda_k \geq \alpha\gamma_k^\nu\}\Big|^4 \\
&\leq \frac{C}{t_n^4 n\alpha^2} \sum_{k\in\mathbb{Z}} x_k^2 I\{\lambda_k \geq \alpha\gamma_k^\nu\}\left\{\frac{1}{n\alpha^2} x_k^2 + \left(\frac{x_k w_k}{|c_k|^2} - 1\right)^2\right\} \\
&= o(1)\frac{1}{t_n^4}\left\{\sum_{k\in\mathbb{Z}} x_k^4 I\{\lambda_k \geq \alpha\gamma_k^\nu\} + \sum_{k\in\mathbb{Z}} x_k^2\left(\frac{x_k w_k}{|c_k|^2} - 1\right)^2 I\{\lambda_k \geq \alpha\gamma_k^\nu\}\right\},
\end{aligned}$$

where the first series converges due to Lemma B.1 and the second can be bounded by Ct_n^2.

Considering $L_{n,4}$, we use the stochastic independence of $Z_{n,j,k}$ and $\mathscr{U}_{j,l} D_{j,l}$ for all $k, l \in \mathbb{Z}$, which results in

$$\begin{aligned}
&\mathrm{E}\big(\mathscr{U}_{j,k} D_{j,k} Z_{n,j,k} \overline{\mathscr{U}_{j,l} D_{j,l} Z_{n,j,l}} \mathscr{U}_{j,p} D_{j,p} Z_{n,j,p} \overline{\mathscr{U}_{j,q} D_{j,q} Z_{n,j,q}}\big) \\
&= \mathrm{E}\big(\mathscr{U}_{j,k} D_{j,k} \overline{\mathscr{U}_{j,l} D_{j,l}} \mathscr{U}_{j,p} D_{j,p} \overline{\mathscr{U}_{j,q} D_{j,q}}\big] \mathrm{E}\big[Z_{n,j,k} \overline{Z_{n,j,l}} Z_{n,j,p} \overline{Z_{n,j,q}}\big).
\end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

The rest of the argumentation is just calculating the expectations using that, for all $j \in \{1, \dots, n\}$, $D_{j,k}, D_{j,l}, D_{j,p}$ and $D_{j,q}$ are uncorrelated with $S_{j,m}$ for all $m \in \mathbb{Z} \backslash \{m \in \mathbb{Z} : |m| = |k|, |l|, |p|, |q|\}$ and stochastically independent of U_j. Finally,

$$\mathrm{E}(S_{j,k} D_{j,k}) = \langle \beta, \phi_k \rangle \mathrm{E} \left\{ \langle \phi_k, X_j \rangle \left(\frac{\langle W_j, \phi_k \rangle}{c_k} - \frac{\langle X_j, \phi_k \rangle}{x_k} \right) \right\} = \langle \beta, \phi_k \rangle \left(\frac{c_k}{c_k} - \frac{x_k}{x_k} \right) = 0 \quad \text{(C.8)}$$

and, in the same way, $\mathrm{E}(\overline{S_{j,k}} D_{j,k}) = \mathrm{E}(S_{j,k} \overline{D_{j,k}}) = 0$, which gives $L_{n,4} = 0$.

With similar arguments as above, which can be found in the supplementary material, we get $L_{n,2} = o\left(\frac{1}{t_n^4} + \frac{1}{t_n^2 n} + \frac{1}{t_n^2} + \frac{1}{t_n \sqrt{n}} \right) + \mathcal{O}\left(\frac{1}{n} + \frac{1}{n^2} \right) = o(1)$ and $L_{n,3} = o\left(\frac{1}{t_n^2 n} \right)$. □

All remaining terms can be estimated with similar techniques. We exemplarily show the idea for Proposition A.5, that is for $R_{n,2}$, in the supplementary material.

D. Proofs of Theorems 3.2 and 3.4

D.1 Proof of Theorem 3.2

For the sake of simplicity, we assume that X is centered. If not, the additional bias term has to be taken into account as well as stated in the assertion of the theorem. We give a short overview of the proof. The used propositions and lemmas are stated and proven in the appendix. For the employed decomposition of the test statistic, we need several (modified) correlation operators of the instruments and X. We define $\mathcal{U}_n$, $\Delta_{W,n} : L_2([0,1]) \to \mathbb{R}$ by

$$\mathcal{U}_n f = \frac{1}{n} \sum_{i=1}^{n} (W_i \otimes U_i) f \quad \text{and} \quad \Delta_{W,n} f = \frac{1}{n} \sum_{i=1}^{n} (W_i \otimes Y_i) f,$$

and set

$$\widetilde{\mathcal{U}}_n = \frac{1}{n} \sum_{i=1}^{n} \langle \cdot, \widetilde{W}_i \rangle U_i = \frac{1}{n} \sum_{k \in \mathbb{Z}} \frac{\overline{c_k}}{w_k} \sum_{i=1}^{n} \langle \phi_k, W_i \rangle \langle \cdot, \phi_k \rangle U_i,$$
$$\widehat{\widetilde{\mathcal{U}}}_n = \frac{1}{n} \sum_{i=1}^{n} \langle \cdot, \widetilde{W}_{n,i} \rangle U_i = \frac{1}{n} \sum_{k \in \mathbb{Z}} \frac{\overline{\widehat{c}_k}}{\widehat{w}_k} I\{\widehat{w}_k \geq \alpha\} \sum_{i=1}^{n} \langle \phi_k, W_i \rangle \langle \cdot, \phi_k \rangle U_i.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

For the test statistic, the following decomposition holds

$$\frac{n}{t_n}T_n = \frac{1}{t_n}\sum_{j=1}^{n}\left|\langle T_{n,1}+T_{n,2}+T_{n,3},X_j\rangle\right|^2 + \frac{1}{t_n}\sum_{j=1}^{n}\langle T_{n,1}+T_{n,2}+T_{n,3},X_j\rangle\langle X_j,R_n\rangle$$
$$+\frac{1}{t_n}\sum_{j=1}^{n}\langle X_j,T_{n,1}+T_{n,2}+T_{n,3}\rangle\langle R_n,X_j\rangle + \frac{1}{t_n}\sum_{j=1}^{n}\left|\langle R_n,X_j\rangle\right|^2, \tag{D.1}$$

where

$$T_{n,1} = \left(\widetilde{\Gamma}_n^{\dagger}\widehat{\widetilde{\mathcal{U}}}_n - \Gamma_{X,n}^{\dagger}\mathcal{U}_{X,n}\right) - \widehat{\Pi}_{\mathcal{K}_n}\left(\widetilde{\Gamma}^{\dagger}\widetilde{\mathcal{U}}_n - \Gamma_X^{\dagger}\mathcal{U}_{X,n}\right),$$
$$T_{n,2} = \left(\widetilde{\Gamma}_n^{\dagger}\widetilde{\Gamma}_n - \Gamma_{X,n}^{\dagger}\Gamma_{X,n}\right)\beta - \widehat{\Pi}_{\mathcal{K}_n}A_n,$$
$$T_{n,3} = \widehat{\Pi}_{\mathcal{K}_n}\left(\widetilde{\Gamma}^{\dagger}\widetilde{\mathcal{U}}_n - \Gamma_X^{\dagger}\mathcal{U}_{X,n} + A_n\right) - \left(\widetilde{\Gamma}^{\dagger}\widetilde{\mathcal{U}}_n - \Gamma_X^{\dagger}\mathcal{U}_{X,n} + A_n\right),$$
$$R_n = \widetilde{\Gamma}^{\dagger}\widetilde{\mathcal{U}}_n - \Gamma_X^{\dagger}\mathcal{U}_{X,n} + A_n$$

and

$$A_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{k\in\mathbb{Z}} D_{i,k}I\{\lambda_k \geq \alpha\gamma_k^{\nu}\}\sum_{\substack{m\in\mathbb{Z},\\|m|\neq|k|}} S_{i,m}\phi_k. \tag{D.2}$$

When subtracting $\frac{n}{t_n}\mathfrak{R}_n$, the last term in (D.1) can be further decomposed to get

$$\frac{1}{t_n}\sum_{j=1}^{n}\left|\langle R_n,X_j\rangle\right|^2 - \frac{n}{t_n}\mathfrak{R}_n = \frac{n}{t_n}R_{n,3} + \frac{n}{t_n}\left(R_{n,2}-\mathfrak{R}_n\right) + \frac{n}{t_n}\left(R_{n,1}+R_{n,4}+R_{n,5}\right), \tag{D.3}$$

where $R_{n,i}$, $i=1,\ldots,5$ as defined in (A.1). In Theorem A.1, we have shown that

$$\frac{n}{t_n}R_{n,3}=\frac{1}{nt_n}\sum_{k\in\mathbb{Z}}x_kI\{\lambda_k\geq\alpha\gamma_k^{\nu}\}\sum_{\substack{i,j=1,\\i\neq j}}^{n}D_{i,k}\Big(\sigma U_i + \sum_{\substack{m\in\mathbb{Z},\\|m|\neq|k|}} S_{i,m}\Big)\overline{D_{j,k}}\Big(\sigma U_j + \sum_{\substack{m\in\mathbb{Z},\\|m|\neq|k|}}\overline{S_{j,m}}\Big)$$

converges weakly to a normal distribution with mean 0 and variance $\mathfrak{V}$, while the remaining terms on the right-hand side of (D.3) are asymptotically negligible using Proposition A.5. Furthermore, the remaining terms on the right-hand side of (D.1) vanish due to Propositions A.2, A.3 and A.4 by using standard arguments that allow to handle also the mixed terms. Finally, the assertion follows from Slutsky's lemma.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0154

D.2 Proof of Theorem 3.4

We only consider the special case $\mu_X = 0$ here. The general case can be proven by similar arguments. Under H_1, $\widehat{\beta}$ is not consistently estimating β such that it converges in probability to $\beta + b$ for some $b \in L^2[0,1]$ with $b = \sigma \sum_{k\in\mathbb{Z}} \frac{\mathrm{E}(U_1\langle X_1,\phi_k\rangle)}{x_k} I\{\lambda_k \geq \alpha\gamma_k^\nu\}\phi_k(t)$ such that $b \neq 0$ under endogeneity by continuity imposed in Assumption 1. Hence, we have

$$T_n = \frac{1}{n}\sum_{i=1}^{n} |\langle \widehat{\beta}_{IV} - (\widehat{\beta} - b), X_i\rangle|^2 - \frac{2}{n}\sum_{i=1}^{n}\langle \widehat{\beta}_{IV} - (\widehat{\beta} - b), X_i\rangle\langle b, X_i\rangle + \frac{1}{n}\sum_{i=1}^{n}|\langle b, X_i\rangle|^2$$
$$= \frac{1}{n}\sum_{i=1}^{n} |\langle \widehat{\beta}_{IV} - (\widehat{\beta} - b), X_i\rangle|^2 - O_p\left(\sqrt{\frac{t_n}{n}}\right) + O_p(1).$$

The standardized version of the first part converges in distribution to a standard normal distribution by similar arguments as in Theorem 3.2 and Corollary 3.3, while the sum of the remainder terms multiplied with $\frac{n}{t_n}$ goes to infinity for $n \to \infty$. Consequently, we have

$$P\left(\frac{n}{\widehat{t}_n}\frac{T_n - \widehat{\mathfrak{R}}_n}{\sqrt{\widehat{\mathfrak{V}}_n}} > u_{1-\gamma}\right) \to 1 \text{ for } n \to \infty.$$

E. Proof of Theorem 4.1

Let $\Phi_{\mathfrak{V}}(\cdot)$ denote the distribution function of the normal distribution with mean zero and variance $\mathfrak{V}$, F_n the distribution function of $\frac{n}{t_n}(T_n - \mathfrak{B}_n - \mathfrak{R}_n)$ and $F^*_{\mathcal{S}_n,n}$ the distribution function of the conditional distribution of $\frac{n}{t_n}(T^*_n - \mathfrak{B}^*_n - \mathfrak{R}^*_n)$ given $\mathcal{S}_n$. By bounding

$$\sup_{t\in\mathbb{R}}\left|F^*_{\mathcal{S}_n,n}(t) - F_n(t)\right| \leq \sup_{t\in\mathbb{R}}\left|F^*_{\mathcal{S}_n,n}(t) - \Phi_{\mathfrak{V}}(t)\right| + \sup_{t\in\mathbb{R}}|F_n(t) - \Phi_{\mathfrak{V}}(t)| =: M_{1,n} + M_{2,n},$$

similar to the example in Section 29 of DasGupta (2008), it is enough to show the convergence of $M_{1,n}$ and $M_{2,n}$. Due to the continuity of $\Phi_{\mathfrak{V}}$, the convergence of $M_{2,n}$ directly follows from Theorem 3.2 and Polya's Theorem, as stated in Section 1.5.3 of Serfling (1980). Again, using Polya's Theorem, it is enough to show for $M_{1,n}$, that for all $\varepsilon > 0$

$\lim_{n\to\infty} \mathbb{P}\left(\left|F^*_{\mathcal{S}_n,n}(t) - \Phi_{\mathfrak{V}}(t)\right| > \varepsilon\right) = 0$. For this we just imitate the proof of Theorem 3.2. Analogously to (D.1), we decompose

$$\frac{n}{t_n}T_n^* = \frac{1}{t_n}\sum_{j=1}^n \left|\left\langle T^*_{n,1} + T^*_{n,2} + T^*_{n,3}, X_j\right\rangle\right|^2 + \frac{1}{t_n}\sum_{j=1}^n \langle T^*_{n,1} + T^*_{n,2} + T^*_{n,3}, X_j\rangle\langle X_j, R_n^*\rangle$$
$$+ \frac{1}{t_n}\sum_{j=1}^n \langle X_j, T^*_{n,1} + T^*_{n,2} + T^*_{n,3}\rangle\langle R_n^*, X_j\rangle + \frac{1}{t_n}\sum_{j=1}^n |\langle R_n^*, X_j\rangle|^2,$$

where, similar to the proof of Theorem 3.2, we get

$$\frac{1}{t_n}\sum_{j=1}^n |\langle R_n^*, X_j\rangle|^2 - \frac{n}{t_n}\mathfrak{R}_n = \frac{n}{t_n}R^*_{n,3} + \frac{n}{t_n}\left(R^*_{n,2} - \mathfrak{R}_n\right) + \frac{n}{t_n}\left(R^*_{n,1} + R^*_{n,4} + R^*_{n,5}\right).$$

Then, $\frac{n}{t_n}(R^*_{n,3} - \mathfrak{B}^*_n - \mathfrak{R}^*_n)$ converges weakly to $\mathcal{N}(0, \mathfrak{V})$ in probability along the same lines of Theorem A.1. The remainder terms can be discussed to be negligible with the same arguments as for the remainder terms in Theorem 3.2. □

References

Alj, A., Azrak, R., and Mélard, G. (2014). On conditions in central limit theorems for martingale difference arrays. *Economics Letters*, 123(3):305–307.

Braakmann, N. and Wildman, J. (2016). Reconsidering the effect of family size on labour supply: the twin problems of the twin birth instrument. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 179(4):1093–1115.

Cardot, H., Mas, A., and Sarda, P. (2007). CLT in functional linear regression models. *Probab. Theory Relat. Fields*, 138(3-4):325–361.

Chiou, J.-M., Yang, Y.-F., and Chen, Y.-T. (2016). Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis*, 146:301–312. Special Issue on Statistical Models and Methods for High or Infinite Dimensional Spaces.

Cuesta-Albertos, J. A., García-Portugués, E., Febrero-Bande, M., and González-Manteiga, W. (2019). Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes. *Ann. Stat.*, 47(1):439–467.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability.* New York, NY: Springer.

Dorn, M. (2021). *Tests auf Exogenität im funktionalen linearen Regressionsmodell unter schwacher Stationarität.* PhD thesis, Bayreuth.

Florens, J.-P. and Van Bellegem, S. (2015). Instrumental variable estimation in functional linear models. *J. Econom.*, 186(2):465–476.

Gaenssler, P., Strobel, J., and Stute, W. (1978). On central limit theorems for martingale triangular arrays. *Acta Math. Acad. Sci. Hung.*, 31:205–216.

García-Portugués, E., Álvarez-Liébana, J., Álvarez-Pérez, G., and González-Manteiga, W. (2020). Goodness-of-fit tests for functional linear models based on integrated projections. In *Functional and high-dimensional statistics and related fields. Selected papers presented at the 5th international workshop on functional and operatorial statistics, IWFOS 2021, Brno, Czech Republic, June 23–25, 2021*, pages 107–114. Cham: Springer.

García-Portugués, E., González-Manteiga, W., and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational and Graphical Statistics*, 23(3):761–778.

Hall, P. and Heyde, C. C. (1980). Martingale limit theory and its application. New York etc.: Academic Press.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Hörmann, S. and Jammoul, F. (2023). Prediction in functional regression with discretely observed and noisy covariates. *Computational Statistics & Data Analysis*, 178:107600.

Johannes, J. (2008). Nonparametric estimation in functional linear model. In *Functional and Operatorial Statistics*, pages 215–221, Heidelberg. Physica-Verlag HD.

Johannes, J. (2009). Nonparametric estimation in functional linear models with second order stationary regressors.

Johannes, J. (2016). Functional linear instrumental regression under second order stationarity.

Mair, B. A. and Ruymgaart, F. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5):1424–1444.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Stat.*, 21(1):255–285.

Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Stat.*, 33(2):774–805.

Neubauer, A. (1988). An a posteriori parameter choice for Tikhonov regularization in the presence of modeling error. *Appl. Numer. Math.*, 4(6):507–519.

Reiss, P. C. (2016). Just how sensitive are instrumental variable estimates? *Foundations and Trends® in Accounting*, 10:204–237.

Ruymgaart, F., Wang, J., Wei, S.-H., and Yu, L. (2011). Some asymptotic theory for functional regression and classification. *Adv. Decis. Sci.*, 2011:22.

Seong, D. and Seo, W.-K. (2021). Functional instrumental variable regression with an application to estimating the impact of immigration on native wages.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* John Wiley & Sons, Hoboken, NJ.

Tsybakov, A. B. (2004). *Introduction à l'estimation non-paramétrique.*, volume 41. Paris: Springer.

Wu, D.-M. (1974). Alternative tests of independence between stochastic regressors and disturbances: Finite sample results. *Econometrica*, 42(3):529–546.

Zhou, H., Yao, F., and Zhang, H. (2022). Functional linear regression for discretely observed data: from ideal to reality. *Biometrika*.

Department of Mathematics, Physics, and Computer Science, University of Bayreuth, 95440 Bayreuth, Germany

E-mail: manuela.dorn@uni-bayreuth.de

Department of Mathematics, Physics, and Computer Science, University of Bayreuth, 95440 Bayreuth, Germany

E-mail: melanie.birke@uni-bayreuth.de

Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany

E-mail: jentsch@statistik.tu-dortmund.de